

# Applied Physics

**Editor: Joan Peuteman**

Peteris Apse-Apsitis

Nikolay Fedosenko

Julia Fedotova

Julia Kasiuk

Dmitry Kovalenko

Ojars Krumins

Nadezhda Kunicina

Andrei Lavysh

Alexander Maskevich

Sergei Maskevich

Alexander Mazanik

Joan Peuteman

Leonids Ribickis

Alexander Rogachev

Natalia Strekal

Vitali Stsiapura

Iosif Sveklo

Anatolijs Zabasta

Anastasija Ziravecka

Riga, 2018

This document has been prepared by the financial support of European Union. The authors from Riga Technical University, KU Leuven, Grodno State University, Gomel State University and Belorussian State University are responsible for the content of this document. This publication reflects the views only of the authors, and it cannot be regarded as the European Union's official position.

The book is developed in a frame of the project “ERASMUS+ Capacity-building in the Field of Higher Education 2015 Call for Proposals EAC/A04/2014 561525-EPP-1-2015-1-LV-EPPKA2-CBHE-JP - ERASMUS+ CBHE.

The textbook is devised for students of applied physics and electrical engineering specialties, who study physics and engineering. The textbook can be useful for students and professionals focusing on abroad range of physics related issues. The book gives overview of current physical applications, measurement methods systems and elements, as well as explains their operating principles.

**Key Action:** Cooperation for innovation and the exchange of good practices

**Action:** Capacity Building in Higher Education

**Action Type:** Joint Projects

**Deliverable:** 2.3. Five electronic courses e-Books by the declared directions



Co-funded by the  
Erasmus+ Programme  
of the European Union

**Project Scientific Manager:** Nadezhda Kunicina

**Project Manager:** Anatolijs Zabašta

**Editor:** Joan Peuteman

**Institution:** KU Leuven

Under the Creative Commons Attribution license, the author(s) and users are free to share (copy and redistribute the material in any medium of format) and adapt (remix, transform and build upon the material for any purpose, even commercially) this work. The licensor cannot revoke these freedoms as long as you follow the license terms.

ISBN - 978-9934-22-153-8 (pdf)

## Contributors

**Peteris Apse-Apsitis**, professor assistant, senior researcher. Head of Division of Industrial Electronic equipment of Institute of Industrial Electronics and Electrical Engineering, Riga Technical University, Faculty of Power and Electrical Engineering, 12/1 Azenes Str. - 510., Riga, LV 1048, Latvia, tel. +371 67089917, peteris.apse-apsitis@rtu.lv

**Nikolay Fedosenko**, docent at Optics Department, Faculty of Physics and Information Technology, Francisk Skorina Gomel State University, 102 Sovetskaya Str., off. 1-11, Gomel, Belarus, 246019, tel. +375 232 57 54 51, fedosenko@gsu.by

**Julia Fedotova**, deputy-director of Research Institute for Nuclear Problems, Belarusian State University, head of the laboratory of Physics of Prospective Materials, 11 Bobrujskaya Str., room 313, 220030 Minsk, Belarus, tel. +375 17 2264224, julia@hep.by

**Julia Kasiuk**, leading researcher of the laboratory of Physics of Prospective Materials, Institute for Nuclear Problems, Belarusian State University, 11 Bobrujskaya Str., room 314, 220030 Minsk, Belarus, tel. +375 17 2264221, julia-nechaj@yandex.ru

**Dmitry Kovalenko**, dean of the Faculty of Physics and Information Technology, Francisk Skorina Gomel State University, 102 Sovetskaya Str., off. 2-9, Gomel, Belarus, 246019, tel. +375 232 57 65 57, dkov@gsu.by

**Ojars Krumins**, Dr., researcher of Institute of Industrial Electronics and Electrical Engineering, Riga Technical University, Faculty of Power and Electrical Engineering, 12/1 Azenes Str.-503, Riga, LV 1048, Latvia, hm.ojars@gmail.com

**Nadezhda Kunicina**, professor, Senior Researcher of Division of Industrial Electronic equipment, Institute of Industrial Electronics and Electrical Engineering, Riga Technical University, Faculty of Power and Electrical Engineering, 12/1 Azenes Str. - 503., Riga, LV 1048, Latvia, tel. +371 67089052, nadezda.kunicina@rtu.lv

**Andrei Lavysh**, senior lecturer of General Physics Department, Faculty of Physics and Technics, Yanka Kupala State University of Grodno, 22 Ozheshko str., Grodno, 230023, Belarus, tel. +375 152 743414, andrewlavysh@gmail.com

**Alexander Maskevich**, professor, Head of the Department of General Physics, Physico-technical Faculty Janka Kupala State University of Grodno, 22 Ozheshko str., Grodno 230023, Belarus, tel. +375 3240923, amaskevich@grsu.by

**Sergei Maskevich**, director of International Sakharov Environmental Institute of Belarusian State University, 23/1 Dolgobrodskaya str., Minsk 220070, Belarus, tel. +375 29 222 58 63, sergei.maskevich@gmail.com

**Alexander Mazanik**, professor assistant of Energy Physics Department, Faculty of Physics, Belarusian State University, Nezavisimosti Av. 4, Minsk 220030, Belarus, tel. +375 17 2095359, mazanikalexander@gmail.com

**Joan Peuteman**, professor, electrical engineering, KU Leuven University, Faculty of Engineering Technology, Mechatronics Group (M-Group), Spoorwegstraat 12, B-8200 Brugge, Belgium, joan.peuteman@kuleuven.be

**Leonids Ribickis**, professor, rector of Riga Technical University, Head of Institute of Industrial Electronics and Electrical engineering, Faculty of Power and Electrical Engineering, Riga Technical University, Scientific Head of Electromechatronics Scientific Laboratory, 1, Kalku Str. -217, Riga, LV 1658, Latvia, tel. +371 67089300, leonids.ribickis@rtu.lv

**Alexander Rogachev**, professor, corresponding member of National Academy of Sciences of Belarus, Director of Scientific Research Institute of Physics and Chemistry, Francisk Skorina Gomel State University, 102 Sovetskaya Str., off. 2-6, Gomel, Belarus, 246019, tel. +375 232 57-82-53, rogachyov@gsu.by

**Natalia Strekal**, professor of the Department of General Physics, Faculty of Physics and Technology, Grodno State University, 22 Ozheshko Str., Grodno, BY 230023, Belarus, tel. +375 336864600, nat@grsu.by

**Vitali Stsiapura**, leading research scientist, Department of Physics and Technics, Yanka Kupala State University of Grodno, 22 Ozheshko str., Grodno 230023, Belarus, tel. +375 152743414, stsiapura@grsu.by

**Iosif Sveklo**, present affiliation - assistant professor, Faculty of Physics of Bialystok University, ul. Ciołkowskiego 1L, 15-245 Białystok, Poland tel. +48-85-7388187, fax. +48-85-7457223, jo@uwb.edu.pl

**Anatolijs Zabasta**, senior researcher of Institute of Industrial Electronics and Electrical Engineering, Riga Technical University, Faculty of Power and Electrical Engineering, 12/1 Azenes Str. - 503., Riga, LV 1048, Latvia, tel. +371 67089568, anatolijs.zabashta@rtu.lv

**Anastasija Zhiravecka**, professor, senior researcher of Division of Industrial Electronic equipment, Institute of Industrial Electronics and Electrical Engineering, Riga Technical University, Faculty of Power and Electrical Engineering, 12/1 Azenes Str. - 509., Riga, LV 1048, Latvia, tel. +371 67089917, anastasija.ziravecka@rtu.lv



# Contents

<b>1</b>	<b>ELECTRICITY AND MAGNETISM.....</b>	<b>14</b>
1.1	Basics of electrostatics.....	15
1.2	Magnetism and electromagnetism.....	17
1.2.1	Faraday's Law .....	17
1.3	Magnetic circuits and basic laws .....	20
1.3.1	Magnetic circuit and its basic parameters .....	20
1.3.2	Interaction of electric and magnetic field.....	22
<b>2</b>	<b>METROLOGY AND MATHEMATICAL MODELING FOR INTELLIGENT SYSTEMS .....</b>	<b>25</b>
2.1	Introduction .....	26
2.2	Metrology: part 1 .....	29
2.2.1	Interval .....	29
2.2.2	Range.....	29
2.2.3	Quantity.....	29
2.2.4	Units .....	31
2.2.5	Quantity value .....	34
2.2.6	Nominal property .....	36
2.3	Metrology: part 2 .....	37
2.3.1	Measurand .....	37
2.3.2	Measurement .....	37
2.3.3	Measurement repeatability .....	40
2.3.4	Intermediate measurement precision.....	40
2.3.5	Measurement reproducibility .....	41
2.3.6	Coverage interval, probability and factor.....	42
2.3.7	Calibration.....	43
2.3.8	Metrological traceability .....	43
2.3.9	Verification and validation.....	44
2.3.10	Metrological comparability of measurement results .....	44
2.3.11	Metrological compatibility of measurement results.....	45
2.3.12	Measurement model and measurement function.....	45
2.3.13	Influence quantity.....	46
2.3.14	Correction.....	46

2.3.15	Blank indication .....	46
2.4	Metrology: part 3 .....	47
2.4.1	Measuring instrument.....	47
2.4.2	Material measure .....	47
2.4.3	Sensor .....	48
2.4.4	Detector .....	48
2.4.5	Measuring chain .....	48
2.4.6	Indication.....	48
2.4.7	Rated operating condition .....	50
2.5	Metrology: part 4 .....	51
2.5.1	Sensitivity.....	51
2.5.2	Selectivity of a measuring system.....	51
2.5.3	Resolution.....	51
2.5.4	Discrimination threshold .....	52
2.5.5	Dead band.....	52
2.5.6	Detection limit.....	52
2.5.7	Stability .....	52
2.5.8	Instrumental drift.....	52
2.5.9	Variation due to an influence quantity .....	52
2.5.10	Step response time.....	52
2.5.11	Instrumental uncertainty.....	52
2.5.12	Accuracy class.....	53
2.5.13	Maximum permissible error .....	53
2.5.14	Null measurement uncertainty .....	53
2.6	Metrology: part 5 .....	54
2.6.1	Calibration diagram.....	54
2.6.2	Measurement standard.....	54
2.6.3	Reference material.....	56
2.6.4	Reference data .....	58
2.7	Classification of measurement errors: part 1 .....	59
2.7.1	Measurement error .....	59
2.7.2	Absolute error.....	60
2.7.3	Measurement .....	61
2.7.4	Random error.....	62

2.7.5	Missed measurement .....	62
2.7.6	Relative error.....	62
2.7.7	Error classification .....	63
2.7.8	Zero error.....	64
2.7.9	Distribution error.....	64
2.8	Classification of measurement errors: part 2 .....	67
2.8.1	Evaluation of indirect measurement errors .....	67
2.8.2	Error dispersion .....	67
2.8.3	Measurement graphical interpretation.....	69
2.9	Measurement systems.....	72
2.9.1	Single parameter systems .....	72
2.9.2	Multiple parameter systems .....	72
2.9.3	Parallel systems .....	73
2.10	Measurement of electric quantities.....	74
2.10.1	Base and derived quantities.....	74
2.10.2	Equipment grouping .....	74
2.10.3	Reference gauges.....	75
2.10.4	Measurement transducers .....	79
2.11	Measurement instruments.....	81
2.11.1	Equivalent schematics of measurement instruments.....	81
2.11.2	Measurement instruments interference .....	81
2.11.3	Magneto-electric instrument .....	81
2.11.4	Electronic MI.....	85
2.11.5	Digital instrument.....	89
2.12	Bibliography and list of acronyms.....	93
<b>3</b>	<b>ELECTRICAL ENGINEERING.....</b>	<b>94</b>
3.1	DC circuits.....	95
3.1.1	Elements of DC circuits .....	95
3.1.2	Ohm's Law.....	96
3.1.3	Theorem of the voltage source (Thevenin's theorem) .....	96
3.1.4	Electrical Power .....	97
3.1.5	Complex electrical circuits and Kirchhoff's laws .....	98
3.2	Circuits of single-phase alternating current.....	99
3.2.1	Circuits with resistors.....	99

3.2.2	AC circuits with coil (reactor or inductor) .....	100
3.2.3	AC circuits with capacitors .....	102
3.2.4	AC circuits with coils and resistors .....	103
3.2.5	AC circuits with capacitor and resistor .....	104
3.3	Three-phase AC systems .....	106
3.3.1	Generation of three-phase voltage.....	106
3.3.2	Connection schemes with symmetric load .....	108
3.3.3	Y Connection of asymmetric load.....	109
<b>4</b>	<b>SPECTROSCOPY, MICROSCOPY, MAGNETOMETRY, ELLIPSOMETRY..</b>	<b>110</b>
4.1	Optical spectroscopy of molecules .....	111
4.1.1	Energy levels and types of molecular spectra .....	111
4.1.2	Vibrational spectroscopy of polyatomic molecules .....	122
4.1.3	Luminescent molecular analysis .....	133
4.1.4	References and bibliography .....	152
4.1.5	List of acronyms.....	154
4.2	Principles of optical microscopy .....	156
4.2.1	Introduction .....	156
4.2.2	Optical magnifications and microscopes.....	157
4.2.3	Magnification of a microscope.....	158
4.2.4	Resolution of optical microscope .....	160
4.2.5	Bright –field and darkfield microscopy .....	162
4.2.6	Polarised light microscopy .....	164
4.2.7	Phase contrast microscopy .....	165
4.2.8	Interference microscopy .....	167
4.2.9	Fluorescence microscopy .....	167
4.2.10	Total internal reflection fluorescence microscopy, (TIRFM) .....	170
4.2.11	Confocal microscope.....	172
4.2.12	Fluorescence lifetime imaging microscopy.....	174
4.2.13	Ultramicroscopy and super-resolution microscopy.....	175
4.2.14	References .....	176
4.3	Scanning probe microscopy.....	178
4.3.1	Introduction .....	178
4.3.2	XYZ – scanner .....	178

4.3.3	Sample – probe detection system .....	180
4.3.4	Feedback concept .....	183
4.3.5	AFM probes and AFM imaging .....	184
4.3.6	Magnetic force measurements.....	186
4.3.7	Electric force microscopy.....	189
4.3.8	Scanning near-field optical microscopy .....	191
4.3.9	Apertureless techniques.....	192
4.3.10	References .....	194
4.4	Magnetometry.....	195
4.4.1	Methods of magnetometry studies .....	195
4.4.2	Measuring and interpretation of magnetization curves .....	198
4.4.3	Analysis of temperature dependences of magnetization .....	199
4.4.4	Magnetic anisotropy and angular dependences of magnetic parameters .....	202
4.4.5	References .....	207
4.5	The essence of ellipsometric method, the fundamental ellipsometry equation .....	208
4.5.1	Introduction .....	208
4.5.2	Ellipsometric measurements .....	209
4.5.3	The methods of measurements using the laser ellipsometer .....	211
4.5.4	Methods of measurement using the spectral ellipsometer .....	213
4.5.5	References .....	219
4.6	Mössbauer spectroscopy.....	220
4.6.1	Mössbauer spectroscopy instrumentation .....	220
4.6.2	Mössbauer spectra of nanocrystalline systems and their interpretation .....	223
4.6.3	References .....	226
4.7	Scanning electron microscopy.....	227
4.7.1	Outlook.....	227
4.7.2	Physical basis of scanning electron microscopy .....	227
4.7.3	Analysis of the elemental composition using electron probe.....	230
4.7.4	Auger-electron spectroscopy .....	233
4.7.5	Cathodoluminescence.....	238
4.7.6	Electron beam probing for study of semiconductor materials and devices.....	240
4.7.7	References .....	246
4.8	Diffraction methods of analysis.....	247
4.8.1	Introduction .....	247

4.8.2	Theoretical basis of diffraction analysis.....	247
4.8.3	Determination of the texture of polycrystalline objects .....	251
4.8.4	Broadening of X-ray reflexes .....	251
4.8.5	Diffraction of electrons .....	253
4.8.6	Scheme of transmission electron microscope .....	254
4.8.7	Low energy electron diffraction .....	259
4.8.8	References .....	261
<b>5</b>	<b>EMC, EMI AND RELIABILITY .....</b>	<b>263</b>
5.1	Introduction to EMC.....	264
5.1.1	Basic concepts .....	264
5.1.2	The EMC directives .....	265
5.1.3	Emission.....	269
5.1.4	Immunity .....	272
5.1.5	References .....	273
5.2	Conducted emission measurements.....	274
5.2.1	Introduction .....	274
5.2.2	Conducted emission measurement .....	277
5.2.3	Reducing the emission levels .....	280
5.2.4	References .....	283
5.3	The practical use of a Farady cage .....	285
5.3.1	Introduction and basic principle.....	285
5.3.2	Protection against lightning strikes .....	286
5.3.3	Electrical and magnetic fields .....	287
5.3.4	Shielding using a Faraday cage.....	289
5.3.5	Faraday cage having walls with apertures.....	290
5.3.6	Miscellaneous.....	292
5.3.7	EMC measurements .....	293
5.3.8	Anechoic chamber.....	295
5.3.9	Performing measurements in an anechoic chamber .....	297
5.3.10	Electromagnetic reverberation chamber.....	298
5.3.11	Semi anechoic chamber.....	299
5.3.12	References: .....	301
5.4	Radiated emission.....	302

5.4.1	Introduction .....	302
5.4.2	Electrical and magnetic fields .....	302
5.4.3	Measuring unintentionally emitted electromagnetic waves .....	304
5.4.4	EMC standards .....	307
5.4.5	The use of shielding .....	310
5.4.6	Shielding of cables .....	313
5.4.7	References .....	317
5.5	Reliability and functional safety .....	318
5.5.1	Introduction .....	318
5.5.2	Functional safety .....	318
5.5.3	Reducing the malfunctioning of an electronic device: redundancy .....	319
5.5.4	Reducing the malfunctioning of an electronic device: watchdog timer .....	322
5.5.5	Reducing the malfunctioning of an electronic device: a parity bit .....	323
5.5.6	Reducing the malfunctioning of a device: the Hamming's algorithm .....	324
5.5.7	Highly Accelerated Lifetime Testing .....	329
5.5.8	References .....	333
<b>6</b>	<b>APPLIED SYSTEM THEORY .....</b>	<b>335</b>
6.1	System theory .....	336
6.1.1	Introduction .....	336
6.1.2	The behaviour of a pendulum .....	336
6.1.3	Electrical systems .....	339
6.1.4	The behaviour of a series RLC circuit .....	342
6.1.5	The use of complex numbers .....	344
6.1.6	References .....	346
6.2	Matlab and Simulink excercises .....	347
6.2.1	Matlab exercise 1: pendulum .....	347
6.2.2	Simulink exercise: pendulum .....	347
6.2.3	Matlab exercise 2 : series RLC circuit .....	347
6.2.4	Simulink excercises: series RLC circuit .....	348
6.3	State space analysis .....	350
6.3.1	Example 1: The behaviour of a pendulum .....	350
6.3.2	Example 2: The behaviour of a satellite .....	351
6.3.3	Example 3: servo motor .....	352

6.3.4	State space representation .....	354
6.3.5	Solutions of state space equations .....	357
6.3.6	Exercise .....	359
6.3.7	References .....	361
6.4	System theory: DSP: Analog and digital filters .....	362
6.4.1	Introduction .....	362
6.4.2	Analog low-pass filters .....	362
6.4.3	Digital low-pass filters .....	372
6.4.4	Practical use of the bilinear transformation .....	375
6.4.5	Design of an IIR digital Butterworth filter .....	377
6.4.6	References .....	382
6.5	EMC filters .....	383
6.5.1	Introduction .....	383
6.5.2	The behaviour of passive filters 1 .....	384
6.5.3	The behaviour of passive filters 2 .....	390
6.5.4	References .....	392



## INTRODUCTION

Physics is often considered to be one of the oldest academic disciplines since physics helps people to understand how our universe behaves. Physics is absolutely a very broad academic discipline including astronomy, mechanics, thermodynamics, optics, acoustics, magnetism, electricity, electromagnetism, energy, engineering physics, ...

A decent use of mathematics is very important but when studying physical systems, it is also very important to have decent observations and measurements of the phenomena one needs to study.

The broad realm of physics includes fundamental physical research which supports a large number of applications. By applying physical insights, it is possible to design and develop a large number of new solutions.

The present course on “applied physics” reveals a broad range of applications. First, a physical foundation on electricity and magnetism has been included to support the entire course. Since performing accurate measurements and having a decent insight in measurement errors is mandatory when studying physical phenomena, an extended chapter deals with measurement principles and measurement errors. Especially the measurement of electric quantities has been studied. In a third chapter, electrical (power) systems have been discussed including electrical power measurements.

Decent observations and measurements are needed when studying physics and engineering topics. Due to this reason, an entire chapter has been devoted to the working principles and the use of spectroscopic methods, microscopy, and magnetometry.

When considering electrical engineering and more precisely the subdiscipline EMC (Electromagnetic compatibility), a large number of measurements are needed to verify whether electronic devices satisfy the EMC normalizations. These normalizations aim to ensure reliable operations of electronic devices. The last chapter deals with system theory and its applicability to design electronic filters. Electronic filters have a broad range of applications including EMC filters intended to reduce EMC related problems.

# **1 ELECTRICITY AND MAGNETISM**

## 1.1 Basics of electrostatics

The electric charges interact with each other, thus the same charges repel but the opposite charges attract each other. Each charge produces an electric field around itself therefore the interaction takes place by means of electric field.

Electric field is a material environment with a forced action onto the charged particles or bodies. Electric field is assumed to picture with the help of electric force lines the direction of which correlates with that of the forces acting. The environment is characterised with a parameter called dielectric permeability. An absolute dielectric permeability is

$$\varepsilon_a = \varepsilon_0 \cdot \varepsilon_r , \quad (1.1)$$

where  $\varepsilon_0$  is a constant parameter equal to a dielectric permeability of vacuum ( $8.86 \cdot 10^{-12}$  F/m);  $\varepsilon_r$  – is a relative dielectric permeability indicating the how many times weaker the charges are interacting in a particular environment than in vacuum.

The force of interaction of two point charges is determined with the Coulomb's Law and directed along a straight line connecting these charges (a point charge is a charge the linear size of which are negligibly small in comparison with the distances under consideration). The Coulomb's Law states that the force of interaction of two point charges is directly proportional to the product of these charges, opposite proportional to the square distance between them and depends on the environment the charges are placed in, i.e.

$$F = \frac{Q_1 \cdot Q_2}{4\pi r^2 \cdot \varepsilon_a} \quad (1.2)$$

where F is the force of the charges interaction, N;  $Q_1, Q_2$  are the charges, C (1 C contains  $6.3 \cdot 10^{18}$  of electrons charges, coulomb is an amount of electricity flowing through cross-section area of a conductor per one second at an unchangeable force of current of 1 A, i.e.  $C = A \cdot s$ ); r is a distance between the charges, m;  $\varepsilon_a$  is an absolute dielectric permeability of the environment, F/m.

An electric field impacts an electrically charged body placed into it. Therefore the electric field can perform a work, i.e. it is described with energy. Each point of the electric field can be characterised with an intensity E and potential  $\varphi$ .

Intensity of electric field E is determined with

$$E = F/Q, \quad (1.3)$$

where F is a force of the field action onto the charge placed into this field. The measurement unit of the intensity of the field is V/m.

The reserve of energy of a electricity quantity unit at a particular point of the electric field is potential. This a work equal to potential energy of a 1 C charge at this point, i.e.

$$\varphi = A/Q, \text{ V} \quad (1.4)$$

the measurement unit is volt (V).

Moving a positive charge  $Q$  from one point of the uniform electric field to another point at some distance  $l$  the forces of the field perform a work  $A = F \cdot l$ . This work is called a voltage across these two points

$$U = A/Q. \quad (1.5)$$

If two points of the field have own potentials,  $\varphi_1$  and  $\varphi_2$  the work to move a single charge between these two points is a difference of two charges

$$U_{12} = \varphi_1 - \varphi_2. \quad (1.6)$$

## 1.2 Magnetism and electromagnetism

Any constant magnet has two poles: North (N) and South (S). A magnetic flux exists between the poles and is described by means of magnetic lines in the direction from N to S (fig.1.1). The force of the magnetic flux corresponds to the number of these magnetic lines. The same type of magnetic flux exists around a wire with an electrical current. The direction of the magnetic flux lines in this case can be determined by means of screwdriver rule.

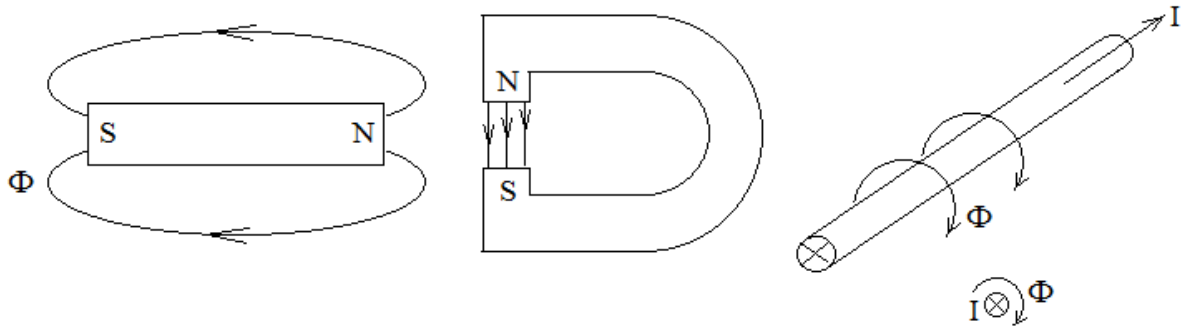


Figure 1.1. Permanent magnet and wire with electric current

Well-known is the fact that like magnetic poles repel but unlike magnetic poles attract each other. Conductors with electric current could be represented as a kind of coil. Then a total magnetic flux is equal to the sum of separate fluxes of all the conductors (fig.1.2) :

$$\Psi = \Phi \cdot w, \quad (1.7)$$

where  $w$  is a number of turns in the winding,  $\Psi$  – value called flux linkage.

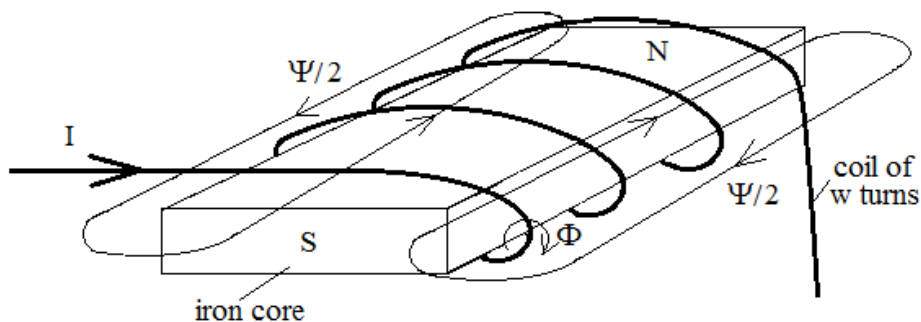


Figure 1.2. Magnetic flux of the coil

### 1.2.1 Faraday's Law

If a constant magnet is moved within a coil with current then the magnetic power lines cross the conductors and an EMF is induced in the coil

$$e = -w \cdot \frac{d\Phi}{dt} . \quad (1.8)$$

Therefore the faster is changing the flux the larger is EMF induced in the coil. The measurement unit for the magnetic flux is Wb (Weber).

Voltage is directed positively from plus to minus, but EMF – in the opposite direction. If both voltage and EMF are in the same direction then

$$u = -e = w \cdot \frac{d\Phi}{dt} . \quad (1.9)$$

Taking this into account Faraday's Law could be rewritten as the following:

$$u = w \cdot \frac{d\Phi \cdot di}{di \cdot dt} , \quad (1.10)$$

where  $w \cdot d\Phi/di$  is the correlation of the flux linkage and current and is marked as inductance

$$L = \frac{w \cdot d\Phi}{di} . \quad (1.11)$$

L is measured in henry (H).

As a result Faraday's Law can be formulated as

$$u = L \cdot \frac{di}{dt} , \quad (1.12)$$

it correlates the voltage of the coil with the speed of changing of its current. If the current in the coil is increasing then the induced electromotive force is of opposite direction to the current and the EMF limits the speed of the current increasing. In the case of current decreasing then the induced EMF is in the same with current direction and the EMF limits the current decreasing.

Taking into account the Faraday's Law the mathematical description of an electrical circuit with a coil and resistor is

$$u = i \cdot R + L \cdot \frac{di}{dt} , \quad (1.13)$$

where i, u are instantaneous values of the current and voltage in the dynamics of their changing. The transient process of current changing in the circuit with coil (Fig.1.3) could be calculated using this expression. Let us consider a connection of coil to a source of constant voltage U. The equation for the transient processes there is the following:

$$u = i \cdot R + L \cdot \frac{di}{dt} . \quad (1.14)$$

This is the first-order equation with the following solution:

$$i(t) = K1 + K2e^{-t/\tau} , \quad (1.15)$$

where K1 is a steady-state value with  $t=\infty$ .

The current through this circuit in steady-state condition will be determined by Ohm's Law:

$$L_{\infty} = K1 = U/R . \quad (1.16)$$

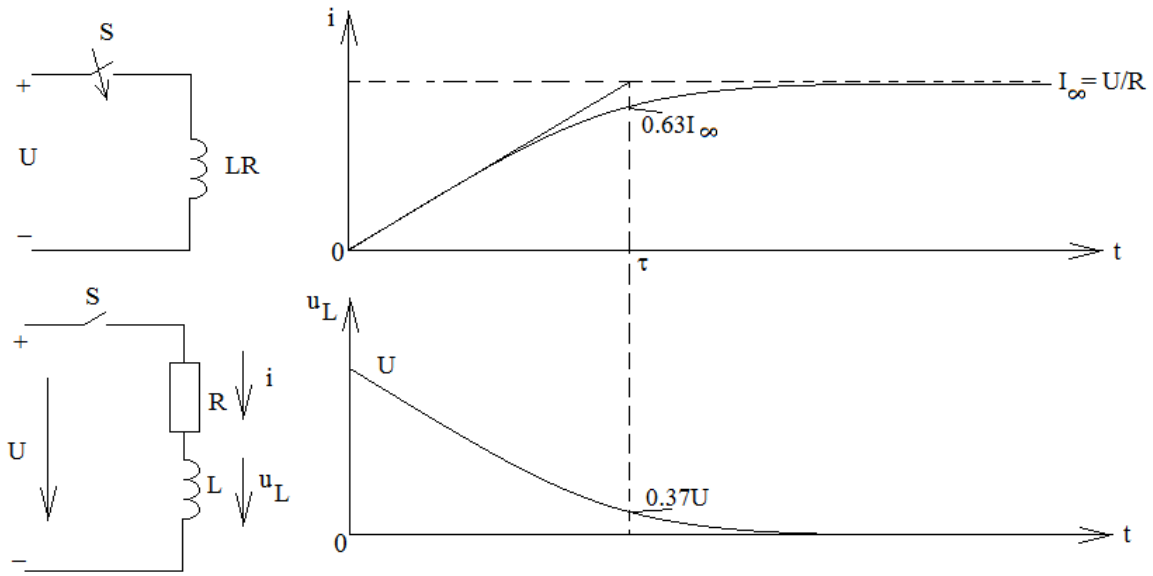


Figure 1.3. Electrical circuit and representation of transient process and current diagrams

The constant value  $K_2$  could be found from the initial condition at  $t=0$  when  $i=0$ :

$$0 = U/R + K_2 \cdot e^0, \quad (1.17)$$

where  $K_2$  is:  $K_2 = -U/R$ .

The result is 
$$i(t) = \frac{U}{R} (1 - e^{-t/\tau}), \quad (1.18)$$

where  $\tau = L/R$  is time constant measured in seconds ( $\tau = \Omega \cdot s / \Omega = s$ ).

Expression (1.18) illustrates that at  $t=0$  the current is equal to zero, but then it increases gradually getting the value up at  $t=\tau$  to the following (Fig.1.3):  $\frac{U}{R} (1 - e^{-t}) = 0.67 \cdot I_{\infty}$ .

The steady-state value is reached at  $t \approx (3 \dots 4)\tau$ .

Thus the conclusion is that the inductivity impedes to the changes of current in the circuit. It is essential to understand that the stored energy in the coil is a product of its flux linkage and current:  $A = 0.5 \cdot \Psi \cdot i = 0.5 \cdot L \cdot i^2$  (J).

In other expression:  $A = \int i \cdot u \cdot dt = \int i \cdot L \cdot \frac{di}{dt} = L \cdot I^2 \cdot 0.5$ .

## 1.3 Magnetic circuits and basic laws

### 1.3.1 Magnetic circuit and its basic parameters

Any magnetic circuit consists of magnetic conductors and a source of magnetic motive force (MMF). One of the best conductors of magnetic flux – steel – is used for magnetic wires (fig.1.4). MMF is created by the winding consisting of a number of turns of conductors ( $w$ ) with current  $I$ :

$$F = I \cdot w. \quad (1.19)$$

As a result of MMF influence the magnetic flux  $\Phi$  is created in the core and it has the following flux density:

$$B = \Phi / s, \quad (1.20)$$

where  $s$  is a cross-section area of the core. As  $\Phi$  is measured in  $V \cdot s$ , but  $S$  in  $m^2$ , the measurement unit for  $B$  is Tesla (T).

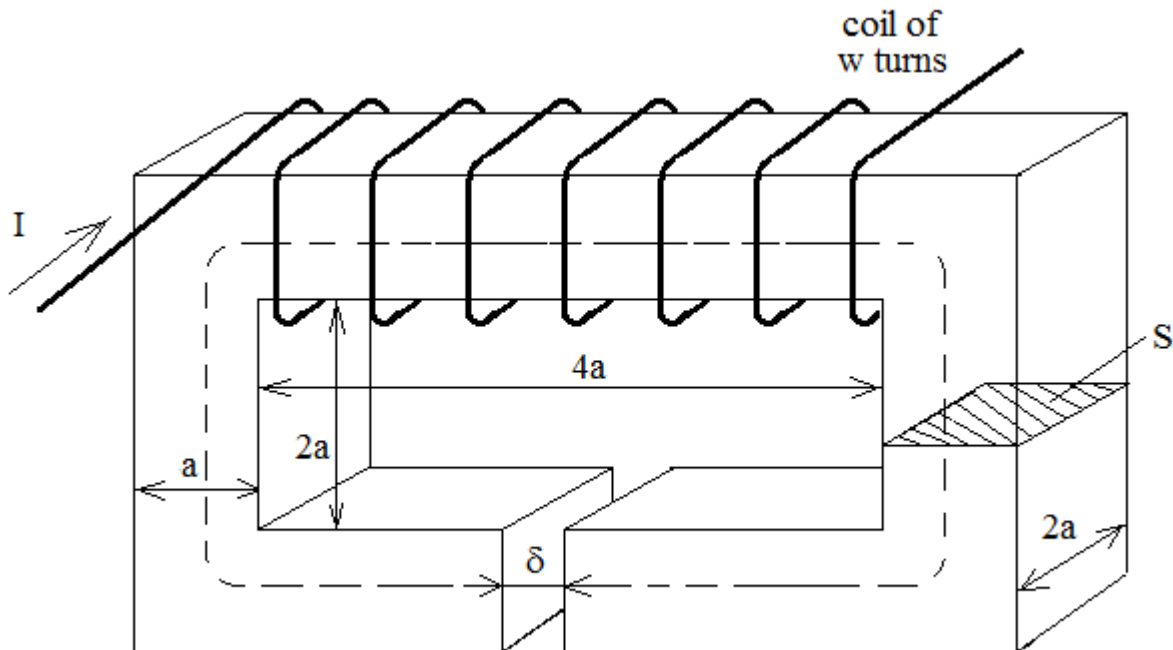


Figure 1.4. Magnetic circuit realisation

Each magnetic conductive material has its own magnetising properties described with magnetisation curve  $B=f(H)$ , which depends on linear magnetising force  $H$  measured as MMF divided by the length  $l$  of the average magnetic line, that is  $A/m$ . The growth of  $B$  is usually started to limit at  $B=1.0T$ , when a saturation process starts (Fig.1.5.).



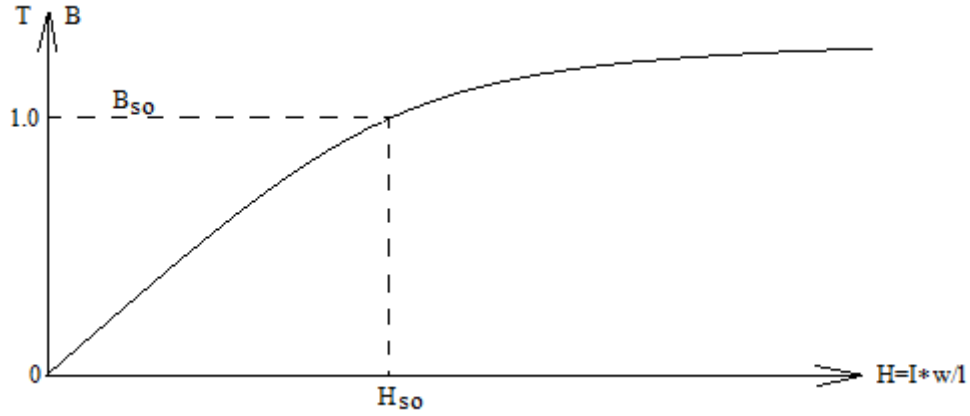


Figure 1.5. Magnetising curve of steel

Logical is to use the linear part of the curve or its part close to the point of saturation for a simulation of a real magnetising system, because less values of magnetising force  $H$  are necessary. For the magnetic system given in fig.1.4 the magnetising flux is determined as

$$\Phi = B(H) \cdot s = B\left(\frac{I \cdot w}{l}\right) \cdot s, \quad (1.21)$$

where flux density  $B$  is a function of  $H = \frac{I \cdot w}{l}$ . If the value of  $H$  is higher than that of saturation  $H_{s0}$  (fig.1.5) then an air gap  $\delta$  should be inside the core of the system (fig.1.4). To overcome this gap a part of general magnetising force is lost

$$H_{\delta} \cdot \delta = \left(\frac{B}{\mu_0}\right) \cdot \delta, \quad (1.22)$$

where  $\delta$  is a length of the air gap,  $B$  – flux density in magnetic conductor,  $\mu_0$  – magnetic permeability of air, which is  $\mu_0 = \frac{B_0}{H_0} = 4\pi \cdot 10^{-7} [\Omega \cdot s/m]$ .

For example, if it is necessary to calculate a magnetic system for a reactor with a constant inductivity  $L$  with the current from zero to  $I_0$  the initial part of the magnetising curve is applied for this purpose and

$$L = \frac{w \cdot B_{s0} \cdot s}{I_0}. \quad (1.23)$$

Therefore if the cross-section area  $s$  is known the required number of turns  $w$  could be calculated and  $H$  can be obtained as  $= I_0 \cdot w / l$ . If this magnetising force is higher than  $H_{s0}$  the difference should be lost in the air gap

$$H_{\delta} = H - H_{s0} = \frac{B}{\mu_0}, \quad (1.24)$$

where from the air gap can be calculated

$$\delta = \frac{I_0 \cdot w \cdot H_{\delta} \cdot 4\pi \cdot 10^{-7}}{H \cdot B_s}. \quad (1.25)$$

If the sizes of the core are defined as in fig.1.4. the cross-section area of the core is  $s = 2 \cdot a^2 = \frac{L \cdot I_0}{w \cdot B_{s0}}$ ,

But an available number of turns of the winding  $w = \frac{8 \cdot a^2 \cdot 0.35}{s_c} = \frac{8 \cdot a^2 \cdot 0.35 \cdot j}{I_R}$ ,

Where  $s_c$  is a cross-section area of the conductor wire,  $j$  is current density (about  $2 \cdot 10^6 \text{ A/m}^2$ ),  $I_R$  – the rated current of the reactor usually lower than  $I_0$ .

### 1.3.2 Interaction of electric and magnetic field

As magnetic flux exists around a current carrying conductor the flux will interact with external magnetic system. A constant magnet (see fig.1.6) creates magnetic flux with density

$$B_1 = \Phi_1 / a \cdot l . \quad (1.26)$$

If the conductor creating its own magnetic flux  $\Phi_2$  is normal to the force lines then according to fig.1.6 the repulsive force exists between the two fluxes (both of them are unidirectional)

$$F = B_1 \cdot I \cdot l, \quad (1.27)$$

where  $l$  is a conductor active length being in accordance with the length of the pole of magnet. The direction of this force is determined with the help of left hand rule which states that if the permanent flux is going into opened palm of the left hand but the fingers are directed as the current in the conductor then the opened thumb points the direction of the movement of the conductor. This phenomenon is the basic principle of electric motors operation.

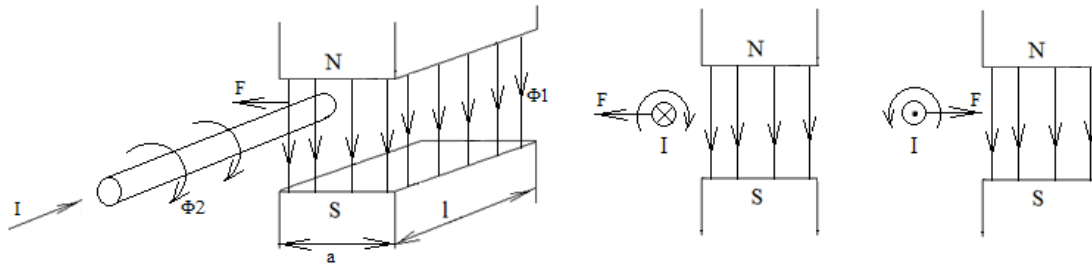


Figure 1.6. Interaction of magnetic flux and current carrying conductor

If the conductor crosses the magnetic force lines then in accordance with Faraday's Law an EMF is induced in the conductor. In fig.1.7 the constant magnet induces magnetic flux  $\Phi_1$  with density  $B_1$ , and the conductor moves normally to its force lines. The following EMF is induced there

$$E = B_1 \cdot l \cdot v , \quad (1.28)$$

where  $v$  is speed of crossing of the force lines  $\Phi_1$  measured in m/s. This happens according to the right hand rule that states if permanent magnetic flux is going to the opened palm of the right hand but the thumb points the direction of the conductor movement in the flux then the

fingers point the direction of the current in the conductor. It describes the basic principle of the AC generator operation.

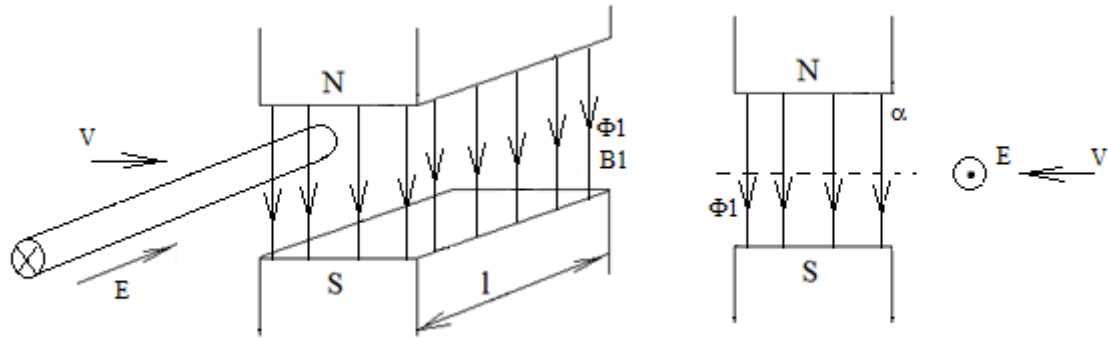


Figure 1.7. Generation of EMF

The generated EMF depends on the angle between the magnetic force lines and the conductor crossing direction; this is angle  $\alpha$  between vectors  $v$  and  $\Phi_1$  (Fig.1.7.) and

$$E = B_1 \cdot l \cdot v \cdot \sin \alpha. \quad (1.29)$$

For the practical realisation of electric generator the conductor rotates in the constant magnetic flux (Fig.1.8), and an EMF of variable both direction and value is induced in the conductor. The highest value of EMF is induced when the conductor is in the upright position:

$$E_m = 2 \cdot B_1 \cdot l \cdot v \cdot w, \quad (1.30)$$

where the conductor is of a frame form and has two edges and could have a lot of turns  $w$ .

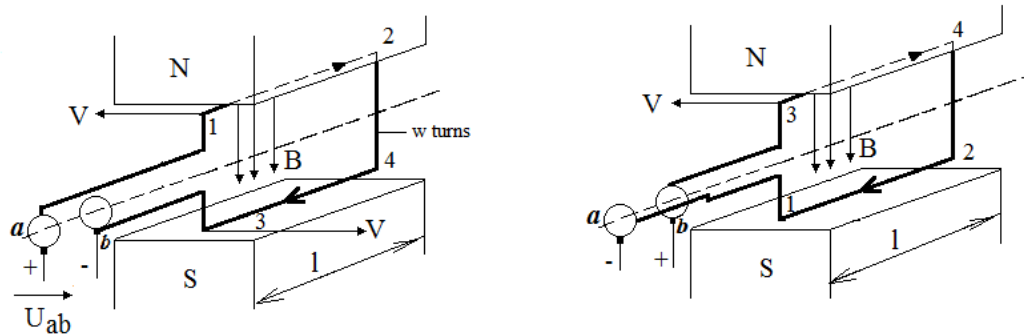


Figure 1.8. Realisation of generator

The EMF is changing periodically depending on angle  $\alpha$ , an instantaneous value is represented by means of sine-form curve

$$e = E_m \cdot \sin \alpha. \quad (1.31)$$

The frame rotates with an angular velocity  $\omega$ . Period of its full turn  $T$  corresponds to angle  $2\pi$ , thus  $2\pi = \omega \cdot T$ ,

where from

$$\omega = 2\pi/T = 2\pi \cdot f, \quad (1.32)$$

where  $f$  is a frequency of the induced EMF. For the assumed standard frequency  $f=50\text{Hz}$  the angular velocity is  $\omega=314\text{ rad/s}$ .

## **2 METROLOGY AND MATHEMATICAL MODELING FOR INTELLIGENT SYSTEMS**

## 2.1 Introduction

Our modern era cannot be imagined without all those machines, computers, discoveries. They all measure, intensively use mathematic calculations, perform activities. Behind all of this bright scene we have various reliable partners – Metrology, science about measuring, and Mathematical Modelling – a science of how to describe processes with mathematical expressions. You might feel them as quite new, however, in fact, metrology, mathematical modelling, intelligent systems have history since beginning of the Mankind.

In ancient times humans were measuring various physical parameters, described processes mathematically, and using their own intelligence capacities, brought us to today's achievements.

But wait! What is intelligence?

Let me declare that intelligence is ability to obtain information, analyse it and react in appropriate way.

So this book is about to how to transfer information from everything, that surrounds us, how to understand, what we have got, what to do with it, how to ensure, that our plans are the best match.

To achieve our goals, we need instruments. First we need a sensor. A device that transforms a physical parameter in to desired format, so we can process it. Even a simplest ruler is a sensor, as it transforms physical dimensions of an object in to values we can interpret and understand. For example, a brick. The same people uses around the world to build their houses. It must have certain dimensions to fit in to building. So man invented a ruler, a sensor. By the way, ruler is a perfect sensor, as it is not altering object it measures. Other sensors do.

Second, we must be able to analyse data, we have measured. Sounds logical. A person, who does measurement, must be able to assess, is this information correct. Is it reliable enough to be used. This is one of most important questions in science. History has many cases, when scientific work or discovery is raised or doomed just because of measurements.

Third thing is the way we use information we got. We should thank our predecessors for discovery of Mathematics, as this is one and only way how to tell others about physical processes. So we can put information, exact values, in to mathematical expressions, formulas, and obtain desired data.

Forth “instrument” now is a decision made on everything we have got. Quite often people use incorrect form of this step. They say: “We do not have enough data to make a decision.” This is wrong. The correct description of situation is: “Based on data we have, we decided to do nothing.” In fact, a decision – do nothing, is completely acceptable and does not mean any harm. Imagine a pilot of a jumbo jet. Flight is normal, everything works as expected, no need for adjustment, so decision – no action. Opposite sample, reactor in nuclear power plant start to generate too much free neutrons. Action should be immediate to slow down process or core will melt and... So decision is important.

Here I decide to advance to fifth important part, a cause, why we have all this stuff. We want to alter, to impact, to change an object. As object here I mean almost everything we have. Whatever we do, build a house, read a book, drive a car, have a sunbath, this is a kind of change.

It is not a secret for everybody, that mankind is lazy. This fact is an engine for inventions. We want to get more for doing less. So it was long time ago, and so it is today. Even Archimedes invented His famous law while relaxing in the bath. Today it is the same, but different. We advanced so far, that some of routine tasks are delegated to machines. Sounds good, but here is also a problem – machine can cover a definite set of duties. Human instead, does not have such limitations. The border here is ability to make a spontaneous decision. Human can, machine can not. It would be nice, if we can delegate some decision making to machines, while we are relaxing in the bath. Now we can and this field have its own name – intelligent systems. Systems, that are able to cognize the target. If follow scientific oriented literature, you can see a constellation of words like, cognitive computing, neural networks, artificial intelligence, intelligent systems and so on. All of them, actually, are just bricks of our today. So it is important to be familiar with them. Field actually is so vast that one book like this can cover only a tiny, but, in fact, extremely important part. Acquisition and processing of data keeping in mind a reason, why it is performed – creation of intelligent system.

Let's declare, what defines an intelligent system. Now we know, that it is capable to cognize an object it is applied to. Transform its physical parameters in to process able data. It is capable to sort out and use only valid data. It has certain mathematical instrument inside, that gives us a solid base for decision making. And finally, there is a way, how object can be impressed. So in brief - Intelligent system is a repetitive process, a cycle, where exist 5 steps:

- Step 1. Perform a measurement
- Step 2. Assess acquired data
- Step 3. Calculate desired values
- Step 4. Make a decision
- Step 5. Perform an impression to environment

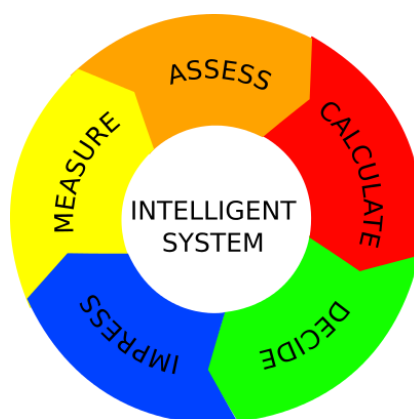


Figure 2.1. Formal cycle of an Intelligent System

What association comes on a pop, when you hear or read this name – Intelligent System? Something very fast operating, powerful and so on. In fact, time for one cycle is not important. Depending on target, time frame varies. A lot. It might take less than a millisecond or more

than a week for one full cycle. Most important thing here is to create a proper design. And this is a thing this book is about. To create a proper design of any kind of intelligent system, you must have absolutely clear understandings about what to measure. How You will measure. What are acceptable measurement discrepancies and how to handle them. You will find answers about how to reliably transfer information from everything, that surrounds us, to assessment mechanism to ensure validity of it. All of mentioned before is crucial for calculations. Mathematics is straight forward. If You enter correct information in to model, you have good result. Good result leads to correct decision. Correct decision makes our life better. You might wonder – and how about error correction? Story is almost the same. How errors are recognized? Based on exact and verified models. So this is exactly about mathematical modelling. A reference model allows to determine, is measurement correct, are values within acceptable bounds, can decision be made on this.

So what about impression. This is also a vast area for development and inventions. If we have a tiny system, all impressions can be done with one actuator, may be a switch or something so simple. Should we have a large scale system, or even a set of systems, decisions and following impressions in one place, can create changes or fluctuations in other. There is a neat geometrical interpretation, or model, of complex IS structure where we have 12 systems in conjunction – dodecahedron.

Let this picture impress your understanding about what to take in account when designing measurements.

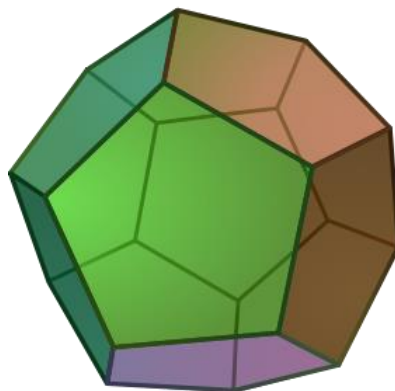


Figure 2.2. Dodecahedron as Schematic Interpretation of Complex IS Structure

I know that some of you will attempt to insert IS cycle directly in to this polyhedron and realize, that edges cannot be associated with the same process for both IS. Vertices even more. This is a good point, as You are starting to model. Don't be confused. Just take in account that in complex environment whole process is (must be) stripped down to smaller components and they are interacting in between at different levels. People, who are involved in programming know this approach as Agile Methodology. However, it is very often covered by some reason poorly written software, You should not be afraid of it. When applied wisely, this allow to reach desired goal and not get trapped in to mayhem. So to be able to design a well performing Intelligent System, you should know about Sensors, Measurements, Mathematical modelling, Decision Making, Actuation. All of this you will have here. At least for the great beginning.



## 2.2 Metrology: part 1

Metrology is scientific study of measurement. For people, who just start to study natural sciences usually this is a moment of big wonder. What can be easier than to measure something? Use a tool, read result and that's it. It's hard to understand in the first approach, that result we got NEVER !!! is absolutely correct. It is correct within some range. We all know, that physics, mathematics, chemistry and so on, are exact sciences where exact data is used. It is true. So our task now is to put together those two things, exactness of science and inexactness of measurements. So there are two approaches in general – most common and easy to understand – Classical Approach, and second – Uncertainty Approach. Beware of a fact, that few terms used in both approaches designate two different concepts. In such cases two different definitions are needed.

Before we advance, it is mandatory to agree on terminology and designations. Terminology section is covered by International Standards ISO 704, ISO 1087-1, and ISO 10241 [1]. Terminology might contain synonyms that are respectively more or less widespread in different countries, however if list is provided, first one is preferred.

Terminology used in metrology:

### 2.2.1 Interval

The term "interval" is used together with the symbol  $[a,b]$  to denote the set of real numbers  $x$  for which  $a \leq x \leq b$ , where  $a$  and  $b > a$  are real numbers. The term "interval" is used here for 'closed interval'. The symbols  $a$  and  $b$  denote the 'end-points' of the interval  $[a,b]$ .

Example

$[-1, 5]$

The two end-points 2 and  $-4$  of the interval  $[-1, 5]$  can be stated as  $2 \pm 3$ . This expression does not denote the interval  $[-1, 5]$ .

### 2.2.2 Range

Term Range is used as range of interval  $[a,b]$  and mean difference between endpoints. It is denoted by  $r[a,b]$ . Term Span is occasionally used instead.

Example :

$$r[-3,7]=7-(-3)=10$$

### 2.2.3 Quantity

Property of a phenomenon, body, or substance, to which a number can be assigned with respect to a reference. The term "Quantity" is often used for kind of quantity where reference can be a measurement unit, a measurement procedure, or a reference material. Symbols for quantities are given in the International Standard ISO/IEC 80000, Quantities and units. The symbols for quantities are written in italics. A quantity as defined here is a scalar. However, a vector or a tensor whose components are quantities is also considered to be a quantity.

Example:

### *Kind of quantity*

Aspect common to mutually comparable quantities. The division of the concept of ‘quantity’ according to ‘kind of quantity’ is to some extent arbitrary. For example - diameter, circumference, and wavelength, are generally considered to be quantities of the same kind, namely, of the kind of quantity called length, or heat, kinetic energy, and potential energy, are generally considered to be quantities of the same kind, namely, of the kind of quantity called energy. Quantities of the same kind within a given system of quantities have the same quantity dimension. However, quantities of the same dimension are not necessarily of the same kind.

### *System of quantities*

This is a set of quantities together with a set of non-contradictory equations relating those quantities. Ordinal quantities are usually not considered to be part of a system of quantities because they are related to other quantities through empirical relations only.

### *Base quantity*

Quantity in a conventionally chosen subset of a given system of quantities, where no subset quantity can be expressed in terms of the others. Base quantities are referred to as being mutually independent since a base quantity cannot be expressed as a product of powers of the other base quantities.

### *Derived quantity*

Quantity, in a system of quantities, defined in terms of its base quantities. It means for example, that in a system of quantities having the base quantities length and mass, density is a derived quantity defined as the quotient of mass and volume (length to the third power).

### *International System of Quantities*

ISQ is based on the seven base quantities length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity. System of quantities is published in the International Standard ISO/IEC 80000, Quantities and units. The International System of Units (SI), is based on the ISQ.

### *Quantity dimension*

Expression of the dependence of a quantity on the base quantities of a system of

quantities as a product of powers of factors corresponding to the base quantities, omitting any numerical factor. A power of a factor is the factor raised to an exponent. Each factor is the dimension of a base quantity. The conventional symbolic representation of the dimension of a derived quantity is the product of powers of the dimensions of the base quantities according to the definition of the derived quantity. In deriving the dimension of a quantity, no account is taken of its scalar, vector or tensor character.

In a given system of quantities,

- quantities of the same kind have the same dimension,
- quantities of different dimensions are always of different kinds, and

- quantities having the same dimension are not necessarily of the same kind.

In the International System of Quantities (ISQ), the dimensions of the base quantities are:

Base quantity	Dimension
length	L
mass	M
time	T
electric current	I
thermodynamic temperature	$\Theta$
amount of substance	N
luminous intensity	J

Thus, the dimension of a quantity  $Q$  is  $\dim Q = L^\alpha M^\beta T^\gamma I^\delta \Theta^\varepsilon N^\zeta J^\eta$  where the exponents, named dimensional exponents, are positive, negative, or zero.

#### *Quantity of dimension one - dimensionless quantity*

Quantity for which all the exponents of the factors corresponding to the base quantities in its quantity dimension are zero. The term “dimensionless quantity” is commonly used for historical reasons. It stems from the fact that all exponents are zero in the symbolic representation of the dimension for such quantities.

The term “quantity of dimension one” reflects the convention according to ISO 31-0:1992, sub clause 2.2.6 in which the symbolic representation of the dimension for such quantities is the symbol 1. The measurement units and values of quantities of dimension one are numbers, but such quantities convey more information than a number. For example, plane angle, solid angle, refractive index, relative permeability, mass fraction, friction factor, Mach number etc.

Quantities of dimension one can also be numbers of entities like number of turns in a coil, number of molecules in a given sample, degeneracy (number of energy levels) in quantum mechanics and other.

#### 2.2.4 Units

Entity used to have a common reference to express a quantity of.

##### *Measurement unit*

Scalar quantity, defined and adopted by convention, with which any other quantity of the same kind can be compared to express the ratio of the two quantities as a number. Measurement units are designated by conventionally assigned names and symbols. Measurement units of quantities of the same dimension may be designated by the same name and symbol even when the quantities are not of the same kind. For example, kilometre per hour and km/h are respectively the name and symbol of both a measurement unit of distance and a unit time, which are

generally not considered to be quantities of the same kind. However, in some cases special measurement unit names are restricted to be used with quantities of specific kind only. For example, the unit  $1/s$  is called hertz when used for frequencies and becquerel when used for activities of radionuclides.

Measurement units of quantities of dimension one are numbers. In some cases, these units are given special names, e.g. radian, steradian, and decibel, or are expressed by quotients such as millmole per mole equal to  $10^{-3}$  and microgram per kilogram equal to  $10^{-9}$ . For a given quantity, the short term “unit” is often combined with the quantity name, such as “mass unit”.

#### *Base unit*

Measurement unit that is adopted by convention for a base quantity. In each system of units there is only one base unit for each base quantity. In the SI, the metre is the base unit of length. The centimetre and the kilometre are also units of length, but they are not base units in the SI. A base unit may also serve for a derived quantity of the same dimension. Rainfall, when defined as areic volume (volume per area), has the metre as a coherent derived unit in the SI. For number of entities, the number one, symbol 1, can be regarded as a base unit in any system of units.

#### *Derived unit*

Measurement unit for a derived quantity. For example, the metre per second, symbol  $m/s$ , and the centimetre per second, symbol  $cm/s$ , are derived units of speed in the SI. The kilometre per hour, symbol  $km/h$ , is a unit of speed outside the SI but accepted for use with the SI. The knot, equal to one nautical mile per hour, is a unit of speed outside the SI.

#### *Coherent derived unit*

Derived unit that, for a given system of quantities and for a chosen set of base units, is a product of powers of base units with no other proportionality factor than one. Coherence can be determined only with respect to a particular system of quantities and a given set of base units. If the metre, the second, and the mole are base units, the metre per second is the coherent derived unit of velocity when velocity is defined by the quantity equation  $v = ds/dt$ , and the mole per cubic metre is the coherent derived unit of amount-of-substance concentration when amount-of-substance concentration is defined by the quantity equation  $c = n/V$ . The kilometre per hour and the knot, given as examples of derived units are not coherent derived units in such a system. The coherent derived unit for every derived quantity of dimension one in a given system of units is the number one, symbol 1. Name and symbol of the measurement unit one are generally not indicated.

#### *System of units*

Set of base units and derived units, their multiples and submultiples, and rules for their use.

#### *Coherent system of units*

System of units, based on a given system of quantities, in which the measurement unit for each derived quantity is a coherent derived unit. A system of units can be coherent only with respect to a system of quantities and the adopted base units. For a coherent system of units, numerical value equations have the same form as the corresponding quantity equations.

### *Off-system measurement unit*

Measurement unit that does not belong to a given system of units

Example:

- The electronvolt (about  $1.602 \times 10^{-19}$  J) is an off-system measurement unit of energy with respect to the SI.
- Day, hour, minute are off-system measurement units of time with respect to the SI.

### *International System of Units - SI*

Coherent system of units based on the International System of Quantities, their names and symbols, and a series of prefixes and their names and symbols, together with rules for their use, adopted by the General Conference on Weights and Measures (CGPM). The SI is founded on the seven base quantities of the ISQ and the base units contained in the following table.

Base quantity	Base unit	
Name	Name	Symbol
Length	metre	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

In quantity calculus, the quantity ‘number of entities’ is often considered to be a base quantity, with the base unit one, symbol 1. For a full description and explanation of the International System of Units, see the current edition of the SI brochure published by the Bureau International des Poids et Mesures.

### *Multiple of a unit*

Measurement unit obtained by multiplying a given measurement unit by an integer greater than one like the kilometre is a decimal multiple of the metre, and the hour is a non-decimal multiple of the second.

### *Submultiple of a unit*

Measurement unit obtained by dividing a given measurement unit by an integer greater than one like the millimetre is a decimal submultiple of the metre, and for plane angle, the second is a non-decimal submultiple of the minute respectively.

### *Unit equation*

Unit equation is mathematical relationship relating base units, coherent derived units or other measurement units like:

- $[Q1] = [Q2] [Q3]$  where  $[Q1]$ ,  $[Q2]$ , and  $[Q3]$  denote the measurement units of  $Q1$ ,  $Q2$ , and  $Q3$ , respectively, provided that these measurement units are in a coherent system of units.
- $J = \text{kg} \cdot \text{m}^2 / \text{s}^2$ , where  $J$ ,  $\text{kg}$ ,  $\text{m}$ , and  $\text{s}$  are the symbols for the joule, kilogram, metre and second, respectively.
- $1 \text{ km/h} = (1/3.6) \text{ m/s}$ .

### *Conversion factor between units*

Conversion factor between units ratio of two measurement units for quantities of the same kind like:

- $\text{km/m} = 1000$  and thus  $1 \text{ km} = 1000 \text{ m}$

The measurement units may belong to different systems of units.

- $\text{h/s} = 3600$  and thus  $1 \text{ h} = 3600 \text{ s}$ ;
- $(\text{km/h})/(\text{m/s}) = (1/3.6)$  and thus  $1 \text{ km/h} = (1/3.6) \text{ m/s}$ .

### 2.2.5 Quantity value

Number and reference together expressing magnitude of a quantity.

A quantity value either is:

- a product of a number and a measurement unit (the unit one is generally not indicated for quantities of dimension one), or
- a number and a reference to a measurement procedure, or
- a number and a reference material.

The number can be real or complex. A quantity value can be presented in more than one way. In the case of vector or tensor quantities, each component has a value as defined above.

### *Numerical quantity value*

Numerical value of a quantity is number in the expression of a quantity value, other than any number serving as the reference. For quantities of dimension one, the reference is a measurement unit which is a number and this is not considered as a part of the numerical quantity value.

For example, in an amount-of-substance fraction equal to  $5 \text{ g/kg}$ , the numerical value is 5 and the unit is  $\text{g/kg}$ . The unit  $\text{g/kg}$  is numerically equal to 0.001, but this number 0.001 is not part of the numerical quantity value that remains 5.

For quantities that have a measurement unit, the numerical value  $\{Q\}$  of a quantity  $Q$  is frequently denoted  $\{Q\} = Q/[Q]$ , where  $[Q]$  denotes the measurement unit like in the case for a mass value of  $5 \text{ kg}$ , the numerical value in kilograms is  $\{m\} = (5 \text{ kg})/\text{kg} = 5$ .

### *Quantity calculus*

Quantity calculus is set of mathematical rules and operations applied to quantities other than ordinal quantities. In quantity calculus, quantity equations are rather preferred to numerical value equations because quantity equations are independent of the choice of measurement units, whereas numerical value equations are not (see ISO 31-0:1992, subclause 2.2.2).

#### *Quantity equation*

Quantity equation is mathematical relationship between quantities in a given system of quantities, independent of measurement units. Generally, quantity equation is the same as formulas used in physics. For example:

- $Q_1 = \zeta Q_2 Q_3$  where  $Q_1$ ,  $Q_2$ , and  $Q_3$  denote different quantities, and where  $\zeta$  is a numerical factor.
- $T = (1/2) mv^2$ , where  $T$  is the kinetic energy and  $v$  the speed of a specified particle of mass  $m$ .
- $n = I \cdot t / F$  where  $n$  is the amount of substance of a univalent component,  $I$  is the electric current and  $t$  the duration of the electrolysis, and where  $F$  is the Faraday constant.

#### *Numerical quantity value equation*

Numerical quantity value equation is mathematical relationship relating numerical quantity values, based on a given quantity equation and specified measurement units.

- In the quantity equation for kinetic energy of a object,  $T = (1/2) mv^2$ , if  $m = 5$  kg and  $v = 4$  m/s, then  $\{T\} = (1/2) \times 5 \times 4^2$  is a numerical value equation giving the numerical value 40 of  $T$  in joules.

#### *Ordinal quantity*

Quantity, defined by a conventional measurement procedure, for which a total ordering relation, according to magnitude, with other quantities of the same kind is defined, but for which no algebraic operations among those quantities are defined. Ordinal quantities can enter into empirical relations only and have neither measurement units nor quantity dimension. Ordinal quantities are arranged according to ordinal quantity scales. For example:

- Wind speed on the Beaufort scale
- Octane number for petroleum fuel
- Earthquake strength on the Richter scale

#### *Quantity scale*

Ordered set of values of quantities of a given kind used in ranking, according to magnitude, quantities of the same kind like:

- Celsius temperature scale
- Time scale
- Rockwell C hardness scale

#### *Ordinal quantity scale*

Also known as conventional reference scale. Quantity scale, defined by formal agreement, on which only comparison of magnitude applies and is used exclusively for ordinal quantities. An

ordinal quantity scale may be established by measurements according to a measurement procedure.

#### *True quantity value*

Quantity value consistent with the definition of a quantity. In the classical approach to describing measurement, a true quantity value is considered unique and, in practice, unknowable. Another approach is to recognize that, owing to the inherently incomplete amount of detail in the definition of a quantity, there is not a single true quantity value but rather a set of true quantity values consistent with the definition. Other approaches dispense altogether with the concept of true quantity value and rely on the concept of metrological compatibility of measurement results for assessing their validity. In the special case of a fundamental constant, the quantity is considered to have a single true quantity value.

#### *Conventional quantity value*

Quantity value attributed by agreement to a quantity for a given purpose. The term “conventional true quantity value” is sometimes used for this concept, but its use is discouraged. A conventional quantity value is only an estimate of a true quantity value. For example standard acceleration of free fall (formerly called ‘standard acceleration due to gravity’) is  $g_n = 9.806\,65\text{ m/s}^2$ .

#### 2.2.6 Nominal property

Property of a phenomenon, body, or substance, that can be identical or not to a comparable property, but cannot be ordered with it according to magnitude.

- Sex of a human being
- Colour of a paint sample
- Colour of a spot test in chemistry
- ISO two-letter country code
- Sequence of amino acids in a polypeptide



## 2.3 Metrology: part 2

Field of knowledge concerned with measurement. Metrology includes all theoretical and practical aspects of measurement, whatever the measurement uncertainty and field of application.

### 2.3.1 Measurand

Quantity intended to be measured. The specification of a measurand requires description of the state of the phenomenon, body, or substance carrying the quantity, including any relevant component and the chemical entities involved. The measurement might change the phenomenon, body, or substance such that the quantity being measured may differ from the measured. In this case adequate correction is necessary. For example, the potential difference between the terminals of a battery may decrease when using a voltmeter with a significant internal conductance to perform the measurement. The open-circuit potential difference can be calculated from the internal resistances of the battery and the voltmeter.

In chemistry, “analyte”, or the name of a substance or compound, are terms sometimes used for ‘measurand’. This usage is erroneous because these terms do not refer to quantities. In chemistry, the measurand can be a biological activity.

### 2.3.2 Measurement

Process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity. Measurement implies comparison of quantities or counting of entities. Measurement presupposes description of the quantity commensurate with the intended use of the measurement result, a measurement procedure, and a calibrated measuring system operating according to a specified measurement procedure.

#### *Measurement principle*

Phenomenon serving as the basis of a measurement. The phenomenon can be of a physical, chemical, or biological nature, like electromagnetic effect applied to the measurement of current, light absorption applied to the measurement of concentration, change of heart rate during exercise.

#### *Measurement method*

Generic description of a logical organization of operations used in a measurement. Measurement methods may be qualified in various ways such as:

- substitution measurement method;
- differential measurement method;
- null measurement method;
- direct measurement method;
- indirect measurement method.

#### *Measurement procedure*

Detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation

to obtain a measurement result. A measurement procedure is usually documented in sufficient detail to enable an operator to perform a measurement. A measurement procedure can include a target measurement uncertainty. A measurement procedure is sometimes called a standard operating procedure, abbreviated SOP.

#### *Reference measurement procedure*

Measurement procedure accepted as providing measurement results fit for their intended use in assessing measurement trueness of measured quantity values obtained from other measurement procedures for quantities of the same kind, or in characterizing reference materials.

#### *Primary measurement procedure*

Measurement procedure used to obtain a measurement result without relation to a measurement standard for a quantity of the same kind.

The Consultative Committee for Amount of Substance – Metrology in Chemistry (CCQM) uses the term “direct primary method of measurement” for this concept. Definitions of two subordinate concepts, which could be termed “direct primary reference measurement procedure” and “ratio primary reference measurement procedure”, are given by the CCQM.

#### *Measurement result*

Set of quantity values being attributed to a measurand together with any other available relevant information. A measurement generally provides information about the set of quantity values, such that some may be more representative of the measurand than others. This may be demonstrated in the form of a probability density function (PDF). A measurement result is generally expressed as a single measured quantity value and a measurement uncertainty. If the measurement uncertainty is considered to be negligible for some purpose, the measurement result may be expressed as a single measured quantity value. In many fields this is the common way of expressing a measurement result.

In the traditional literature measurement result was defined as a value attributed to a measurand and explained to mean an indication, or an uncorrected result, or a corrected result, according to the context.

#### *Measurement accuracy*

In classical approach measurement accuracy is closeness of agreement between a measured quantity value and a true quantity value of the measurand. The concept ‘measurement accuracy’ is not given a numerical value, but a measurement is said to be more accurate when it offers a smaller measurement uncertainty. Measures of measurement accuracy are found in ISO 5725.

The term “measurement accuracy” should not be used for “measurement trueness” and the term “measurement precision” should not be used for “measurement accuracy”.

In uncertainty approach measurement accuracy is closeness of agreement between measured quantity values that are being attributed to the measurand.

### *Measurement trueness*

Trueness of measurement mean closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value. A reference quantity value can be a true quantity value of the measurand or an assigned quantity value of a measurement standard with negligible measurement uncertainty. Measurement trueness cannot be expressed numerically, but measures are given in ISO 5725. Measurement trueness is inversely related to only systematic measurement error. The term “measurement trueness” should not be used for description of measurement accuracy and vice versa.

### *Measurement precision*

Measurement precision is closeness of agreement between indications obtained by replicate measurements on the same or similar objects under specified conditions. Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.

The ‘specified conditions’ can be repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement. Measurement precision is used to define measurement repeatability, intermediate measurement precision, and measurement reproducibility. Sometimes “precision” is erroneously used to mean measurement accuracy.

### *Measurement error*

Error of measurement is difference of measured quantity value and reference quantity value. The error concept can be used when there is a single reference quantity value to refer to, which occurs if a calibration is made by means of a measurement standard of negligible measurement uncertainty or if a conventional quantity value is given, or if the measurand is supposed to be represented by a unique true quantity value or a set of true quantity values of negligible range. The sign of the difference must be noted.

Measurement error should not be confused with production error or mistake.

### *Systematic measurement error*

Systematic error is a component of measurement error that in replicate measurements remains constant or varies in a predictable manner. The reference quantity value for a systematic measurement error is a true quantity value, or a measured quantity value of a measurement standard of negligible measurement uncertainty, or a conventional quantity value.

Systematic measurement error, and its causes, can be known or unknown. A correction can be applied to compensate for a known systematic measurement error. Systematic measurement error equals the difference of measurement error and random measurement error.

### *Measurement bias*

A systematic measurement error on its estimate, with respect to a reference quantity value.

### *Random measurement error*

Random error is a component of measurement error that in replicate measurements varies in an unpredictable manner. The reference quantity value for a random measurement error is the average that would ensue from an infinite number of replicate measurements of the same measurand. Random measurement errors of a set of replicated measurements form a distribution that can be described by its variance, and has an expectation of zero. Random error equals the difference of error of measurement and systematic error of measurement.

### *Measured quantity value*

Measured value is quantity value representing a measurement result. For a measurement involving replicate indications, each indication can be used to provide a corresponding measured quantity value. This set of individual measured quantity values can be used to calculate a resulting measured quantity value, usually with a decreased associated measurement uncertainty.

When the range of the true quantity values believed to represent the measurand is small compared with the measurement uncertainty, a measured quantity value can be considered to be the best estimate of an essentially unique true quantity value and is often an average or median of individual measured quantity values obtained through replicate measurements. In the case where the range of the true quantity values believed to represent the measurand is not small compared with the measurement uncertainty, a measured value is often an estimate of an average or median of the set of true quantity values. Terms “result of measurement” and “estimate of the value of the measurand” or just “estimate of the measurand” are used for “measured quantity value”.

### 2.3.3 Measurement repeatability

Measurement precision under a set of repeatability conditions of measurement.

#### *Repeatability condition of measurement*

Repeatability condition of measurement in a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time. In chemistry, the term ‘intra-serial precision condition of measurement’ is sometimes used to designate this concept.

### 2.3.4 Intermediate measurement precision

Measurement precision under a set of intermediate precision conditions of measurement.

#### *Intermediate precision condition of measurement*

This is a condition of measurement in a set of conditions that includes the same measurement procedure, same location, and replicate measurements on the same or similar objects over an extended period of time, but may include other conditions involving changes. The changes can include new calibrations, calibrators, operators, and measuring systems. A specification should contain the conditions changed and unchanged, to the extent practical. In chemistry, the term

inter-serial intermediate precision condition of measurement' is sometimes used to designate this concept.

### 2.3.5 Measurement reproducibility

Measurement precision under reproducibility conditions of measurement.

#### *Reproducibility condition of measurement*

Condition of measurement in a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects. The different measuring systems may use different measurement procedures. A specification should give the conditions changed and unchanged, to the extent practical.

#### *Measurement uncertainty*

Uncertainty of measurement is a parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used. Measurement uncertainty includes components arising from systematic effects, such as components associated with corrections and the assigned quantity values of measurement standards, as well as the definitional uncertainty. Sometimes known systematic effects are not corrected for but are instead treated as uncertainty components.

The parameter may be, for example, a standard deviation called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability. Measurement uncertainty comprises, in general, many components. Some of these may be evaluated by Type A evaluation of measurement uncertainty from the statistical distribution of the quantity values from series of measurements and can be characterized by experimental standard deviations. The other components, which may be evaluated by Type B evaluation of measurement uncertainty, can also be characterized by standard deviations, evaluated from probability density functions based on experience or other information.

#### *Type A evaluation of measurement uncertainty*

This is evaluation of a component of measurement uncertainty by a statistical analysis of quantity values obtained under defined measurement conditions. For various types of measurement conditions, see repeatability condition of measurement, intermediate precision condition of measurement and reproducibility condition of measurement.

#### *Type B evaluation of measurement uncertainty*

This is evaluation of a component of measurement uncertainty determined by means other than a Type A evaluation of measurement uncertainty like:

- Evaluation based on information
- Associated with authoritative published quantity values;
- Associated with the quantity value of a certified reference material;
- Obtained from a calibration certificate and incorporation of drift;
- Obtained from the accuracy class of a verified measuring instrument;
- Obtained from limits deduced through personal experience.

#### *Definitional uncertainty*

Minimum measurement uncertainty resulting from the inherently finite amount of detail in the definition of a measurand. Any change in the descriptive detail of a measurand requires another model leading to another measurand having another definitional uncertainty. In some other literature the concept ‘definitional uncertainty’ is termed “intrinsic uncertainty”.

#### *Standard measurement uncertainty*

Standard uncertainty is measurement uncertainty expressed as a standard deviation.

#### *Combined standard measurement uncertainty*

Standard measurement uncertainty that is obtained from the measurement results of the input quantities in a measurement model.

#### *Uncertainty budget*

Statement of a measurement uncertainty, of the components of that measurement uncertainty, and of their calculation and combination. The uncertainty budget should include the measurement model, estimates and measurement uncertainties of the quantities in the measurement model, covariances, type of applied probability density functions, degrees of freedom, type of evaluation of measurement uncertainty, and any coverage factor.

#### *Target measurement uncertainty*

Measurement uncertainty specified as a goal or optimum and decided on the basis of the intended use of measurement results.

#### *Expanded measurement uncertainty*

Product of a combined standard measurement uncertainty and a factor larger than the number one. Expanded measurement uncertainty is termed “overall uncertainty” and also simply “uncertainty” in documents. The term ‘factor’ in this definition refers to a coverage factor.

### 2.3.6 Coverage interval, probability and factor

#### *Coverage interval*

This is interval containing the set of true quantity values of a measured with a stated probability, based on the information available. A coverage interval does not need to be centred on the measured quantity value. A coverage interval should not be termed ‘confidence interval’ to avoid confusion with the statistical concept. A coverage interval can be derived from an expanded measurement uncertainty.

#### *Coverage probability*

Probability that the set of true quantity values of a measurand is contained within a specified coverage interval. This definition pertains to the Uncertainty Approach. The coverage probability is also termed as “level of confidence”.

#### *Coverage factor*

Number larger than one by which a combined standard measurement uncertainty is multiplied to obtain an expanded measurement uncertainty.

### 2.3.7 Calibration

Operation that, under specified conditions, in a first step establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication.

A calibration may be expressed by a statement, calibration function, calibration diagram, calibration curve, or calibration table. In some cases, it may consist of an additive or multiplicative correction of the indication with associated uncertainty. Calibration should not be confused with adjustment of a measuring system, often mistakenly called ‘self-calibration’, nor with verification of calibration. Sometimes the first step alone in the above definition is perceived as being calibration.

#### *Calibration hierarchy*

Sequence of calibrations from a stated reference to the final measuring instrument or measuring system, where the outcome of each calibration depends on the outcome of the previous calibration. Be aware that measurement uncertainty necessarily increases along the sequence of calibrations. The elements of a calibration hierarchy are one or more measurement standards or calibrators and measuring systems operated according to measurement procedures. For this definition, the ‘stated reference’ can be a definition of a measurement unit through its practical realization, or a measurement procedure, or a measurement standard.

A comparison between two measurement standards may be viewed as a calibration if the comparison is used to check and, if necessary, correct the quantity value and measurement uncertainty attributed to one of the measurement standards.

### 2.3.8 Metrological traceability

Property of a measurement result whereby the result can be related to a stated reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty. For this definition, a ‘stated reference’ can be a definition of a measurement unit through its practical realization, or a measurement procedure including the measurement unit for a non-ordinal quantity, or a measurement standard. Metrological traceability requires an established calibration hierarchy. Specification of the stated reference must include the time at which this reference was used, along with any other relevant metrological information about the reference, such as when the first calibration in the calibration hierarchy was performed.

For measurements with more than one input quantity in the measurement model, each of the input quantities should itself be metrologically traceable and the calibration hierarchy involved may form a branched structure or a network. The effort involved in establishing metrological traceability for each input quantity should be commensurate with its relative contribution to the measurement result.

Metrological traceability by itself does not ensure adequate measurement uncertainty or absence of mistakes. A comparison between two measurement standards may be viewed as a calibration if the comparison is used to check and, if necessary, correct the quantity value and measurement uncertainty attributed to one of the measurement standards.

The abbreviated term “traceability” is sometimes used for ‘metrological traceability’ as well as for other concepts, such as ‘sample traceability’ or ‘document traceability’ or ‘instrument traceability’, where the history (‘trace’) of an item is meant. Therefore, the full term is preferred.

#### *Metrological traceability chain*

Sequence of measurement standards and calibrations that is used to relate a measurement result to a stated reference. A metrological traceability chain is defined through a calibration hierarchy. The metrological traceability chain is used to establish metrological traceability of the measurement result.

#### *Metrological traceability to a measurement unit*

Metrological traceability where the stated reference is the definition of a measurement unit through its practical realization. The expression ‘traceability to the SI’ means metrological traceability to a measurement unit of the International System of Units.

### 2.3.9 Verification and validation

#### *Verification*

Provision of objective evidence that a given item fulfils specified requirements, taking any measurement uncertainty into consideration.

For example:

- Confirmation that a given reference material as claimed is homogeneous for the quantity and measurement procedure concerned, down to a test portion having a mass of 10 mg;
- Confirmation that stated performance properties or legal requirements of a measuring system are achieved;
- Confirmation that a stated target measurement uncertainty can be met.

The item to verify may be, e.g., a process, measurement procedure, material, compound, or measuring system.

The specified requirements may be, e.g., that a manufacturer’s specifications are met.

In legal metrology, verification pertains to the examination and marking and/or issuing of a verification certificate for a measuring instrument. Verification should not be confused with calibration or validation.

In chemistry, verification of identity of entity involved, or of activity, requires a description of the structure or properties of that entity or activity.

#### *Validation*

This is verification, where the specified requirements are adequate for a stated use. For example, a measurement procedure, ordinarily used for the measurement of nitrogen concentration in water, may be validated also for the measurement of nitrogen concentration in human serum.

### 2.3.10 Metrological comparability of measurement results

This is comparability of measurement results that are metrologically traceable to the same reference like measurement results, for the distances from Earth to Moon and from Paris to London, are comparable when they are both metrologically traceable to the same measurement unit, for instance the metre.



Metrological comparability of measurement results does not necessitate that the measured quantity values and associated measurement uncertainties compared are of the same order of magnitude.

#### 2.3.11 Metrological compatibility of measurement results

Property of all pairs of measurement results for a specified measurand such that the absolute value of the difference of the measured quantity values is smaller than some chosen multiple of the standard measurement uncertainty of that difference. Metrological compatibility of measurement results replaces the traditional concept of “staying within the error”, as it represents the criterion for deciding whether two measurement results refer to the same measured or not. If in a set of measurements of a measurand, thought to be constant, a measurement result is not compatible with the others, either the measurement was not correct (e.g. its measurement uncertainty was assessed as being too narrow) or the measured quantity changed between measurements.

Correlation between the measurements influences metrological compatibility. If the measurements are completely uncorrelated, the standard measurement uncertainty of their difference is equal to the root mean square sum of their standard measurement uncertainties, while it is lower for positive covariance or higher for negative covariance.

#### 2.3.12 Measurement model and measurement function

##### *Measurement model*

Mathematical relation among all quantities known to be involved in a measurement. A general form of the measurement model is the equation  $h(Y, X_1, \dots, X_n) = 0$ , where  $Y$ , the output quantity in the measurement model, is the measured that is to be inferred from information about input quantities in the measurement model  $X_1, \dots, X_n$ .

In more complex cases where there are two or more output quantities, the measurement model consists of more than one equation.

##### *Measurement function*

Function of quantities, the value of which, when calculated using known quantity values for the input quantities in a measurement model, is a measured quantity value of the output quantity in the measurement model.

If the measurement model  $h(Y, X_1, \dots, X_n) = 0$  can explicitly be written as  $Y = f(X_1, \dots, X_n)$ , where  $Y$  is the output quantity in the measurement model, the function  $f$  is the measurement function. More generally,  $f$  may symbolize an algorithm, yielding for input quantity values  $x_1, \dots, x_n$  a corresponding unique output quantity value  $y = f(x_1, \dots, x_n)$ . The measurement function is also used to calculate the measurement uncertainty associated with the measured quantity value of  $Y$ .

##### *Input quantity in a measurement model*

Quantity that must be measured, or a quantity, the value of which can be otherwise obtained, in order to calculate a measured quantity value of a measurand. For example, when the length of a rod is being measured, temperature, length, and the linear thermal expansion coefficient of

the rod are input quantities in a measurement model. An input quantity in a measurement model often is an output quantity of a measuring instrument or measuring system.

#### *Output quantity in a measurement model*

Quantity, the measured value of which is calculated using the values of input quantities in a measurement model

#### 2.3.13 Influence quantity

Quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result.

For example:

- Frequency in the direct measurement with an ammeter of the constant amplitude of an alternating current;
- Amount-of-substance concentration of bilirubin in a direct measurement of haemoglobin amount-of-substance concentration in human blood plasma;
- Temperature of a micrometer used for measurement of length of a rod, but not the temperature of the rod itself which may enter in the definition of the measurand;
- Background pressure in the ion source of a mass spectrometer during a measurement of amount-of-substance fraction.

An indirect measurement involves a combination of direct measurements, each of which may be affected by influence quantities.

#### 2.3.14 Correction

Modification, applied to a measured quantity value, to compensate for a known systematic effect. The modification can take different forms, such as an added or a factor, or can be deduced from a table.

#### 2.3.15 Blank indication

Indication obtained from a phenomenon, body, or substance similar to the one under investigation, but for which the quantity of interest is assumed to have zero value.

## 2.4 Metrology: part 3

### 2.4.1 Measuring instrument

Device used for making measurements, alone or in conjunction with supplementary device(s). A measuring instrument alone may be considered to be a measuring system. A measuring instrument may be an indicating measuring instrument or a material measure.

#### *Measuring system*

Set of one or more measuring instruments and often other devices, including any reagent and supply, assembled and adapted to give measured quantity values within specified intervals for quantities of specified kinds. A measuring system may consist of only one measuring instrument.

#### *Indicating measuring instrument*

Measuring instrument providing an output signal carrying information about the value of the quantity being measured like ammeter, micrometre, thermometer etc. An indicating measuring instrument may provide a record of its indication. The output signal may be presented in visual or acoustic form. It may also be transmitted to one or more other devices.

#### *Displaying measuring instrument*

Indicating measuring instrument where the output signal is presented in visual form.

#### *Scale of a displaying measuring instrument*

Part of a displaying measuring instrument, consisting of an ordered set of marks together with any associated numbers or quantity values.

### 2.4.2 Material measure

Measuring instrument reproducing or supplying, in a permanent manner during its use, quantities of one or more given kinds, each with an assigned value. The indication of a material measure is its assigned value.

Examples of material measure:

- weight piece;
- volume measure (supplying one or several quantity values, with or without a quantity scale);
- standard electric resistor;
- line scale (ruler);
- gauge block;
- standard signal generator.

#### *Measuring transducer*

Device, used in measurement, that provides an output quantity having a specified relation to the input quantity like thermocouple, current transformer, strain gauge, pH electrode, Bourdon tube, bimetal strip and others.

### 2.4.3 Sensor

Element of a measuring system that is directly affected by the phenomenon, body, or substance carrying the quantity to be measured.

For example:

- Sensing coil of a platinum resistance thermometer;
- Rotor of a turbine flow meter;
- Bourdon tube of a pressure gauge;
- Float of a level-measuring instrument;
- Photocell of a spectrometer;
- Thermotropic liquid crystal which changes colour as a function of temperature.

In some fields the term "detector" is used for this concept.

### 2.4.4 Detector

Device or substance that indicates the presence of a phenomenon, body, or substance when a threshold value of an associated quantity is exceeded. Such devices are halogen leak detector or litmus paper and others.

In some fields the term "detector" is used for the concept of sensor. In chemistry the term "indicator" is frequently used for this concept.

### 2.4.5 Measuring chain

Series of elements of a measuring system constituting a single path of the signal from the sensor to the output element like an electro-acoustic measuring chain comprising a microphone, attenuator, filter, amplifier, and voltmeter or a mechanical measuring chain comprising a Bourdon tube, system of levers, and mechanical dial.

#### *Adjustment of a measuring system*

Set of operations carried out on a measuring system in order that it provide prescribed indications corresponding to given values of the quantity to be measured. Types of adjustment include zero adjustment of a measuring system, offset adjustment, and span adjustment (sometimes called gain adjustment).

Adjustment of a measuring system should not be confused with calibration. After an adjustment, a measuring system usually must be recalibrated.

#### *Zero adjustment of a measuring system*

Adjustment of a measuring system providing a null indication corresponding to a null value of the quantity to be measured.

### 2.4.6 Indication

Quantity value provided by a measuring instrument or a measuring system. The indication may be presented in visual or acoustic form or may be transferred to another device. The indication is often given by the position on the display for analog outputs, the displayed or printed number for digital outputs, the code pattern for code outputs, or the assigned quantity value for material

measures. The indication and the value of the quantity being measured are not necessarily values of quantities of the same kind.

#### *Indication interval*

Set of quantity values bounded by the extreme possible indications. An indication interval is usually stated in terms of its smallest and greatest quantity values, for example, 99 V to 201 V. In some fields the term is ‘range of indications’.

#### *Nominal indication interval*

Set of quantity values, bounded by rounded or approximate extreme indications, obtainable with a particular setting of the controls of a measuring instrument or measuring system and used to designate this setting. A nominal indication interval is usually stated as its smallest and greatest quantity values, for example, “100 V to 200 V”. Where one of the extreme quantity values is zero, the nominal indication interval is sometimes stated solely as the greatest quantity value, for example a nominal indication interval of 0 V to 100 V is expressed as “100 V”. In some fields the term is “nominal range”.

#### *Range of a nominal indication interval*

Absolute value of the difference between the extreme quantity values of a nominal indication interval. For a nominal indication interval of  $-10\text{ V}$  to  $+10\text{ V}$ , the range of the nominal indication interval is 20 V. Range of a nominal indication interval is sometimes termed “span of a nominal interval”.

#### *Nominal quantity value*

Rounded or approximate value of a characterizing quantity of a measuring instrument or measuring system that provides guidance for their appropriate use.

For example:

- 100  $\Omega$  as the nominal quantity value marked on a standard resistor;
- mol/l as the nominal quantity value for amount-of-substance concentration of a solution of hydrogen chloride, HCl;
- 40  $^{\circ}\text{C}$  as a maximum Celsius temperature for storage.

The term “nominal value” is also used for ‘nominal property value’.

#### *Measuring interval or working interval*

Set of values of the quantities of the same kind that can be measured by a given measuring instrument or measuring system with specified instrumental uncertainty, under defined conditions.

In some fields the term is “measuring range” or “measurement range”. The lower limit of a measurement interval should not be confused with the detection limit of the measuring instrument.

#### *Steady state condition*

Operating condition of a measuring instrument or measuring system in which the calibration remains valid for a measurand varying with time.

#### 2.4.7 Rated operating condition

Condition that must be fulfilled during measurement in order that a measuring instrument or measuring system perform as designed. The rated operating condition generally specifies intervals of values for the quantity being measured and for any influence quantity.

##### *Limiting condition*

Extreme condition that a measuring instrument or measuring system is required to withstand without damage, and without degradation of specified metrological properties, when it is subsequently operated under its rated operating conditions.

Limiting conditions for storage, transport or operation can differ. Limiting conditions can include limiting values of the quantity being measured and of any influence quantity.

##### *Reference condition*

This separates two different cases. For performance evaluation this is condition of use prescribed for evaluating the performance of a measuring instrument or measuring system or for comparison of measurement results. For instrumental uncertainty this is condition of use for a measuring instrument or measuring system, under which the specified instrumental uncertainty is the smallest possible. A reference condition specifies an interval of values of the measurand and of the influence quantities.

## 2.5 Metrology: part 4

### 2.5.1 Sensitivity

Quotient of the change in the indication and the corresponding change in the value of the quantity being measured. The sensitivity can depend on the value of the quantity being measured. The change considered in the value of the quantity being measured must be large compared with the resolution.

### 2.5.2 Selectivity of a measuring system

Selectivity is capability of a measuring system, using a specified measurement procedure, to provide measurement results, for one or more measurands, that do not depend on each other nor on any other quantity in the system undergoing measurement.

For example:

- Capability of a measuring system including a mass spectrometer to measure the ion current ratio generated by two specified compounds without disturbance by other specified sources of electric current.
- Capability of a measuring system to measure the power of a signal component at a given frequency without being disturbed by signal components or other signals at other frequencies.
- Capability of a receiver to discriminate between a wanted signal and unwanted signals, often having frequencies slightly different from the frequency of the wanted signal.
- Capability of a measuring system for ionizing radiation to respond to a radiation to be measured in the presence of concomitant radiation.
- Capability of a measuring system to measure the amount-of-substance concentration of creatininium in blood plasma by a Jaffé-procedure without interference from the glucose, urate, ketone, and protein concentrations.
- Capability of a mass spectrometer to measure the amount-of-substance abundance of the  $^{28}\text{Si}$  isotope in silicon from a geological deposit.

In physics, there is only one measurand, the other quantities are of the same kind as the measurand, and they are input quantities to the measuring system. In chemistry, the measured quantities often involve different components in the system undergoing measurement and these quantities are not necessarily of the same kind. In chemistry, selectivity of a measuring system is usually obtained for quantities with selected components in concentrations within stated intervals. Selectivity as used in physics is a concept close to specificity as it is sometimes used in chemistry.

### 2.5.3 Resolution

Resolution is smallest change in a quantity being measured that causes a perceptible change in the corresponding indication of for displaying devices smallest difference between indications that can be meaningfully distinguished. The resolution can depend on, for example, noise (internal or external) or friction. It may also depend on the value of the quantity being measured.

#### 2.5.4 Discrimination threshold

Largest change in the value of a quantity being measured that causes no detectable change in the corresponding indication. The discrimination threshold may depend on, for example, noise (internal or external) or friction. It can also depend on the value of the quantity being measured and how the change is applied.

#### 2.5.5 Dead band

Maximum interval through which the value of a quantity being measured can be changed in both directions without producing a detectable change in the corresponding indication. The dead band can depend on the rate of change.

#### 2.5.6 Detection limit

In chemistry measured quantity value, obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component in a material is  $\beta$ , given a probability  $\alpha$  of falsely claiming its presence. In some literature the abbreviation LOD is used. Default values recommended for  $\alpha$  and  $\beta$  equal to 0.05.

#### 2.5.7 Stability

Ability of a measuring instrument or measuring system to maintain its metrological properties constant with time. Stability may be quantified in several ways like in terms of the duration of a time interval over which a metrological property changes by a stated amount or in terms of the change of a property over a stated time interval.

#### 2.5.8 Instrumental drift

Continuous change in an indication, related neither to a change in the quantity being measured nor to a change of any recognized influence quantity. For a material measure, the drift is a change of the value of the supplied quantity that is not due to a change of any recognized influence quantity.

#### 2.5.9 Variation due to an influence quantity

Difference of the indications for the same measured quantity value, or of the quantity values supplied by a material measure, when an influence quantity assumes successively two different quantity values.

#### 2.5.10 Step response time

Duration between the instant when an input quantity value of a measuring instrument or measuring system is subjected to an abrupt change between two specified constant quantity values and the instant when the corresponding indication settles within specified limits around its final steady value.

#### 2.5.11 Instrumental uncertainty

Component of measurement uncertainty arising from the measuring instrument or measuring system in use, and obtained by its calibration. Information relevant to instrumental uncertainty may be given in the instrument specifications.



#### 2.5.12 Accuracy class

Class of measuring instruments that meet stated metrological requirements that are intended to keep measurement errors or instrumental uncertainties within specified limits under specified operating conditions. An accuracy class is usually denoted by a number or symbol adopted by convention. A related concept is measurement accuracy. Measuring instrument includes material measure.

#### 2.5.13 Maximum permissible error

Extreme value of the measurement error, with respect to a known reference quantity value, permitted by specifications or regulations for a given measurement, measuring instrument, or measuring system. Generally, the two extreme values are taken together and are termed 'maximum permissible errors' or 'limits of error'. The term "tolerance" should not be used to designate 'maximum permissible error'.

##### *Datum error*

Measurement error, with respect to a known reference quantity value, of a measuring instrument or measuring system at a specified measured quantity value.

##### *Zero error*

Datum error for zero measured quantity value.

#### 2.5.14 Null measurement uncertainty

Measurement uncertainty for zero measured quantity value. The null measurement uncertainty is associated with a null or near zero indication and covers the interval where one does not know whether the measurand is too small to be detected or the signal of the measuring instrument is due only to noise.

## 2.6 Metrology: part 5

### 2.6.1 Calibration diagram

Graphical expression of the relation between indication and corresponding measurement result. A calibration diagram is the strip of the plane defined by the axis of the indication and the axis of measurement result, that represents the response of the measuring instrument to different measured quantity values. A one-to-many relation is given, and the width of the strip for a given indication provides the instrumental uncertainty. Alternative expressions of the relation include a calibration curve and associated measurement uncertainty, a calibration table, or a set of functions. This concept pertains to a calibration when the instrumental uncertainty is large in comparison with the measurement uncertainties of the measurement standards.

#### *Calibration curve*

Expression of the relation between indication and corresponding measured quantity value. A calibration curve expresses a one-to-one relation that does not supply a measurement result as it bears no information about the measurement uncertainty.

### 2.6.2 Measurement standard

Realization of the definition of a given quantity, with stated quantity value and measurement uncertainty, used as a reference.

For example:

- 1 kg mass standard;
- 100  $\Omega$  standard resistor;
- caesium frequency standard;
- standard hydrogen electrode;
- set of reference solutions of cortisol in human serum having certified concentrations and measurement uncertainties of cortisol;
- reference material providing certified values with measurement uncertainties for the mass concentration of each of ten different proteins.

The “realization of the definition of a given quantity” can be provided by a measuring system, a material measure, or a certified reference material. A measurement standard is frequently used as a reference to assign measurement results to other quantities of the same kind, thereby establishing metrological traceability through calibration of other measuring standards, measuring instruments, or measuring systems. In many cases, measurement standards are realizations of the definition of a measurement unit. The standard measurement uncertainty associated with a measurement standard is always a component of the combined standard measurement uncertainty in a measurement result obtained using the measurement standard. Frequently, this component is small compared with other components of the combined standard uncertainty.

Several quantities of the same kind or of different kinds may be realized in one device which is commonly also called a measurement standard.

The word “embodiment” is sometimes used in the English language instead of “realization”. In science and technology, the English word “standard” is used with two different meanings: as a widely adopted written standard, specification, technical recommendation or similar document (in French “norme”) and as a measurement standard (in French “étalon”). The quantity value and measurement uncertainty must be ensured at the time when the measurement standard is used.

#### *International measurement standard*

Measurement standard recognized by signatories to an international agreement and intended to serve worldwide.

#### *National measurement standard*

Measurement standard recognized by national authority to serve in the country.

#### *Primary measurement standard*

Measurement standard whose quantity value and measurement uncertainty are established using a primary measurement procedure. For example:

- Primary measurement standard of amount-of-substance concentration prepared by dissolving a known amount of substance of a chemical component to a known volume of solution.
- Primary measurement standard for pressure based on separate measurements of force and area.
- Primary measurement standard for isotope amount-of-substance ratio measurements, prepared by mixing known amount of substance of specified isotopes.

#### *Secondary measurement standard*

Measurement standard whose quantity value and measurement uncertainty are assigned through calibration with respect to a primary measurement standard for a quantity of the same kind. The relation may be obtained directly between the primary measurement standard and the secondary measurement standard, or involve an intermediate measuring system calibrated by the primary measurement standard and assigning a measurement result to the secondary measurement standard. A measurement standard having its quantity value assigned by a ratio primary measurement procedure is a secondary measurement standard.

#### *Reference measurement standard*

Measurement standard designated for the calibration of working measurement standards for quantities of a given kind in a given organization or at a given location.

#### *Working measurement standard*

Measurement standard that is used routinely to calibrate or verify measuring instruments or measuring systems. A working measurement standard is usually calibrated with respect to a reference measurement standard. Working measurement standard is also named as “check standard” or “control standard”.

#### *Travelling measurement standard*

Measurement standard, sometimes of special construction, intended for transport between different locations like a portable battery-operated caesium-133 frequency standard.

#### *Transfer device*

Device used as an intermediary to compare measurement standards. Sometimes, measurement standards are used as transfer devices.

#### *Intrinsic measurement standard*

Measurement standard based on a property of a phenomenon or substance, sufficiently stable and reproducible for its intended use. For example:

- Triple-point-of-water cell as an intrinsic measurement standard of thermodynamic temperature;
- Intrinsic measurement standard of electric potential difference based on the Josephson effect;
- Intrinsic measurement standard of electric resistance based on the quantum Hall effect;
- Sample of copper as an intrinsic standard of electric conductivity.

The quantity value of an intrinsic measurement standard is assigned by consensus and does not need to be established by relating it to another measurement standard of the same type. Its measurement uncertainty is determined by considering two components - that associated with its consensus quantity value and that associated with its construction, implementation and maintenance. An intrinsic measurement standard usually consists of a system produced according to the requirements of a consensus procedure and subject to periodic verification. The consensus procedure may contain provisions for application of corrections necessitated by the implementation. Intrinsic measurement standards that are based on quantum phenomena usually have outstanding stability.

The adjective ‘intrinsic’ does not mean that this measurement standard may be implemented and used without special care or that this measurement standard is immune to spurious effects.

#### *Conservation of a measurement standard*

Also named as maintenance of a measurement standard is set of operations necessary to preserve the metrological properties of a measurement standard within stated limits. Conservation commonly includes periodic verification of predefined metrological properties or calibration, storage under suitable conditions, and specified care in use.

#### *Calibrator*

Measurement standard used in calibration.

### 2.6.3 Reference material

Material, sufficiently homogeneous and stable regarding one or more properties, used in calibration, in assignment of a value to another material, or in quality assurance. ‘Reference material’ comprises materials embodying quantities as well as nominal properties. Examples allowing to understand what is reference material describing quantities:

- Water of stated purity, the dynamic viscosity of which is used to calibrate viscometers;

- Human serum without an assigned quantity value for the concentration of the inherent cholesterol, used only as a measurement precision control material;
- Fish tissue containing a stated mass fraction of a dioxin, used as a calibrator.

Examples of reference materials describing properties:

- Colour chart indicating one or more specified colours;
- DNA compound containing a specified nucleic acid sequence;
- Urine containing androstenedione.

In this definition, “value” covers both quantity value and ‘nominal property value’. Reference materials with or without assigned quantity values can be used for precision control whereas only reference materials with assigned quantity values can be used for calibration or measurement trueness control.

Some reference materials have quantities that are metrologically traceable to a measurement unit outside a system of units. Such materials include vaccines to which International Units (IU) have been assigned by the World Health Organization.

The specifications of a reference material should include its material traceability, indicating its origin and processing. In a given measurement, a reference material can only be used for either calibration or quality assurance. A reference material is sometimes incorporated into a specially fabricated device.

For example:

- Substance of known triple-point in a triple-point cell;
- Glass of known optical density in a transmission filter holder;
- Spheres of uniform particle size mounted on a microscope slide;
- Array of Josephson junctions.

#### 2.6.3.1 Certified reference material

Reference material, accompanied by documentation issued by an authoritative body and referring to valid procedures used to obtain a specified property value with uncertainty and traceability. For example, human serum with assigned quantity value for the concentration of cholesterol and associated measurement uncertainty stated in an accompanying certificate, used as calibrator or measurement trueness control material. Certified reference material has enclosed documentation in the form of a ‘certificate’. Valid procedures for the production and certification of certified reference materials are given in standards documentation.

In this definition, “uncertainty” covers both ‘measurement uncertainty’ and ‘uncertainty of a nominal property value’, such as for identity and sequence, expressed as probabilities. “Traceability” covers both ‘metrological traceability’ of a quantity value and ‘traceability of a nominal property value’.

‘Certified reference material’ is a specific concept under ‘reference material’. Specified quantity values in certified reference materials require metrological traceability with associated measurement uncertainty.

### 2.6.3.2 Commutability of a reference material

Property of a reference material, demonstrated by the closeness of agreement between the relation among the measurement results for a stated quantity in this material, obtained according to two given measurement procedures, and the relation obtained among the measurement results for other specified materials. The reference material in question is usually a calibrator and the other specified materials are usually routine samples. The measurement procedures referred to in the definition are the one preceding and the one following the reference material (calibrator) in question in a calibration hierarchy. The stability of commutable reference materials is monitored regularly.

### 2.6.4 Reference data

Data that is critically evaluated and verified, obtained from an identified source, and related to a property of a phenomenon, body, or substance, or a system of components of known composition or structure. For example, data for solubility of chemical compounds. “Data” is commonly used in the singular sense, instead of “datum”.

#### *Standard reference data*

Reference data issued by a stated recognized authority.

#### *Reference quantity value*

Quantity value, generally accepted as having a suitably small measurement uncertainty, to be used as a basis for comparison with values of quantities of the same kind. A reference quantity value with associated measurement uncertainty is usually referred to:

- A material, e.g. a certified reference material;
- A device, e.g. a stabilized laser;
- A reference measurement procedure;
- A comparison of measurement standards.

A reference quantity value must be metrologically traceable.

## 2.7 Classification of measurement errors: part 1

### 2.7.1 Measurement error

We all every day perform measurements. Beginning with rough estimation of distance to something, and ending with application of precise instruments. What makes difference? Error, fault, precision, accuracy etc. In Chapter 1 we met many expressions defining exact thing. When we perform measurement, it is actually an experiment of physics. At least it can be treated so. We all know that two species of science - physics and mathematics, go together hand-by-hand. We express physics with mathematical formulas. When we cannot perform straight experiment, it is possible to apply equivalent model a get desired result. It is usually assumed, that measured result is accurate and satisfy our requirements. However, it is not true, never was and won't be in the future. We simply sacrifice quality of measurement to ease of use. Where is the difference – all mathematical calculations are exact. If done properly, they do not contain error. If we are able to calculate that distance is 1 meter, then it is. If we do physical experiment and apply instrument, then acquired result contains error. It is expressed as difference between obtained value  $X$  and reference value  $X_0$ .

$$\Delta = X - X_0 \quad (2.1)$$

Volume of this error relate to accuracy of the applied instrument. You might say: “Ok, I have a ruler exactly one-meter-long so I can measure everything within that meter!” Yes, You can with some accuracy and that should be kept in mind. This is one of direct measurements that give desired information within certain error range. If 1 m long ruler have 1mm divisions on it, so we can assume, that everything we measure has accuracy of one millimetre. And how about things we cannot measure directly? Indirect measurements are that for. And that's it. There is no other type of measurements in the world – direct and indirect only.

So when we perform measurement, it depends of instrument accuracy and external components like temperature, humidity, pressure, magnetic flux, ionizing radiation etc. Basically, is the tool we use affected by any external event and how much? For example, one-meter-long ruler in room temperature appear to be longer than in South pole during Antarctic winter. The same relate also to indirect measurements where we should take in account all influences to tools we use for measurements to obtain data for calculation of desired value.

So in scientific measurements it is mandatory to provide total error of obtained result. For example,  $L = (1000 \pm 1) \text{ mm}$ .

Now you might have question, how many digits should the measurement contain, how many are for error. Every value contains so called significant digits. Those, who are specifying the value where last one is one that contains error. If we have leading zeroes before one, they cannot be treated as significant unless two digits are taken for absolute error. For example:

Number	Count of significant digits
1000	4
$800 \cdot 10$	3

$9 \cdot 10^3$	1
$0,2 \cdot 10^4$	1
$0,03 \cdot 10^5$	1

From perspective of mathematics, all digits are significant however with physical measurements it is not true. Where is difference – if we use values without specifying absolute error, then it means  $\pm 1$  of last significant digit. From here it is clearly visible, that it is not possible to measure a value with tool that has smallest division equal to it or larger. However technically it is possible to measure distance of 1 mm with 1 m long ruler with 1 mm divisions, from metrology perspective such measurement should be treated as failure and discarded.

Another important thing is how measured value is expressed.

Let's take for example

Value	Accuracy
1234 mm	1 mm
$200 \cdot 10^1$ mm	1 cm
$5 \cdot 10^3$ mA	1 A
$0,2 \cdot 10^4$ mV	1 V
$0,08 \cdot 10^5$ m $\Omega$	1 $\Omega$

When physical values are measured, it is not allowed to display value without absolute error rounded to one or two significant digits. Always keep in mind, that dimension of absolute value is the same as for measurement.

Samples of measurement value display:

$I = 3,45 \pm 0,4$  – bad format. Should be  $3,5 \pm 0,4$

$I = 3,45 \pm 0,15$  – correct, two significant digits used

$U = 127,5 \pm 0,04$  – bad format. Should be  $127,50 \pm 0,04$  – tailing zero added to keep range

$Z = 1024,7 \pm 40$  – bad format. Should be  $(100 \pm 4) \cdot 10$ .

### 2.7.2 Absolute error

Till now only one component of absolute error is mentioned – smallest division of instruments scale. Confusion about measurement value equal to smallest division is reasonable and that's why absolute error ( $\Delta$ ), named also as margin of error, actually contains several components. They are:



$$\Delta = \Delta S + \Delta M + \Delta P \quad (2.2)$$

Where:  $\Delta S$  stands for error from measurement system. Depending on system used usually value of  $\Delta S$  is within range  $0,2...0,5*[\Delta]$  where  $[\Delta]$  is maximal acceptable error of measurement system. Such approach is caused by need to deal with errors caused by methodology of measurement and subject itself. A fact that measurement system changes during runtime allow to add certain metrological resilience to the system. If multiple measurement systems have been used to obtain value of one physical parameter, then each of them have their own  $\Delta S$  and absolute error value. By type of appearance this error can be as constant or alternating. Alternating also can be of two types – progressing or periodic. If error is constant, then it is easy to take it in account. More complicated situation is when error appear to be alternating, progressing and also periodic. Then a mandatory action is to have or obtain expression that describe changes  $\Delta S$  during measurement session.

$\Delta M$  stands for error introduced by method used for measurement. It is determined by analysis of obstacles affected by parameter changes during experiment, value transformation function, methodology of data acquisition, interference between value measured and measurement device and other factors not related to measurement system itself. It is clear that best instrument is the one that does not interfere with measured object, however almost all measurements, where instrument contact measurable object, interference exist. Depending on method used, this sort of error can contain a separate component named as instrument error ( $\Delta i$ ). This component relates to any discrepancy caused by non-linear conversion of measurable value in indirect measurement, limited range of sensor used, inertia of measurement system, ageing of reference object etc.

$\Delta P$  describes any private or personal error derived from nominal value of measurement device scale division. In some literature this type of error is named also as subjective error. Usually it is 0.2 to 0.3 but not more than 0.5 of scale division.

### 2.7.3 Measurement

If measurement target itself can satisfy expected results while being in certain range, it is necessary to add absolute error values to clarify validity of data. Visual interpretation of this statement is given in Fig. 2.3

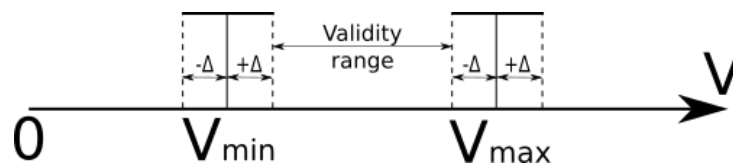


Figure 2.3. Absolute error application to measured values

Let's have a system where a direct measurement of value  $V$  is performed and it is expected to have it within validity range. So how it is possible to determine, that obtained value is valid? From measurement statement it is predefined that acceptable value is in between  $V_{min}$  and

$V_{\max}$ . As declared before, absolute error must be added to acquired values. So it is visible that absolute error value narrows clear validity range. In the same time nature of error states, that broader range with  $V_{\min} - \Delta$  and  $V_{\max} + \Delta$  also can contain valid values. To reduce any confusion  $-\Delta$  and  $+\Delta$  range around obtained value is called as ambiguous.

From here it is clearly visible, that if ambient parameters change significantly during one measurement session, it is mandatory to split session in sections and calculate absolute error for each of them.

Do not be confused by symbol  $\Delta$  used for description of absolute error and way's it is obtained. Basically that's because of two methods it can be get. It means – if we do not have reliable reference value  $X_0$ , absolute error can be calculated. Reference value  $X_0$  can be obtained in few following ways:

- Theoretically obtained or science backed declared value;
- Attested value based on research performed by national or international authority;
- Agreed and attested value backed by research and series of experiments performed by team of scientists and/or engineers;
- Expectation or average value of  $n$  measurement series, where  $n$  is large enough count of series to be acceptable as exhaustive and representative.

#### 2.7.4 Random error

If absolute error described before, is expressible and remains almost constant during one measurement session, then another type of error – random error ( $\delta$ ) affect all measurements and renders “do only one measurement” approach unreliable. This is why metrology demands at least 5 or more sequential measurements to be performed to acquire valid data for analysis. Random error has a dodgy character – it might be bigger than calculated absolute error in one measurement and disappear in next. It might have various signs in sequential measurements. For example – a measurement session is performed carefully for a running device to measure value  $Q$  expected to be  $230 \pm 5$ . It appears that within set of measurements there are values 224 and 236 – so outside ambiguous range with value of 1. So here this 1 is a random error added to absolute error. Influence of random error should be kept as small as possible and this can be achieved through larger amount of measurements.

#### 2.7.5 Missed measurement

In some occasions interference of random ( $\delta$ ) and measurement system error ( $\Delta_s$ ) cause an obtained value  $X$  to be far out of expected range. Such case is named as rough error or miss. In some literature a name “blunder” is used. Such value is clearly extractable from measurement series as it is stand-alone and mathematical processing of measurements allow to determine them to exclude from further use.

#### 2.7.6 Relative error

Relationship between absolute error of measurement and measured value itself is named as relative error. It is expressed as

$$\delta X = \Delta X / X, \quad (2.3)$$

Relative error by itself is a dimensionless value and displays level of inaccuracy of obtained result. Sometimes it is expressed in percent –  $\delta X \cdot 100\%$ .

### 2.7.7 Error classification

Errors in direct measurements can be schematically classified as given below in Fig.2.4.

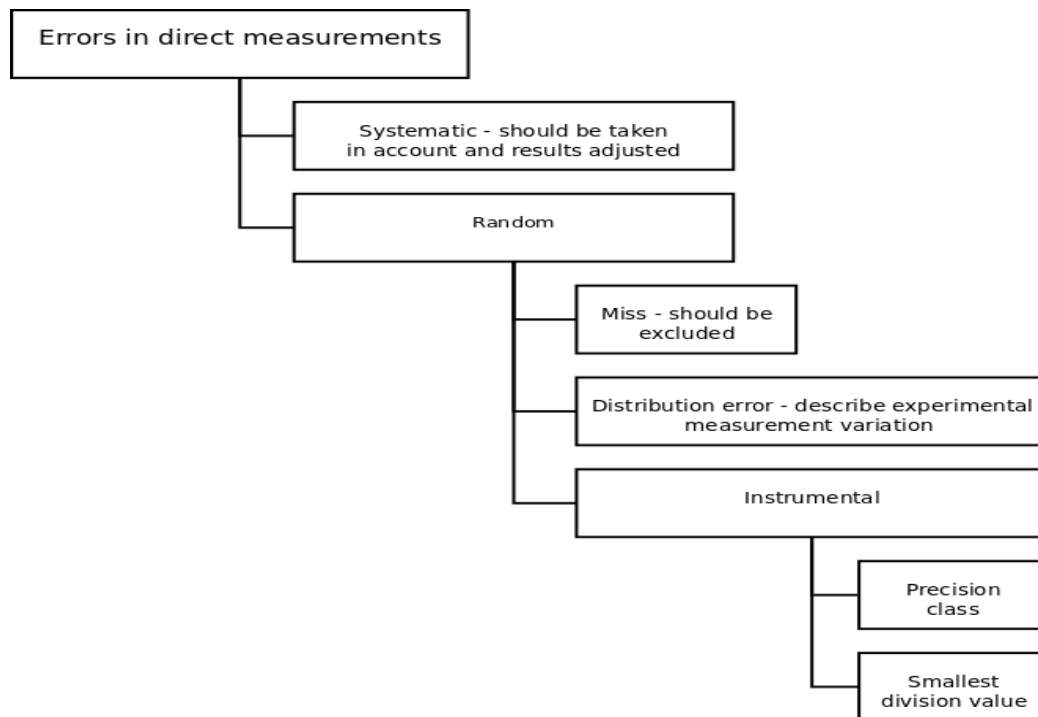


Figure 2.4. Error classification

Errors introduced by instruments come from two aspects:

First is instruments precision class  $\tau$  that depends on physical phenomenon used, components used to manufacture and design. Error component from this aspect can be calculated as  $(\Delta x)_i = (\tau / 100)X$ . Precision class in percent is printed on instruments scale (Fig.2.4.) or provided in accompanying document. If no information is provided or available on any source, precision class can be taken as 4 however some cheap multimeter devices in some cases have error rate exceeding 10%. Such tools should be avoided for scientific measurements.

Second is value of smallest division of the instruments scale. This is very important when selecting instrument for measurement as in some cases device available can contain a nonlinear scale and measured value is not in the “best match” range. In this case error component can be expressed as  $(\Delta X)_{SC} = \frac{1}{2} h$  where h is smallest division.

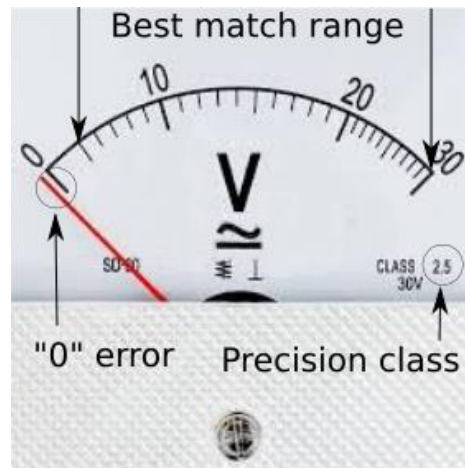


Figure 2.5. Instruments scale

In Fig.2.5, a precision class 2.5 instruments are displayed and best match range is from 5 to 30 volts with smallest division of 1 volt. So it cannot be used as reliable source of values for measurements.

#### 2.7.8 Zero error

Another specific source of error is “0” error and it describe situation when instrument does not display “0” when it should. As it is visible in Fig. 2.5, deflection of a dart from “0” value is almost the same as smallest value of scale in range from 20 to 30 volts. This sort of error should be eliminated as far as possible before measurements and each instrument should have a construction for that. Instruments without “0” error elimination capability or broken, should not be used for measurements.

Specific “0” error case is for digital instruments with very high input impedance a sensitivity. With open circuit it is possible and allowed to have a non zero value on display as specific zeroing and calibration routine should be performed. Zeroing and calibration routine is described by manufacturer of device and presented in accompanying document.

#### 2.7.9 Distribution error

Distribution error is the one that cover obtained value differences of multiple sequential measurements of the same physical phenomenon.

### 2.7.9.1 Cornfield's method

Simplest method to calculate distribution error  $(\Delta x)_d$  is Cornfield's method containing following steps.

- From a set of measurements  $\{X_1, X_2 \dots X_n\}$  take lowest ( $X_{min}$ ) and highest ( $X_{max}$ ) values and calculate average value  $x$ :

$$x = \frac{X_{max} + X_{min}}{2}, \quad (2.4)$$

- Express absolute error as

$$\Delta x_d = \frac{X_{max} - X_{min}}{2}, \quad (2.5)$$

and result can be logged as  $X = x \pm \Delta x_d$  with reliability probability  $\alpha = 1 - \left(\frac{1}{2}\right)^{n-1}$  describing chance that true value is within trusted range  $\langle x - \Delta x_d; x + \Delta x_d \rangle$ .

Reliability probability display, how many measurements of equivalent measurement series falls to trusted range. This method's drawback is also clearly visible as it directly depends on count of measurements in a set. So for 5 measurement set probability  $\alpha$  is 0,9375 but for 25 – 0,99999994. So basically trusted range  $\pm \Delta x_d$  cannot be expanded other ways than by increasing count of measurements.

### 2.7.9.2 Students method

Sequence of operations in this method is following:

- From a set of measurements  $\{X_1, X_2 \dots X_n\}$  with missed measurements already excluded, calculate average value  $x$  according to expression:

$$x = \frac{\sum_{i=1}^n (X_i)^2}{n}, \quad (2.6)$$

- Use obtained value for mean square error calculation by using expression:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (X_i - x)^2}{(n-1) * n}}, \quad (2.7)$$

Obtain reliability probability value  $\alpha$  in the same way, as it is described for Cornfield's method. According to measurement count  $n$  and  $\alpha$  from table read Students coefficient  $t_{\alpha n}$ .

- Now express error of measurement average value  $\Delta x = t_{\alpha n} * \sigma_x$

So desired result  $X = (x \pm \Delta x)$  within reliability probability  $\alpha$ .

In scientific papers trusted range often is declared as  $\Delta x = \sigma X$  with probability value  $\alpha = 0.7$  and it is named as standard and trusted range then is not provided. Usage of Students method is necessary when value of physical parameter is needed to be within desired reliability probability.

In the industry this probability is replaced with accuracy class of the instrument. So for majority of research cases, where precision of measurement is not very crucial, Cornfield's method is completely acceptable. Even more – absolute error can be calculated as

$$\Delta x = \sqrt{(\Delta x)_i^2 + (\Delta x)_{sc}^2 + (\Delta x)_d^2} \quad (2.8)$$

If all error components are calculated for the same reliability probability. While all error values are rounded to one significant digit, in praxis it is sufficient to take one largest error value from those three if it is at least three times bigger than any other. So basically factor, the error taken is related to, will determine precision of the experiment.

## 2.8 Classification of measurement errors: part 2

### 2.8.1 Evaluation of indirect measurement errors

When indirect measurement is performed, its precision is affected by error from measurement of physical value a desired value is derived from. Let's say desired value  $Y$  is a known function of directly measured value  $X$ .

$$Y = f(X) \quad (2.9)$$

As we already know, every  $X$  measured contains absolute error  $\Delta x$  affecting precision of  $Y$ . Addition of  $\Delta x$  mean also addition to result of function in form  $\Delta Y \approx f' \Delta x$ . In case, when indirectly measured value is a function of a set of separate values, then for each argument  $x_i$  we get

$$\Delta Y \approx \frac{\partial f}{\partial x_1} \Delta x_1 + \frac{\partial f}{\partial x_2} \Delta x_2 + \dots + \frac{\partial f}{\partial x_n} \Delta x_n, \quad (2.10)$$

Such approach gives maximal possible error value as it reflects maximum deflection from average value and that in real is unlikely so indirect measurement error can be calculated as

$$\Delta Y \approx \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2 \Delta x_i^2}, \quad (2.11)$$

In real measurements relative precision of various  $x_i$  might differ significantly so might be reasonable to assess for a set of values  $i = 1 \dots, m-1, m, m+1, \dots n$  for inequality

$$\frac{\partial f}{\partial x_m} \Delta x_m > 3 \frac{\partial f}{\partial x_i} \Delta x_i, \quad (2.12)$$

In case when it is true – the largest error value is more than three times larger than next largest, error of indirect measurement  $\Delta Y$  can be assumed to be set by only  $\Delta x_m$  thus

$$\Delta Y \approx \left| \frac{\partial f}{\partial x_m} \right| \Delta x_m, \quad (2.13)$$

### 2.8.2 Error dispersion

Error dispersion from metrology perspective plays a significant role while attempting to determine true value of the measurement. As declared before, to obtain a really true value of certain physical parameter, it should comply to two rules:

- Physical parameter is not changing during whole measurement session;
- Count of measurements performed tends to be infinite.

As we know, both statements are not true. They can not be true as all physical parameters are fluctuating because of entropy, and measurement techniques deliver satisfactory count of readings in certain period of time. So what we can talk about is probability  $p^*(x_i)$  that measurement performed will deliver value within expected range. So expected measurement value is

$$m_x = M[x] = \int_{-\infty}^{\infty} x p(x) dx, \quad (2.14)$$

with mean square error (deviation)

$$\delta_x = \sqrt{\int_{-\infty}^{\infty} (x - m)^2 p(x) dx} \quad (2.15)$$

In case, when any systematic error is excluded, it is possible to declare obtained value  $m_x$  as true value for  $Q$ , where  $Q$  is true value of measurand.

While for measurement singletons error calculations are described before, in the industry we should deal with multiple and continuous measurements. As measurement techniques evolve, situation when one parameter is measured many times without any noticeable alterations in system becomes a common case. Modern monitoring systems perform so many readings of one physical value, so count in terms of metrology can be treated as infinite. Basically, thank to modern devices it is possible to perform almost infinite count of measurements with accuracy that renders part of error calculations as obsolete. However, it is possible to chop series of measurements in separate discrete items, it is more reasonable not to treat measurements as singletons, but as a trusted range where they all fall with certain probability. This means, that for a set of measurements we can express also probability of greatness of difference from average value understanding with this also probability of difference from true value.

What is important, error dispersion and difference probability have the same density function around true value as measurement values itself. That's why it is possible with increased amount of measurements reduce error volume.

Probabilities depending on approach chosen, have four most common distribution functions. Graphical interpretation is presented below in figure 2.6.

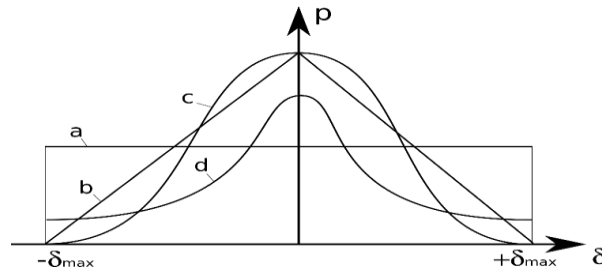


Figure 2.6. Error dispersion probability functions

Function a represent even or equal Simpsons distribution function and mean that in acceptable range probability of error is the same. If such approach is used, the measurand and its absolute error can be expressed without specifying probability of accuracy.

Function b is named as triangular Simpsons distribution and is derived from two even distribution functions. For measurements with multiple errors where each refers to equal or triangular distribution rule, a normalized integrated function for absolute error must be obtained. In most cases it tends to be similar to normal distribution.

Function c is named as normal, or Gauss distribution and is typical for most cases. Function d is named as Students differential distribution function  $S(t,k)$  and it depends on parameter  $t_p = \frac{x-Q}{S_{\bar{x}}}$ , where  $S_{\bar{x}} = \frac{\delta_x}{\sqrt{n}}$  and degree of freedom  $k = n-1$ . Values of this function are already calculated and can be found in table format. So if a limited count of measurements has been



made, and measurements itself are distributed according to normal function, then probability to obtain true value within trusted range, can be obtained according to Students distribution.

$$P[(x - t_p S_{\bar{x}}) \leq Q \leq (x + t_p S_{\bar{x}})] = 2 \int_0^{t_p} S(t, k) dt, \quad (2.16)$$

This permits to provide a measurement result, true value Q, in format:

$$Q = x \pm t_p S_{\bar{x}}, \text{ with probability P in percent.}$$

In fact, for measurement count in series exceeding 20, Students distribution tends to become the same as normal.

### 2.8.3 Measurement graphical interpretation

A need to display a set of measurements in graphical format is common. We all know how easy is to understand a picture, and how difficult it is to comprehend large table of numbers a picture is created from.

There are many ways how data can be displayed, but in science sequence of building a graphic is such:

- Determine a scale and axis. Depending on values you need to display, linear or logarithmic scale should be selected. Those are two most common however depending on what is measured and what is required for display, different free form scales can be chosen. Figure 2.7. below gives an overview of scale diversity;
- Put measurement value points in the grid. Measurement value points must be displayed with error area for science related graphics. This statement might be confusing and almost impossible to fulfil in case when error is tiny when compared to grid. In such case smallest grid division is used for error range border;
- Depending on target, connect points with broken lines, smooth line or approximated “ideal” line. First, broken line, method can be used when process, where measurements are made is straight linear between measurements. Smoothed line can be used when displaying a set of measurements that follow certain non-linear rule and it is necessary to demonstrate compliance. Samples are given in Figure 2.9 below.

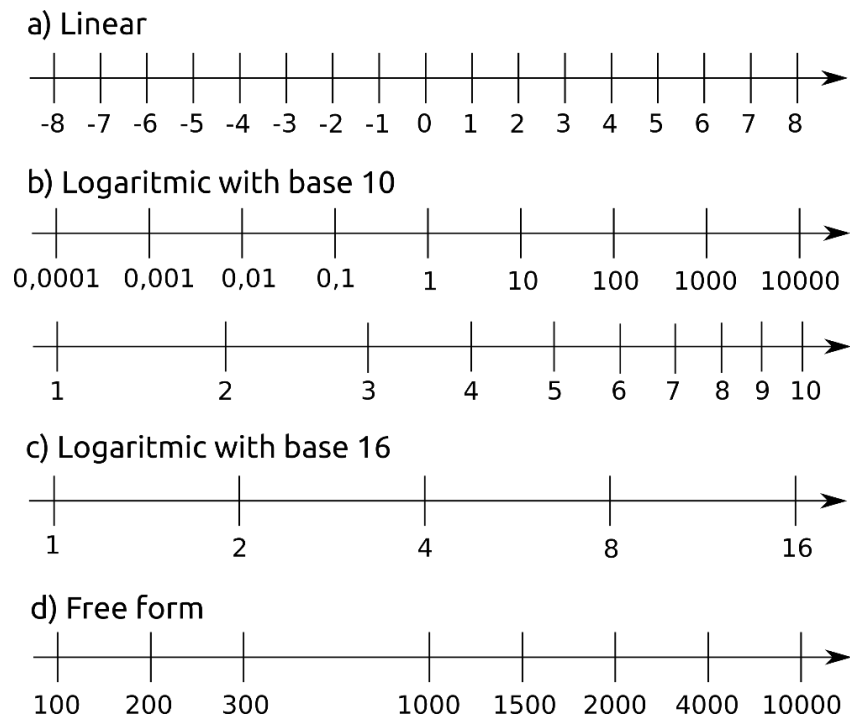


Figure 2.7. Diverse scales used for measurement display

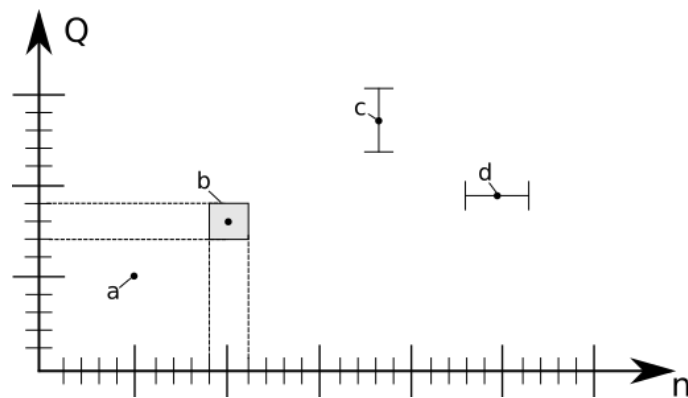


Figure 2.8. Measurement display diversity: a) generic point; b) point with error area c) and d) points with error range for only one dimension

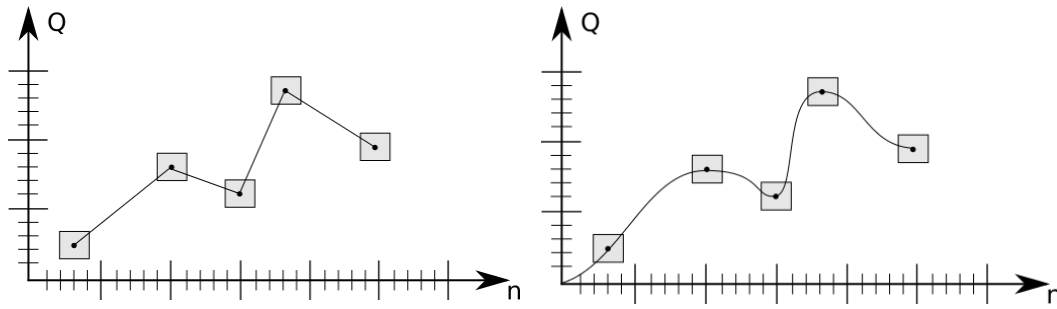


Figure 2.9. Graphical interpretation of measurement

Approximated line does not connect measurements, but represent expected, or theoretically ideal process. It can be used in conjunction of error areas displaying that measurement error area, and ideal process pseudo error area overlap. This allow also asses measurement set in general. It is possible to obtain a measurement that is valid from error calculation perspective but does not represent process measured. Statement might rise a question – how to interpret such results. Answer can be found in section about measurement systems in following pages. All cases are provided in Figure 2.10 below.

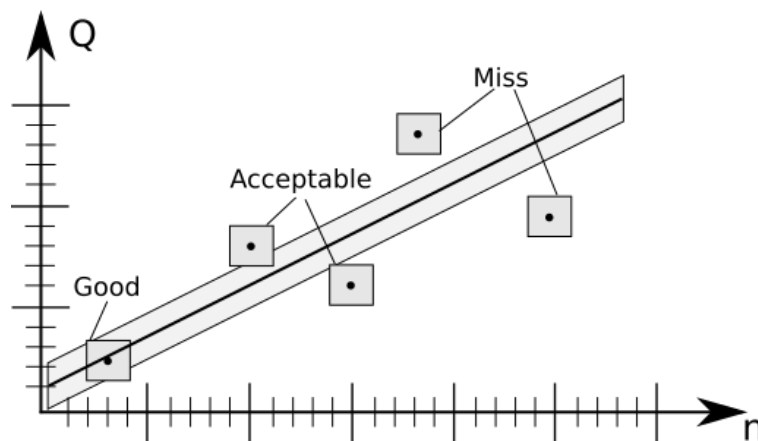


Figure 2.10. Approximated “ideal” line with actual measurement points

## 2.9 Measurement systems

### 2.9.1 Single parameter systems

Such systems are the simplest and actually covers process where attempt to obtain a true value only of one physical parameter is performed. For example – a voltage is measured in power outlet. Depending on way, how measurement is made, it is necessary to declare, is obtained value a separate measurement – singleton, or set of  $n$  measurements within time range. That's because we already know basic differences of error calculations.

One reason is accuracy, but the most important case from industry perspective, why measurements are made, is monitoring. So from this comes out need to determine time range and minimum measurement count during period of time. So actually we are starting to change paradigm when with phrase – measurement, we understand a set of value  $Q$  readings in certain period of time. As count of readings within one measurement can be large enough, distribution function is reduced to normal. All others are for cases with small measurement count. So besides of all errors and probabilities, it is common to declare additionally a time frame, when measurements are made. So this statement joins measurement representation in “ideal” form with readings. Basically, if a large amount of measurements is taken, we can easy obtain expected value (known also as knowledge base or reference value in intelligent systems) and its error range, thus every next measurement can be assessed against expectations. From here comes possibility to create a regulation mechanism, as from differences we can consider about action. Let's analyse three cases in figure with “ideal” line and measurement points. Ideal line represents an expected value for our measurement system, that is assembled to monitor certain parameter  $Q$ . When measurement is within error range of expected value, we can say that it is normal operation of the system. When measurements are outside error range, however error ranges overlap, this is clear signal that system desires closer attention as it might mean early warning about possible malfunction. Clear miss in measurements actually can signal two things – first. System is malfunctioning. Second – it is necessary to determine what part is malfunctioning. While we are talking about single parameter systems, it is impossible to determine faulty component. A multiple parameter systems are that for.

### 2.9.2 Multiple parameter systems

Main difference from single parameter system is that measurements have been made in the same moment of time. This broadens meaning of term – measurement, as in this case we talk about multiple values in conjunction. Basically multiple parameter systems do not only direct or indirect measurements, but is instrument to determine relations between different events. As in single parameter system, multiple ones have trusted ranges and errors, but as this is one, but multiple parameter system, measurement and error distribution follow to the same rule for all values. Probability that true value of measurement is within expected range is equal for all measurements. Technically this mean that it is possible to assemble self-controlling device. As we are measuring multiple parameters, that follow certain correlation rule, a clear miss in measurement of one parameter, while others are within “good” range, allow to conclude about type of malfunction and its most probable location.

### 2.9.3 Parallel systems

From metrology perspective parallel systems are completely separate as they have different error volume and distribution. Key element for parallel system usage is time. As example – a voltage and current is measured in electric circuit. Voltmeter of class 2,5 and ammeter of class 0,1 is used. While this is the same circuit, devices are different. From task perspective, it is interesting to measure both physical parameters, voltage and current, for the same moment of time. We know how current relate to voltage. So from here comes desired result - value with error and probability to be true value within section of time.

So for indirect measurements depending on desired result a single parameter, multiple parameter or parallel systems can be used.

Important aspect is scale, size of measurement system itself. While parallel systems can be very large and gather information from thousands of measurement points, multiple parameter systems are most often a single device with various inputs, so count of simultaneous readings of values is limited by measurement device construction.

## 2.10 Measurement of electric quantities

### 2.10.1 Base and derived quantities

So now about how to measure one or another quantity. All quantities are separated in two groups non-electric and electric. Such separation is defined by nature and bonds tight to today's common approach to automation, where almost everything is electric. With measurements of electrical quantities everything is more or less clear and basic. We have only one base quantity and it is electric current that is measured in amperes and have unit symbol A. All other units are derived.

List of derived quantities is:

- Electrical potential and potential difference, measured in Volts and have symbol V;
- Electromotive force is also measured in Volts and have symbol V;
- Electrical resistance, measured in Ohms and have Greek symbol ( $\Omega$ );
- Electrical impedance, measured in Ohms and have Greek symbol ( $\Omega$ );
- Electrical charge, electrical flux and amount of electricity are measured in Coulombs and have symbol C;
- Electrical capacitance, measured in Farads and have symbol F;
- Electrical inductance, measured in Henry and have symbol H;
- Electrical power, measured in Watts and have symbol W;
- Magnetic flux, measured in Weber and have symbol Wb;
- Density of magnetic flux, measured in Tesla and have symbol T.

So for all electric measurements base unit is one Ampere and global assumption is that current is 1A when during 1 second  $1.602176487 \times 10^{19}$  electrons flow through a square of a conductor. Basically  $1.602176487 \times 10^{19}$  electrons are assumed to be one Coulomb and expression of the current is

$$i(t) = \frac{dq}{dt} \quad (2.17)$$

However, in fact, this is not used as general reference as counting such large numbers of small particles is far too complicated. It is just representing phenomenon of electrical current on it's natural way as motion of electrons.

So it can be declared, that measurement of any electrical quantity is actually measurement of a current flowing in the measurement area, and this statement is compliant with SI where we have current as base value.

Another base statement is that every acquisition of a value of quantity, measurement, is performed using dedicated technical equipment.

### 2.10.2 Equipment grouping

All technical equipment used belongs to one of groups:

Gauges, measurement transducers, measurement instruments, measurement information systems, integrated measurement assembly.

- Gauges are components with in-time constant parameters within known error range and are used to provide reference value. Simple resistor can be used as a gauge in specific cases.
- Measurement transducers (MT) are tools used to convert value of the measured quantity in to another desired form that is convenient for transmission or processing. Transducers connected directly to a point of quantity value acquisition are named as primary transducers. Transducers used to transmit acquired value proportionally are named as scale transducers.
- Measurement instrument (MI) is a tool that applied directly to a point of quantity value acquisition provide sufficient and recognizable information to end user. MI that has continuous relation between acquired and displayed quantity value are named as analogue. MI that encodes acquired value in to discrete units are named as digital. MI's are also grouped by result output type. MI's used only to acquire a quantity value without recording functionality are called as displaying. MI's equipped with recording component are named as registering. MI's performing internally one or several conversions of acquired value are named as direct conversion instruments. MI's where besides direct conversion a reverse conversion is performed, are named as balanced conversion or comparison instruments.
- Measurement information system (MIS) is a set of devices orchestrated by dedicated management component and used for collection, processing and delivery to display or store of large variety of quantity values. Such systems often contain analytics and are able to interact with external parties.
- Integrated measurement assembly (IMA) mostly is a component of a large scale manufacturing line that performs data acquisition from various measurement points and is able to alter several parameters automatically. In case when inside of a plant multiple manufacturing line IMA's are governed by common external utility, they can be treated as components of MIS.

### 2.10.3 Reference gauges

Reference gauges are samples of equipment dedicated to preserve certain initial values for comparison in later steps or resetting of measurement equipment to initial state. Reference gauges usually belong to one of groups – benchmark or working. Benchmark gauges are used exclusively for measurement tool calibration purposes. Working gauges are used during measurement data acquisition. Reference gauges usually belong to one of three precision classes where class 1 is the most accurate. Depending on reference gauge type, they might be single value or multi value device. Single value reference gauges usually are resistors, inductance coils, constant capacity condensers, chemical normal elements and stabilized voltage sources.

- Reference resistors are made from materials that have significant resistance in combination with excellent thermal parameters. Usually it looks like a coil with windings from manganin and copper. Such resistors are named as wire-wound. Cheaper alternative is foil resistor. Reference resistors have tolerance class from 0.0005 to 0.1 and nominal resistance in range from  $10^{-5}\Omega$  to  $10^{10}\Omega$ . Reference resistors usually have four connectors – two are voltage connectors and are labelled as U or V and two current connectors usually labelled as I or A.

During measurements current connectors are used to attach reference resistor to schematics and voltage connectors are connected to measurement instrument. Sample of 100 Ohm reference resistor and internal wiring diagram is given in Fig. 2.11 where a) shows a one of versions how reference resistor looks, b) displays internal parts when cover is removed and c) provides internal connection diagram.

In case when reference resistors are used in alternating current schemes, internal capacitance and inductance should be taken in account.

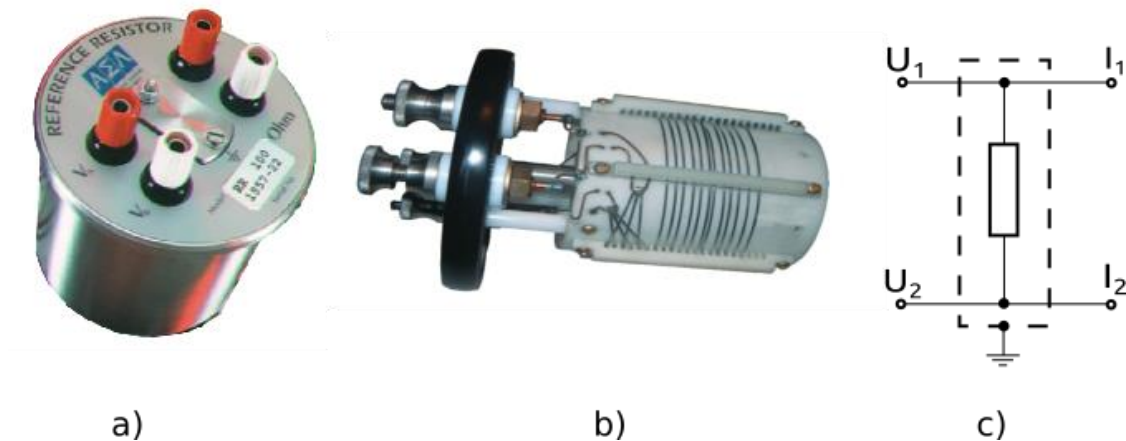


Figure 2.11. Reference resistor and internal connection diagram

- Reference inductance coils (RIC) usually are made from wire that is wound on special frame or reel. Inductance coils typically have inductance from  $10^{-6}$  to 1H. Tolerance of RIC usually is within range from 0.05 to 0.5. While inductance coils are used with alternating current, it is necessary to determine working frequency range. Another type of RIC is mutual inductance coils. It is made as two coils on common frame and mostly are manufactured with inductance of  $10^{-4}$  to  $10^{-2}$  H and working frequency 50 kHz.
- Reference capacitors are usually low capacitance devices with value up to 10nF. As insulator between electrodes in most cases air or isinglass-stone is used. For high voltage appliances reference condensers are filled with sulphur fluoride ( $\text{SF}_6$ ) gas.

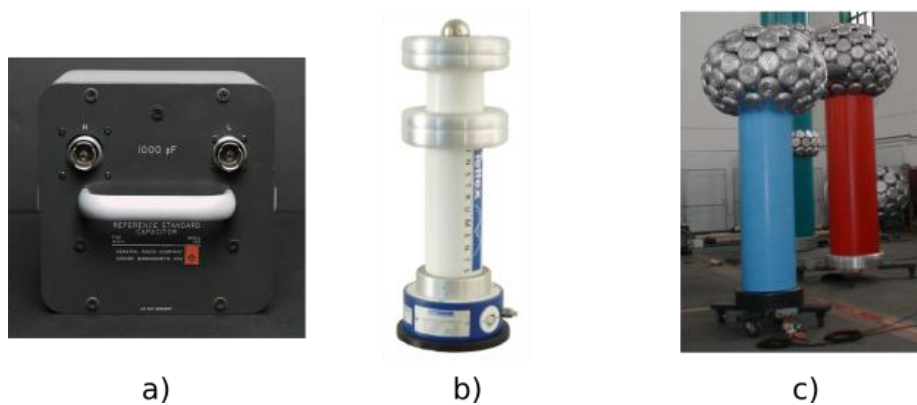


Figure 2.12 Reference capacitors



In Fig 2.12 three sample capacitors are shown where a) is 1000pF nitrogen insulated capacitor and b) and c) are high pressure SF<sub>6</sub> capacitors for 25kV and 800kV working voltages accordingly.

- Standard cells are electro chemical sources of electromotive force with well-known value and stability at reference temperature of 20°C. Standard cells have been manufactured with saturated or non-saturated electrolyte. Standard cells have tolerance class between 0.0002 and 0.02. However, they have electromotive force, due to comparable large internal resistance, standard cells cannot be used as power source or loaded with high currents. Electromotive force  $E_{20}$  of a typical standard cell is 1.018638V at reference temperature, however depending on manufacturer, it varies between 1.0185V and 1.0187V. In case, when ambient (working) temperature  $t$  is not equal to reference temperature (20°C), an expression is used to calculate electromotive force  $E_t$  for working temperature.

$$E_t = E_{20} - 40.8 * 10^{-6}(t - 20) - 0,95 * 10^{-6}(t - 20)^2 + 0.01 * 10^{-6}(t - 20)^3 \quad (2.18)$$

Today standard cells have been replaced by stabilized voltage sources however they still remain as one and only independent reference. Image of internals of standard cell is given below.



Figure 2.13. Standard cell also known as Weston cell

- Stabilised voltage sources are complex electronic devices made to provide electromotive force at desired voltage with extremely high precision. For example, 1.5000±0.0001V. The same as standard cell stabilised voltage source is not a power supply so for normal operation it demands low draining current, not more than 1mA and mains voltage ±10% of nominal.

Multi value reference devices are able to provide more than one reference value. Typically for one physical phenomenon. Multi value reference devices are measurement generators, different kinds of special calibrators, variable capacity condensers, variable inductance variometers, resistance, capacitance and inductance boxes. Typically, multi value reference devices have worse accuracy than single value devices.

- Measurement generators, named also as arbitrary / function generators, are sources of alternating current with known generated signal form where amplitude, frequency and other parameters can be altered within predefined range. Depending on signal form generated measurement generators are grouped in sine wave generators, determined spectrum noise generators, pulse generators, special form signal generators. Measurement generators mostly are used as working reference devices and should be calibrated against single value benchmark. Operational ranges are starting from  $10^{-3}\text{Hz}$  and does not exceed  $10^{10}\text{Hz}$ . Depending on signal generation methodology for full analogue generators frequency is  $\pm(0.1 - 3)\%$  and amplitude  $\pm(1-6)\%$  of range. Digital generators can provide signals with very high precision  $\pm 0.0001\%$  as they rely on high accuracy time domain generators – quartz or even caesium clocks, but depending on schematics used inside, at some ranges they might have wider signal spectrum due to digital to analogue conversion. Refer to manufacturer manuals before use of such device to determine precision for desired working range.



Figure 2.14. Measurement generator

- Measurement calibrators are sources of signal with stabilised voltage and / or current parameters. Calibrators like measurement generators are used mostly as working references and should be calibrated against single parameter benchmark before use. Calibrators are able to provide predefined waveforms or direct current with predefined parameters. Such devices are widely used in automated testing systems and when supplied with control unit are able to cover multiple checkpoints. Measurement calibrators in general are able to provide precise voltage from  $1\mu\text{V}$  to  $1\text{kV}$ . Calibrators should not be loaded and used as power sources. Operational range for current typically varies from  $1\mu\text{A}$  to  $100\text{mA}$ . For actual values refer to manufacturer manuals. Accuracy of a calibrator is usually around  $\pm 5 \cdot 10^{-3}\%$  of operation range.

Frequency calibrators are usually combined with voltage calibrators, however does not have so broad voltage and frequency range than measurement generators.

Phase calibrator provides two signal channels with shift in phase from  $0^\circ$  to  $359.9^\circ$  where last digit is determined by sampling step (in case mentioned sampling step is  $0.1^\circ$ ). Depending on technology used, phase calibrator has an internal comparably narrow range ( $1\text{Hz} - 20\text{kHz}$ ) measurement generator or can be paired with external measurement generator to cover broader range of signals and frequencies.

Variable capacity condenser usually is an air / gas insulated device with working range from 0 to 50  $\mu\text{F}$  and accuracy  $\pm 0.2\%$  of selected range.

- Resistance, inductance and capacitance boxes are widely used as multi value reference device that permit to select desired value using rotary switches, one per decade. Such boxes might have extremely wide working range – smallest and biggest range difference can be up to  $10^{12}$  however usually it does not exceed  $10^6$ . Such boxes are mainly used as working reference with precision class 0.005 to 1. When using this type of equipment user should be aware, that decades used inside might have various precision classes. According to precision class of the decade,  $\gamma$  error value can be calculated as

$$\delta_n = \pm \gamma \frac{N_{max}}{N} \quad (2.19)$$

where  $N_{max}$  is maximum value for range and  $N$  is value selected. This basically means that in case, when box is used as reference and all decades should be engaged, actual working value must be verified against benchmark because error value in greater decade can be greater than adjustment introduced by smaller decade.

For example – resistance box with six selectors and working range from 1  $\Omega$  to 999999  $\Omega$  need to be set up for reference value of 985135  $\Omega$ . Due to errors in bigger decades, selectors on box display different value, 985128 when desired is displayed on MI. So absolute error in this case is 7  $\Omega$  only that for box with precision class 0.005 is very good.

Sample images of combined resistance+capacitance and inductance boxes are given in Fig.2.15.



Figure 2.15. Resistance, capacitance and inductance boxes

#### 2.10.4 Measurement transducers

Measurement transducers are separated in two groups – mechanical transducers (MT) or electrical transducers (ET). This describes the way how measurand is altered for future processing. Transducers provide useable output as response on quantity, property or condition in it's input. Transducers depending on their construction are forwarding and / or converting. Transducer inputs are sensitive to mechanical, electrical, magnetic, optical, chemical, acoustic, thermal and nuclear phenomenons or their combinations. Transducer as a device can contain

multiple components, that can be treated as transducers separately so characteristics of it might be assembled from characteristics of components.

- Mechanical transducers are devices that uses non-electric components to forward or convert measurand Typically, MT are rugged, cheap, accurate, linear within working range and does not demand external power source to operate. All this allow to build various MI's and use them where acquisition of information for automated future processing is not crucial. MT might be an essential part of a larger measurement system or hybrid transducer where cooperate with another transducer. Unfortunately, MT have serious disadvantages like friction, narrow operational frequency range and in some cases comparably high energy drain (interference) from measurand. This does not comply to main requirement for any MI, that demand to affect measurand as less as possible. All this narrows set of cases where generic MT can be applied.
- Electrical transducers are used to convert any non-electric quantity in to electrical quantity. ET have another wide used common name – sensor. Depending of phenomenon used ET converts quantity on its input to electromotive force, resistance, inductance, capacitance or any other that can be measured conveniently. However, sensors in most cases are basic ET, in some specific cases they can be hybrid, where measurand is first transformed by MT in to quantity that allow to apply precise ET instead of direct conversion with another, inaccurate method. As for MT also ET have definite reproducible function that describe relationship between input and output quantities. So output to input in known time domain is predictable at determined degree of accuracy regarding to environmental conditions.

## 2.11 Measurement instruments

### 2.11.1 Equivalent schematics of measurement instruments

Regardless of construction, every MI used to measure electric quantities has impedance. Impedance comes from internal resistance, inductance and capacitance of the circuit. Graphical diagram is given below.

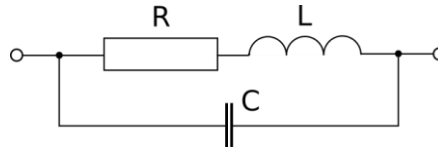


Figure 2.16. Equivalent schematic of an MI

Equivalent impedance of an MI according to diagram can be calculated as

$$Z = Z_R + Z_L + Z_C = R + j\omega L - j\frac{1}{\omega C} \quad (2.20)$$

From here it is clearly visible that frequency of alternating current is important and thus MI should be carefully selected prior use according to expected frequency. Values of equivalent schematics and maximum working frequency is usually provided by manufacturer of the device.

### 2.11.2 Measurement instruments interference

Most important thing to keep in mind is that MI is not allowed to interfere with parameter it is measuring. This statement actually is quite hard to achieve as all have impedance and there is current needed to flow through MI. So this is a thing to take in account when measuring electrical quantities inside of a system. So for current measurements impedance of MI should be significantly smaller than measured system (MS) has.

$$Z_{MI}^I \ll Z_{MS}^I \quad (2.21)$$

For voltage measurements it is opposite. For voltmeters internal impedance must be high as possible so:

$$Z_{MI}^U \gg Z_{MS}^U \quad (2.22)$$

How much should be the difference to consider impact of MI as not considerable? Answer is simple – MI should not alter any significant digit. So from here it is possible to declare that impact of MI to measurement should not exceed 1/10 of last significant digit value. For example, if a measurement of voltage is performed and smallest significant digit is second, mean 10V, then impact of MI should not exceed 1V.

### 2.11.3 Magneto-electric instrument

There are multiple methods how to acquire value of a current. As we know, around every conductor with current exists a magnetic field. If this conductor is located in the magnetic field created by another source, let it be a permanent magnet, both fields start to interact. This

phenomenon allows to assemble an electro-mechanic ammeter that depending on components used has a determined measurement range, sensitivity and accuracy. This is one of the most basic constructions that have name magneto-electric instrument with moving frame. It belongs to electro-mechanic group of MT.

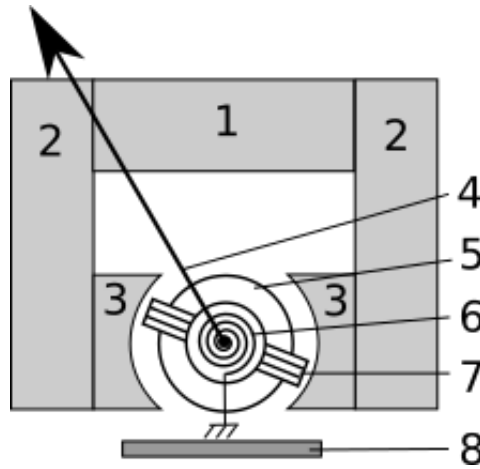


Figure 2.17. Diagram of a magneto-electric instrument

Construction of a magneto electric instrument contains: 1 - permanent magnet, 2 - magnetic flux guides, 3 - magnetic pole endings, 4 – dart, 5 - fixed base, 6 – spring, 7 – coil, 8 - magnetic shunt.

For operation this type of MI demand a current flowing through a coil. This create a momentary torque  $M_t$  inside of a system. Basically torque comes from interaction of two magnetic fields, one created by permanent magnet and appear to be constant, and second – created by current in the coil of the MI. Energy  $W_M$  of magnetic field in the gap, that create torque, depends on linkage between fields  $\psi$  and electric current  $i$ , where linkage depends on magnetic induction  $B$  in the gap between pole endings and coil frame, coil square  $s$ , count of windings in the coil  $w$ , and magnetic flux direction angle  $\alpha$ .

$W_M = \Psi_i, \Psi = Bsw\alpha$  thus momentary torque is

$$M_t = \frac{dW_m}{d\alpha} = Bswi \quad (2.23)$$

In case, when current in the coil is sinusoidally alternating, torque obtains form

$$M_t = BswI_m \sin\omega t \quad (2.24)$$

While mass of the coil and spring of the moving part is creating counterforce, mechanical system has certain self-oscillation frequency. Typically, this frequency is around 1Hz and characterize time needed for needle to reach maximal working position (other end of scale). So from here it is clear that for frequencies over self-oscillation such MI is unable to display correct value and can be used only for direct current measurements or it will follow slow oscillations displaying momentary value.

Every spring used in MI to return needle to zero position will create a counter torque  $M_c$  that depend on energy accumulated in spring  $W$  and turning angle  $\alpha$  :

$$M_c = -W\alpha. \quad (2.25)$$

From here it is possible to express sensitivity to current of a magneto-electric instrument as

$$S_I = \frac{B_{sw}}{W} \text{ and turning angle for direct current } I \text{ equals to } \alpha = S_I I.$$

This means that turning angle is dependent on polarity, current direction.

However, this type of MI is hard to apply directly to intelligent system, it is very simple and reliable device without need of external power source. It can be used for measurements in cases when other measurement methods are under suspect of providing incorrect data or being tampered. Even more, magneto-electric instrument as shown above can be monitored by optical recognition device and provide desired information over distance.

For example, of an application let's take magneto electric MI and instead of a needle put a mirror to the moving part. From low power laser 3 beam 4 is targeted to mirror of MI 1 and reflected to optical sensor array 2 graduated accordingly.

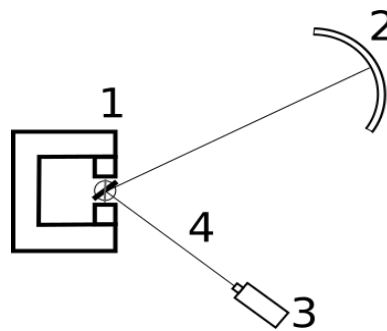




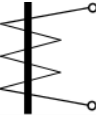
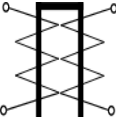
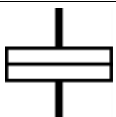
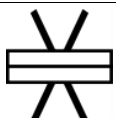
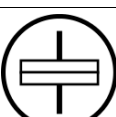



Figure 2.18. Optical magneto-electric MI



Because of comparably large distance between mirror and sensor array, small motion of the frame gives noticeable deflection of laser “needle”. Such composition allows to measure electrical values in place where extremely high sensitivity is needed and by some reason, like ionizing radiation, other types of transducers cannot be used.

Magneto-electric instrument described here is the most basic type. All other instruments differ only by magnetic system – mean count of coils and sources of interacting force. To describe type of MI all of them usually have symbolic description printed on front plate. List of symbols is given in table below.

	Magneto – electric instrument with moving frame
---	---

	Magneto – electric logometer
	Magneto – electric instrument with moving permanent magnet
	Magneto - electric logometer with moving permanent magnet
	Electro – magnetic MI
	Electro – magnetic logometer
	Electro – dynamic MI
	Ferro – dynamic MI
	Electro – dynamic logometer
	Ferro – dynamic logometer



	Inductive MI
	Electro – static MI

#### 2.11.4 Electronic MI

Electronic MI's (EMI) are separate class that was created to avoid drawbacks of basic magneto-electric MI like low impedance and following high energy drain from measurand. EMI's are fully analogue devices with external power source needed to operate and set of internal schematics to broaden it's application. As common case, high accuracy magneto – electric instrument is used to display values so name – electronic voltmeter is widely used.

If basic MI is applied directly, then EMI have to convert measurand. To obtain high input impedance (typically over 1MW), EMI utilize direct conversion transducers that introduce their own error. Such instruments usually have precision class not better than 1. To achieve lower error rates, EMI can be equipped with more complicated and harder to use balanced equivalent transducers but have precision class from 0.2 to 2.5.

Typical simplified structure of EMI is given in Fig. 2.19.

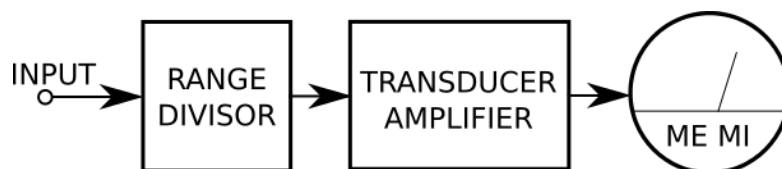


Figure 2.19. Structure of electronic measurement instrument

Range Divisor is simple voltage divider that feeds direct current amplifier with a determined fraction of input signal. Already well known magneto-electric MI is used to display obtained result. In general, magneto electric instrument used here for display is generic and range less so result displayed is actually an angle of dart deflection. As this is typical magneto-electric MI, deflection angle  $\alpha$  is

$$\alpha = k_{RD} k_{TA} S_{MI} U_{IN} = k_V U_{IN}, \quad (2.25)$$

where  $k_{RD}$  is conversion coefficient of range divisor,  $k_{TA}$  is conversion coefficient of transducer amplifier,  $S_{MI}$  is sensitivity of instrument itself and  $U_{IN}$  is voltage at input. Coefficients of all components actually are joinable in one parameter – conversion coefficient of voltmeter  $k_V$ .

Sequential structure of EMI allow to build devices with extremely high sensitivity and very broad operation range by varying conversion coefficient  $k_V$ , but this approach has also drawbacks. Sensitivity of instrument typically depend on amplification coefficient, but due to imperfection, amplifier start to voluntary adjust signal as  $k_{TA}$  varies in time. Such phenomenon

is called as “zero drift”. To avoid this phenomenon, for high accuracy EMI amplification coefficient  $k_{TA}$  is around 1, mean – no amplification, so main task of an amplifier in such cases is to provide high input impedance.

To broaden operation range in low voltage area, mean, enhance it from 1mV to 1mV, instruments might have “zero adjustment” capability to set  $k_{TA}$  at desired level right before measurement session.

Another solution to improve EMI sensitivity and usability for extremely low voltages is to use dedicated modulator and demodulator technique in transducer. Voltage in electronic MI input is fed to modulator M driven by frequency generator G (Fig.2.20). Modulated signal is amplified by alternating current amplifier A, that blocks direct current component of signal thus avoiding “zero drift” phenomenon. Demodulator DM afterwards convert amplified signal in to sequence of impulses with proportional amplitude representing input signal. Important thing is that modulator and demodulator, steerable rectifier, is driven by the same generator so they work synchronously.

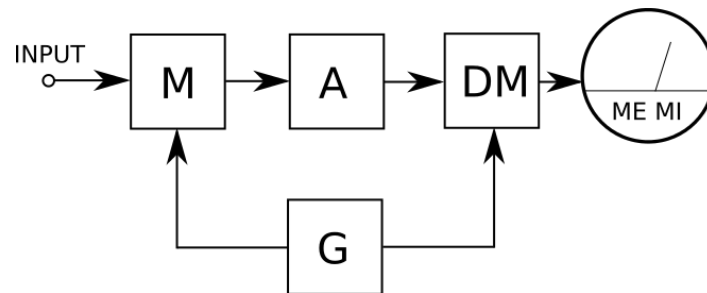


Figure 2.20. Structure of modulator – demodulator of electronic MI

Average voltage at exit of demodulator is proportional to voltage at input so

$$U_{avg} = kU_{IN}. \quad (2.26)$$

As system is not affected by “zero drift” phenomenon, amplification coefficient  $k$  can be set at quite high values  $k=1 \cdot 10^5$  or even more. This allow to have electronic voltmeter capable to measure voltages starting from 1mV with error  $\pm 0.5 - 6\%$  of selected range.

By default all electronic voltmeters are suitable for direct current only, so capability to measure any other form of signal is achieved by addition of the transducer. If sine waveform alternating current can be converted in to suitable direct current by simple rectifier, then for other cases additional schematics is needed.

Depending on schematics used, MI for alternating current can be set to display amplitude, average value or actual value.

- Voltmeters capable to display **amplitude value** have scheme called peak detector. Duty of such scheme is to keep maximum amplitude value at output so observer is able to read information from MI. Generic scheme of peak detector and time / voltage diagrams are given below.

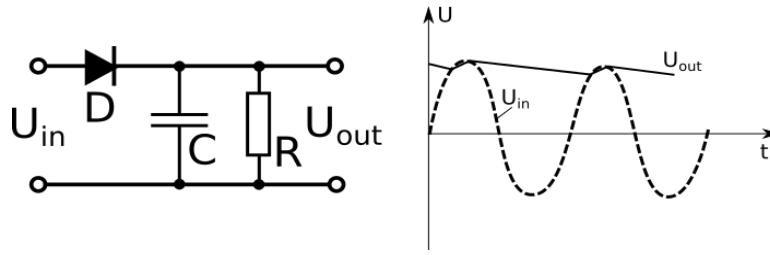


Figure 2.21. Generic peak detector and voltage diagram

As it is possible to see from time – voltage diagram, voltage at output is still alternating at some level but significantly less than it is in input, so amplitude value can be read within certain error range. Capacitor C and drain resistor R should be taken so, that discharge time is much bigger than time between two peaks of AC. So this approach sets certain working frequency range. In the same time large C and large R will increase inertia of the instrument and reduce its capability to follow changes of the peak values.

Peak detector shown in Fig. 2.21 is also named as open. As it allows the DC component  $U_0$  of the signal to flow in, if its direction matches rectifier diode D. In such cases input voltage  $U_{in}$  is

$$U_{in} = U_0 + U_m \sin \omega t \quad (2.27)$$

Due to rectifier, in cases when  $U_0 < 0$ , DC component will be ignored. This aspect should be taken in account in measurement planning stage when expected values are declared.

To avoid any undesired effects of DC component, a “closed” type peak detector can be used. There are various generic schemes, how this is achieved but they all lead to desired goal – deflection angle of measurement instrument needle  $\alpha$  is proportional to amplitude at the input,  $\alpha = k_v U_m$ .

Another thing to take in account is a fact that scale of the instrument is graduated explicitly for sinusoidal current. So if form of the measured signal is not precise sine wave (within error range), then value read should be adjusted using coefficient  $k_a$ . So if value displayed is  $U_d$ , then  $U_m$  is

$$U_m = k_{sin} U_d = \sqrt{2} U_d. \quad (2.28)$$

Value  $k_{sin}$  here is amplitude coefficient for sine form voltage. So for non-sine form signal value will be

$$U = \frac{\sqrt{2} U_d}{k_a}. \quad (2.29)$$

- **Average value** voltmeters contain current rectifiers where amplified signal is fed. This allow to increase sensitivity of an instrument and reduce non linear influences of rectifier. So deflection angle  $\alpha$  of the needle is proportional to average value  $U_{x\,avg}$  of rectified current caused by voltage  $U_x$

$$\alpha = k_v \frac{1}{T} \int_0^T |u_x(t)| dt = k_v U_{x\,avg}. \quad (2.30)$$

Scale of such instruments are also graduated for sinusoidal waveform signals so in case, when

non sinusoidal signals are measured, true average value  $U$  is obtained by adjustment of value displayed.

$$U = \frac{k_s U_d}{k_{sin}}, \quad (2.31)$$

where  $k_s$  is coefficient of non-sine signal form and  $k_{sin}$  is coefficient of sine signal.

- **Actual value** voltmeters are built on alternating voltage converter that strictly follow the rule

$$U_{out} = k U_{in}^2. \quad (2.32)$$

Schematics of such converter can vary depending on instrument manufacturer but for all of them deflection angle of MI needle is proportional to square of  $U_{in}$  - actual value of voltage applied:

$$\alpha = k_v \frac{1}{T} \int_0^T u_{in}^2(t) dt = k_v U_{in}^2. \quad (2.33)$$

So it is clearly visible that such instrument have “squared” scale and to reduce inconvenience of this, for voltmeters with linear scale, additional converter is engaged. Additional converter is used in negative feedback circuit.

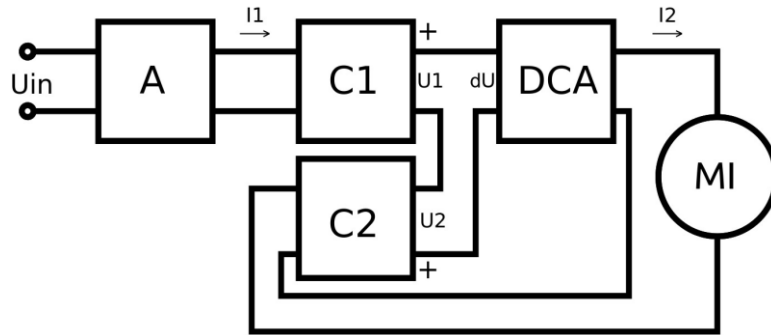


Figure 2.22. Dual square converter actual value voltmeter structure

As shown in Fig. 2.22, voltage  $U_{in}$  applied to input is amplified by alternating current amplifier A and create current  $I_1$  in the first square converter  $C_1$  and converted voltage  $U_1$  in exit. Current  $I_2$  at exit of direct current amplifier (DCA) caused by difference of voltages on exits of converters and is fed back to second converter. Both converters  $C_1$  and  $C_2$  have their own conversion coefficients  $k_1$  and  $k_2$  so voltages at converter exit are:

$$U_1 = k_1 I_1^2; U_2 = k_2 I_2^2. \quad (2.34)$$

While current  $I_1$  from amplifier A is proportional to voltage applied, then it can be expressed as

$$I_1 = k_a U_{in}. \quad (2.35)$$

So voltage  $U_1$  at the exit of first square converter ( $C_1$ ) is

$$U_1 = k_1 k_a^2 U_{in}^2 \quad (2.36)$$

While amplification coefficient of direct current amplifier is high enough, it leads to state when signal  $dU$  at DCA input is  $dU = U_1 - U_2 \approx 0$ . So if  $U_1 \approx U_2$ , then

$k_1 k_a^2 U_{in}^2 = k_2 I_2^2$  and deflection angle of the instrument needle  $\alpha$  is

$$\alpha = S_{MI} I_2 = S_{MI} k_a \sqrt{\frac{k_1}{k_2}} U_{in} = k_v U_{in}. \quad (2.37)$$

So deflection angle of the needle  $\alpha$  is proportional to voltage applied at the input.

Such actual value voltmeters typically have working voltage range starting from several millivolts and frequency from 1Hz to 1GHz. Depending on working range selected and MI used accuracy of measurement can vary in range  $\pm(0.5 \text{ to } 25)\%$ .

Beside of voltmeters explained above, in production variety of specific instruments can be met. They might have various converters inside so allowing to measure large variety of signals with known accuracy. Tools containing a set of additional schematics, usually are called as **universal voltmeters**. User using rotary switch select desired work mode and perform measurement. Typically universal voltmeters are good enough for general daily use but do not offer exceptional precision and working range. To measure voltage of non periodic or short signals, dedicated **pulse voltmeters** are used. This type of tool contain a schematic capable to capture and lock peak value of a short impulse following after comparably long gap. Working range of such voltmeters seldom exceed 20V however depending on schematics used, they are capable to capture impulses up to only 10ns wide. Time of the impulse  $t$  and gap  $T$  is declaring a time domain so for sequential impulses working frequency is

$$f = \frac{1}{t+T}. \quad (2.38)$$

Usually working frequency start from 0.1Hz and goes up to 1GHz.

For pulse voltmeters one of crucial things is compatibility to spectre of the signal. As we know, impulses are assembly of sinusoidal signals. Shorter impulses contain broader spectrum, so in cases when short-pulse signal is applied and have frequency close to maximum of the instrument, signal building spectrum exceed capabilities of the tool so error level is greater than at low frequency. This mean that voltage shown is lower than actual value.

Another specific tool is **selective voltmeter** that have schematic capable to chop out from input signal actual value of specific component. Mean, they are working in comparably narrow frequency range that can be set. Such voltmeters are used to measure “strength” of signal filtered out from sequence of pulses or noise where signal to noise ratio is poor and can not be measured by general use universal voltmeter. Typical sphere of application for such instrument is measurement of signals carried over by radio. Depending of schematics used, selective voltmeters have working voltages from several microvolts and does not exceed 10V. Error range lies between  $\pm(5 \text{ to } 15)\%$  for frequencies between 20Hz and 2MHz.

#### 2.11.5 Digital instrument

Another type of MI's is digital one. All digital MI's (DMI) have one general attribute – their impedance is very high. Multiple times higher than best magneto-electric MI has. DMI's does not demand current to display a value. Besides fact, that all digital MI's demand an external source of power to operate, they do not display direct value of quantity measured, but an approximated equivalent.

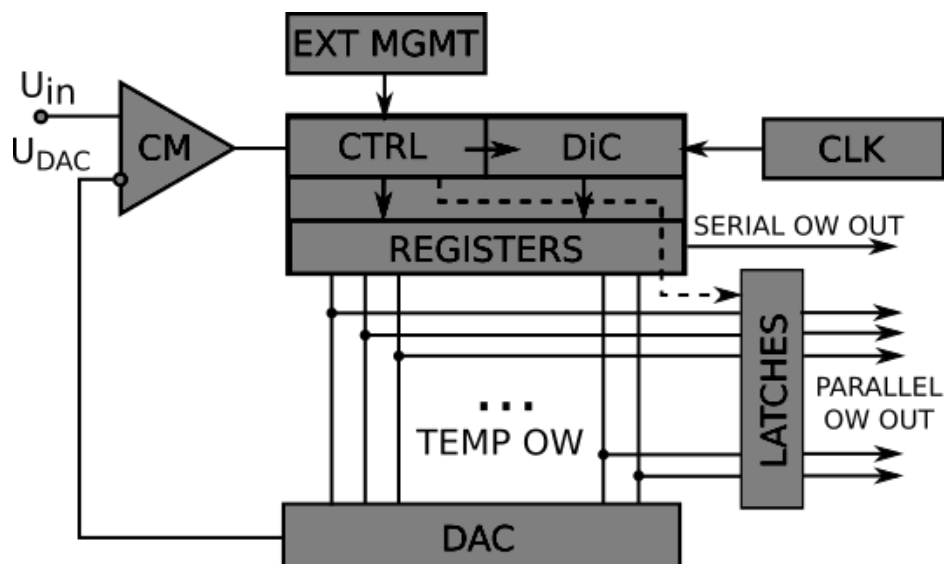
For operation DMI contain two main components an analog-digital converter (ADC) and digital counter (DiC). While ADC by design is able to operate with only one range and type of signals, usually DMI contain a variety of completely analogue transducers functionally similar to those, that are described before in section about EMI. Their duty is to convert input signal proportionally to working range of ADC. So it is ADC duty then to achieve that analogue signal at input is represented as digital (binary) code at exit. In some literature this binary code is named also as operational word (OW). So after acquisition of OW it is possible to process it for display or transmission.

Functionally, all ADC's treat input signal as explicitly DC so if alternating signal need to be measured, it should be converted into DC in the same way it is for EMI.

Today depending on application two types of ADC are most widespread however there are more on field with their pros and cons.

- Successive approximation converter

This type of converter utilize count-conversion methodology. By command using DiC a generation of temporary OW is started. Speed of temporary OW generation depend on clock CLK working frequency. Digital value is converted by digital-analog converter DAC and fed to comparator CM. Comparator will return to control schematics CTRL a DiC stop signal when  $U_{DAC}$  at DAC exit is equal or greater than  $U_{in}$ . Stop signal serves as flag – value obtained, so temporary OW is valid for transmission. OW will be held on latches until next OW is obtained. DiC reset and start signal depending on systems design can appear periodically in control schematics or upon request from external management. Schematics of successive



approximation converter is given below.

Figure 2.23. Structure of successive approximation A/D converter

- Parallel (flash) converter

Parallel converter does not use counter but calibrated resistor array and reference voltage for operation. Signal applied at converters input is fed to every comparator. Comparator returns “true” value if signal at input is over it's reference value, and dedicated encoder convert it in to OW. However parallel converter is unbeatable leader in conversion speed (20 million conversions per second possible), it has significant drawback in perspective of OW size that usually does not exceed 8 bits and has maximum of 11 due to it's construction. Basically every additional bit doubles the count of comparators and resistors needed in reference array.

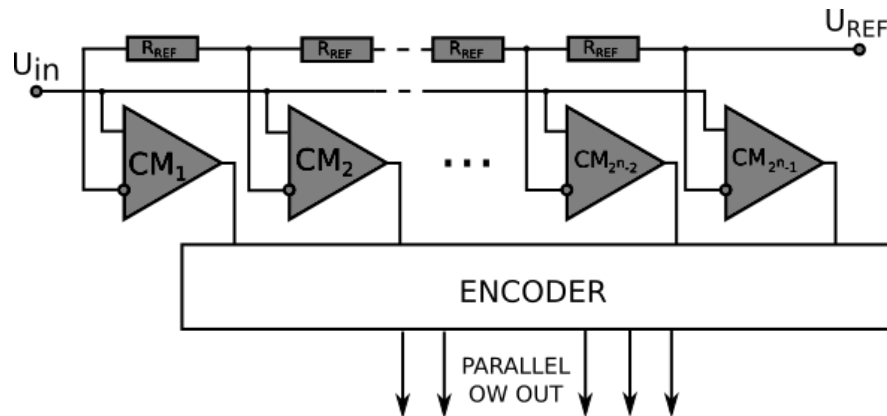
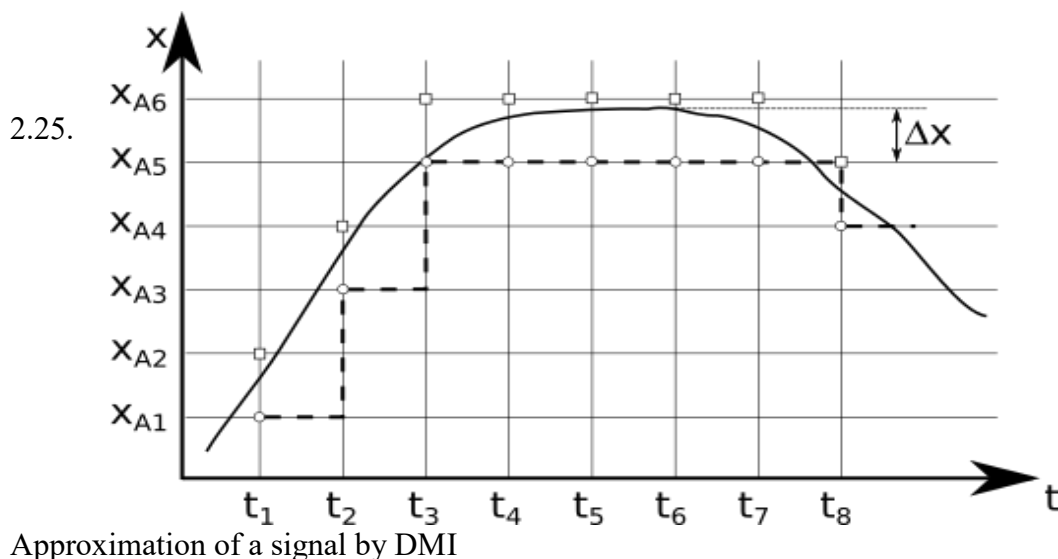


Figure 2.24. Structure of parallel A/D converter

Regardless of ADC construction, all DMI show value of measurand with error in last digit.

As way how value is delivered eliminate reading error, it still contain errors introduced by conversion. This mean that value shown can be lower than true value by one level of approximation. Simplified visual interpretation of such conversion is given below.



Figure

Let's take a random freeform signal  $x$  and approximate it in the way DMI does. Scale of image does not represent any specific DMI but is drawn large to display principles. So at moment of time  $t_1$  analogue to discrete signal conversion is performed. Depending on ADC used, level  $x_{A1}$

will be declared as measured value for parallel converter (round dots) or  $x_{A2}$  for DiC based converters. So both cases have the same style of approximation but value differs by one step. As it is visible in fig. 2.25 for moment of time,  $t_6$  for example approximation always introduce error  $\Delta x$  between true signal value and assumption of DMI.

In fact approximation error  $\Delta x$  does not play significant role in modern equipment where OW size is 16 or 24 bits and DiC based ADC is able to perform conversion in less than 1ms. Bigger impact on DMI is caused by noise that comes with measured signal. Due to DMI's high input impedance and selectivity of ADC, noise make high precision measurements problematic and shift position of last significant digit to one, that is not changing in certain period of time. For example, observer perform voltage measurement with tool able to display measured value with 10 digits. Due to DMI construction, last 5 digits (positions from 1 to 5) change rapidly and are almost unreadable. Value at position 6 does not change so it can be considered as last significant digit. Basically this mean that signal to noise ratio determine last significant digit and is tied to range selected.

For ADC's with DiC two approaches for OW generation is used:

- Direct count method expect counter to increase value of OW by one every clock cycle. This method is used in cheap DMI's with slow response time and it's main advantage is simplicity.
- Floating “one” method means setting a value 1 in OW oldest (biggest) bit and comparison of result. If  $U_{DAC}$  is over  $U_{in}$ , bit value is set to 0 and next bit is populated by 1. Value of position remain 1 if  $U_{DAC}$  appear to be below  $U_{in}$ . However method is more complicated than direct count, it is much faster as it need fixed count of clock cycles to achieve result. Technically with clock running at 32MHz it takes  $0.75\mu s$  to fill register and obtain OW.

However DMI's parameter - time to obtain OW ( $t_{ow}$ ) and provide measurement value might sound as not very important, it can be considered as true only in case when observer is human and a static value is desired. Today this appear to be only one specific case of many. While DMI is capable to provide a large array of values thus showing dynamics of measured process, observer is able to do more accurate decisions. In complex systems DMI usually is a key component that provide automated management system, acting as observer, with data about system state. So DMI is a true footstone of every Intelligent System (IS), component that adds intelligence to the system. So if time to make decision should be kept as short as possible, speed of DMI is extremely crucial.

Another key factor of every MI, and DMI is not an exclusion, is maximum working frequency ( $f$ ). It is directly related to ADC sampling rate. According to Nyquist-Shannon sampling theorem (know also as Kotelnikov's theorem), highest frequency of spectrum in signal applied as  $U_{in}$ , that is taken in account, is half of OW generation frequency

$$f = \frac{1}{2t_{ow}}. \quad (2.39)$$

This aspect, in fact, dramatically reduce maximum working frequency of DMI where direct count DiC is used.



## 2.12 Bibliography and list of acronyms

1. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM). 3<sup>rd</sup> edition. 127 pages.

2. <http://www.bipm.org/en/publications/si-brochure/>

List of acronyms:

BIPM - International Bureau of Weights and Measures

CCQM - Consultative Committee for Amount of Substance – Metrology in Chemistry

CGPM - General Conference on Weights and Measures

CODATA - Committee on Data for Science and Technology

GUM - Guide to the Expression of Uncertainty in Measurement

IAEA - International Atomic Energy Agency

ICSU - International Council of Scientific Unions

IEC - International Electrotechnical Commission

IFCC - International Federation of Clinical Chemistry and Laboratory Medicine

ILAC - International Laboratory Accreditation Cooperation

ISO - International Organization for Standardization

IUPAC - International Union of Pure and Applied Chemistry

IUPAC-CIAAW - International Union of Pure and Applied Chemistry – Commission on Isotopic Abundances and Atomic Weights

IUPAP - International Union of Pure and Applied Physics

JCGM - Joint Committee for Guides in Metrology

JCGM/WG 1 - Working Group 1 of Joint Committee for Guides in Metrology

JCGM/WG 2 - Working Group 2 of Joint Committee for Guides in Metrology

OIML - International Organization of Legal Metrology

VIM - International Vocabulary of Metrology -- Basic and General Concepts and Associated Terms

WHO - World Health Organization

MT – Measurement Transducer

MI – Measurement Instrument

### **3 ELECTRICAL ENGINEERING**

## 3.1 DC circuits

### 3.1.1 Elements of DC circuits

An elementary DC electrical circuit consists of a DC source of electrical energy, load – a consumer of the energy, connecting wires, terminals with the help of which the wires are connected to the source as well as of commutating (like switch) and protecting (e.g. fuses) devices. The DC source of electrical energy consists of a unidirectional source of EMF  $E$  and its internal resistance  $R_i$  (Fig.3.1). One pole of the source has a constant positive polarity but the other – constant negative. The DC source of EMF could be provided by chemical accumulators, photovoltaics, fuel-cells, etc. Electrical current is a constant movement of charged particles through the close (with switch  $S$  in ON condition) circuit directed from the positive pole of the source of EMF to the negative one. In respect to the load the current is a result of an action of the potential difference across the load which is called voltage  $U$ . EMF and voltage are measured in volts [V], electric current – in amperes [A]. For DC circuits all electrical values are marked with the use of uppercase letters  $I$ ,  $U$ , etc.

Because of the inside structure of the conductors and other elements any electrical circuit and the EMF source itself resists to the movement of the current; this phenomenon is called an electrical resistance ( $R$ ). The unit of measurement of the electrical resistance is Ohm [ $\Omega$ ]. The resistance of the conductor depends on its material, its length as well as cross-section area. The resistance of the load is calculated for gaining a certain job on the load by the action of the current and this resistance is much higher than that of the energy source and the wires.

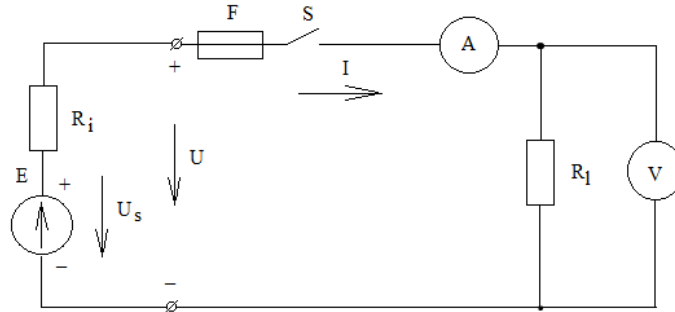


Figure 3.1. A Simple DC electrical circuit

Resistivity of a conductor is used to evaluate electrical features of material. Resistivity is different for various materials and these data could be found in special handbooks. Resistivity is designated as  $\rho$ , measured in [ $\Omega \cdot m$ ] and has the following correlation with the resistance of the whole conductor:

$$R = \frac{\rho \cdot l}{s}, \quad (3.1)$$

where  $l$  – is the length of the conductor, m;  $s$  – cross-section area of the conductor,  $m^2$ .

All materials are divided into 3 groups:

- 1 – conductors of electrical current, which have small level of resistivity  $\rho$ ,
- 2 – insulators with high level of  $\rho$ ,

3 – semiconductors which could have different resistivity depending on specific conditions.

For the most useful conduction materials  $\rho$  has the following meanings:

- for copper wires  $\rho=0.0175 \cdot 10^{-6} \Omega \cdot \text{m}$ ,
- for aluminum wires  $\rho=0.027 \cdot 10^{-6} \Omega \cdot \text{m}$ ,
- for steel wires  $\rho=0.1 \cdot 10^{-6} \Omega \cdot \text{m}$ .

The material usually used for the wires is steel or copper; the resistance of the wires ordinary is much lower than that of the load then the voltage drop across the wires is very low and therefore can be neglected at calculations.

### 3.1.2 Ohm's Law

The correlation of EMF (voltage), resistance and current in the given circuit is described by so called Ohm's Law: the current in the circuit is directly proportional to the EMF (voltage) and reversibly proportional to the whole resistance of the circuit:

$$I = \frac{E}{R_i + R_l} = \frac{U_s}{R_i + R_l}. \quad (3.2)$$

Ohm's law is valid not only for the whole circuit but also for any of its part. At any part of the electric circuit the voltage is supplied in the direction from the point of higher potential to the point of lower potential, i.e. across the load the voltage is directed from the positive terminal of the source to the negative and calculated as

$$U = I \cdot R_l, \quad (3.3)$$

where  $U$  – is a difference of potentials in volts [V]. Ohm's law for a part of electrical circuit could be formulated as follows: current through a part of an electrical circuit is equal to the voltage, supplied to the terminals of this part divided by its resistance.

### 3.1.3 Theorem of the voltage source (Thevenin's theorem)

This theorem describes features of a source of voltage containing EMF and internal resistance in series. The theorem states that any network with respect to a pair of terminals may be replaced by a voltage source (equal to the open-circuit voltage between the terminals) in connected in series with a resistance. If the switch  $S$  in Fig.3.1. is in turn off condition the current is equal to zero and the voltage supplied to the terminals corresponds to EMF ( $U_s=E$ ). This is so called open-circuit condition. If the terminals of the source are connected shortly by a wire, a short connected circuit takes place, then

$$I_{sh} = E / R_i. \quad (3.4)$$

The internal resistance of the source could be determined from this (2.4) expression. Nevertheless the current here is very high. That is the reason why this condition of the circuit is not used, so the circuit is usually loaded with an external load.

In the case of external loading the voltage between the terminals is  $U = I \cdot R_l$ . The voltage supplied to the internal resistance of the source is:

$$U_{ri} = I \cdot R_i = E - U, \quad (3.5)$$

Therefore the internal resistance of the source is equal to  $R_i = (E - U)/I$ .

The relation between  $U$  and  $I$  is called a loading characteristic of the source and it is shown in Fig.3.2.

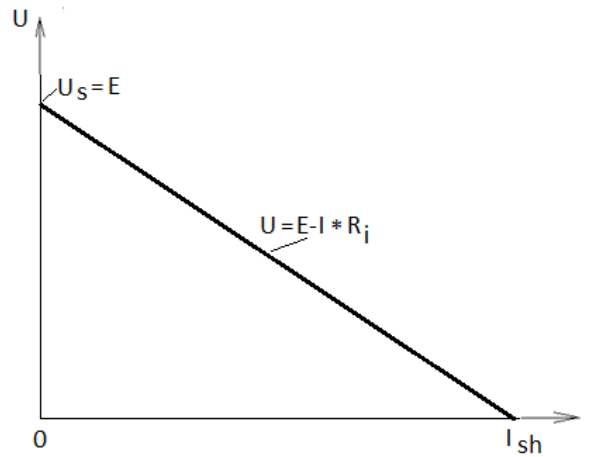


Figure 3.2. The loading characteristic of DC source

#### 3.1.4 Electrical Power

Electrical power of any element of the circuit is a product of its current and voltage:

$$P = U \cdot I, \text{ W (Watt)}. \quad (3.6)$$

Taking into account Ohm's law two other versions of realisation of the power could be used:

$$P = (I \cdot R) \cdot I = I^2 \cdot R, \quad (3.7)$$

$$P = U \cdot U/R = U^2/R. \quad (3.8)$$

It should be mentioned that in DC circuits all the power of the resistive elements is transformed into heat. In respect to the connecting wires a specific indicator – current density  $j$  - is used for the calculations of cross-section area of the wires for appropriate temperature conditions with current in the wire. For individual connecting wires this indicator is about  $j = (3 \dots 4) \cdot 10^6 \text{ A/m}^2$ ; for the wires in a coil it is  $j = 2 \cdot 10^6 \text{ A/m}^2$ .

For the protection of power source and whole electrical supply network an especially low resistance of wire with higher resistivity is inserted in series into the circuit. It is so called fuse (see fig.3.1) – circuit protection element. If the current in the circuit is inadmissibly high a power of higher density is dissipated in the resistance of the fuse (because of small size of the fuse's wire) and the wire is melted so open the circuit and protecting the network wires and supply.

Power generated by the EMF source is the following:

$$P_E = E \cdot I. \quad (3.9)$$

Relation between the power of the load and that of the source is called an efficiency factor:

$$\eta = \frac{P_L}{P_E}. \quad (3.10)$$

For the circuit in Fig.3.1.  $\eta = \frac{U \cdot I}{E \cdot I} = \frac{U}{E}$ .

Voltage drop across the internal resistance of the source could be defined as the losses of voltage which decrease the efficiency factor.

### 3.1.5 Complex electrical circuits and Kirchhoff's laws

A real electrical circuit could include more than one source of EMF as well as resistors (loads) connected in different ways. In Fig.3.3 one example of a complex circuit containing three resistors and two EMF is shown. Three indifferent ways directed currents are possible in this case. Current  $I_1$  is running to point 1 and currents  $I_2$  and  $I_3$  are running in the direction from it. There are three branches of current in the given scheme.

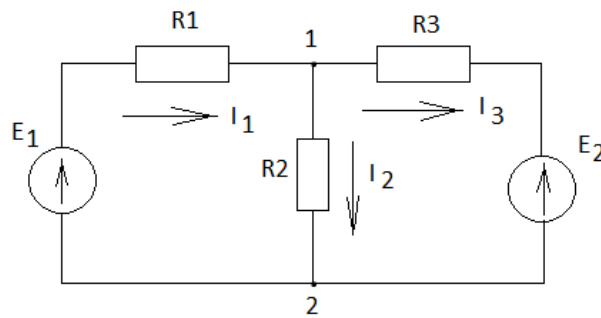


Figure 3.3. A complex DC electric circuit

Kirchhoff's Current Law states that the algebraic sum of all currents towards a node is zero. If the current runs to the node a symbol "plus" (+) is applied, if the current runs from it then it is "minus". Therefore in accordance with KCL:

$$+I_1 - I_2 - I_3 = 0, \quad (3.11)$$

or  $I_1 = I_2 + I_3$ .

This circuit can be also considered as a combination of two loops. For the loops Kirchhoff's Voltage Law states that the algebraic sum of the EMFs around the loops is equal to the sum of voltage drops. Thus for the left loop:

$$E_1 = I_1 \cdot R_1 + I_2 \cdot R_2, \quad (3.12)$$

For the right loop

$$-E_2 = I_3 \cdot R_3 - I_2 \cdot R_2. \quad (3.13)$$

Further applying together KVL, KCL and Ohm's law the current in each branch can be calculated as well as on the basis of these laws of electrical circuits a number of methods and approaches to solve complex electrical circuit are developed.

## 3.2 Circuits of single-phase alternating current

### 3.2.1 Circuits with resistors

If a circuit of alternating current contains a resistor  $R$  (see fig.3.4) the instantaneous value of the current  $i$  is changing in accordance with the changes of instantaneous value of the supply voltage  $= U_m \sin \omega t$  :

$$i = \frac{U_m}{R} \sin \omega t = I_m \sin \omega t. \quad (3.14)$$

A pointer of a voltmeter  $V$  connected in parallel to the input terminals and in series connected ammeter  $A$  could not follow to the quick changes of the instantaneous values. The devices are performed following the principle of electromagnetic influence, and they measure the real active value of the signals (so called root-mean-square value or RMS):

$$U = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} U_m^2 \sin^2 \omega t \cdot d\omega t} = I_m \sqrt{1/2}. \quad (3.15)$$

As we can see the correlation between the maximum and RMS values of the sine-form signal is

$$k_m = \frac{I_m}{I} = \frac{U_m}{U} = \sqrt{2}. \quad (3.16)$$

For RMS value  $= I \cdot R$  .

If to multiply the instantaneous value of the current and voltage the instantaneous value of the power is obtained as the curve in fig.3.4, which is only positive and changing with double frequency of electric supply network. As all instantaneous values of the power are positive this circuit has only active power, the average value per period of which is

$$P_{av} = \frac{1}{\pi} \int_0^{\pi} U_m I_m \sin^2 \omega t \cdot d\omega t = U \cdot I. \quad (3.17)$$

In the circuit under consideration the phasors of current and voltage are congruent in phase, i.e. they are of the same direction and rotate with the same velocity.

Active power is measured in watts (W) by means of wattmeter with two coils – the coil of current and the coil of voltage (Fig.3.5). The deviation of the indicator depends on the torque produced by interaction of the magnetic flux of both coils similar to the case of DC power.

Measurement of the RMS values of the current and voltage is realised by electro-magnetic devices on the basis of interaction of magneto-motive force of the coil with current and iron armature with a pointer (Fig.3.5.).

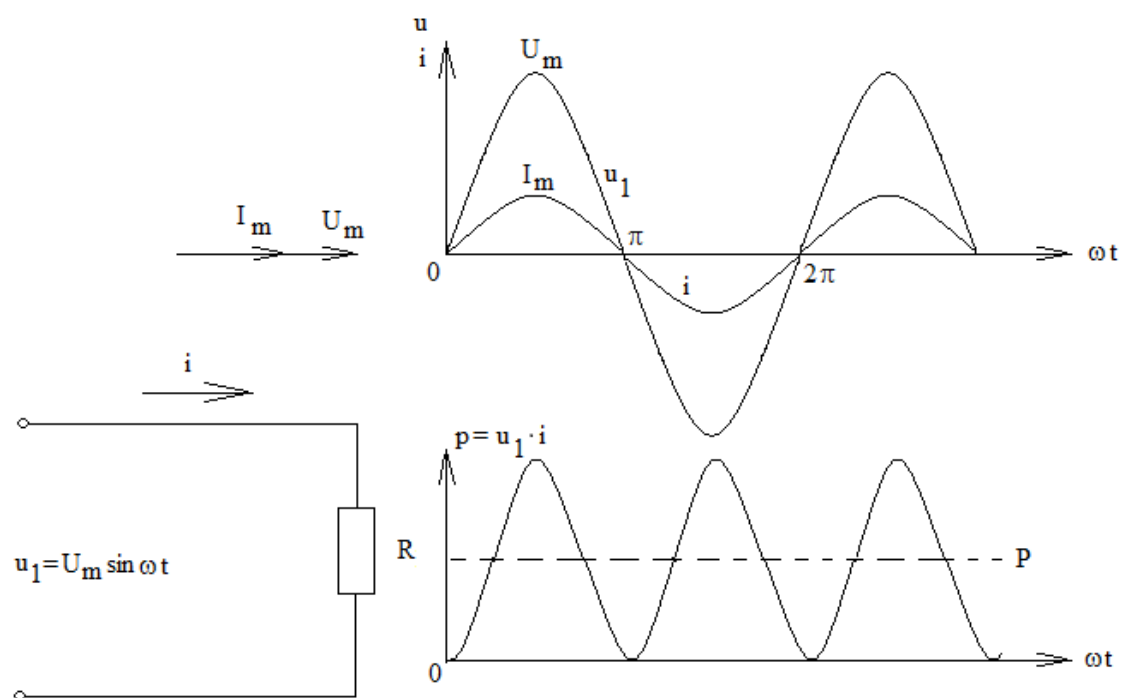


Figure 3.4. AC circuit with resistance and diagram of the its signals

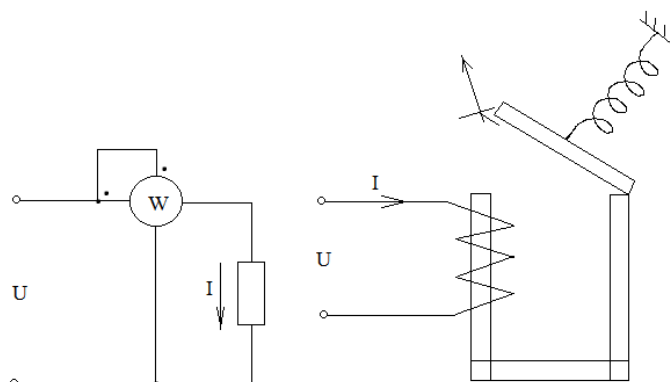


Figure 3.5. Measuring of power of AC circuit

### 3.2.2 AC circuits with coil (reactor or inductor)

In the case when the circuit contains an inductor only it looks like in Fig. 3.6.



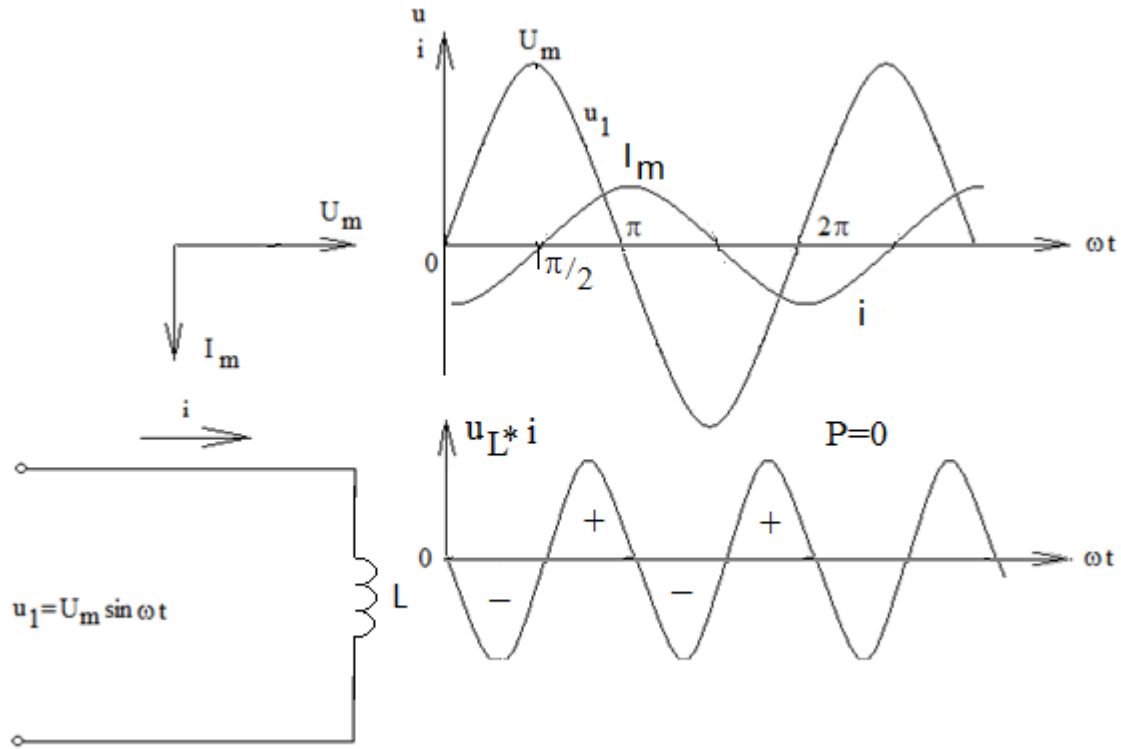


Figure 3.6. AC circuit with a coil and diagrams

As it was already said the voltage of the coil is the following:

$$u_L = U_m \sin \omega t = L \frac{di}{dt}. \quad (3.18)$$

With this (3.18) expression we can derive another one for an instantaneous value of the current:

$$i = \frac{U_m}{L} \int \sin \omega t \cdot d\omega t = -\frac{U_m}{\omega L} \cos \omega t = -I_m \cos \omega t. \quad (3.19)$$

This expression allows concluding that at  $t=0$  when  $u_L=0$  the current has a negative maximum value  $-I_m$  (Fig.3.6.), but at  $\omega t=\pi/2$ , when  $u_L=U_m$ , the current is  $i=0$ , i.e. the current lags behind the voltage by  $90^\circ$  in phase.

The maximum value of the current is calculating dividing  $U_m$  by  $\omega L$ . Therefore  $\omega L$  corresponds to the impedance of the coil. But the multiplication of  $U_L$  and  $i$  corresponds to a sine-form signal with double frequency, the average value of which is zero. It means that no active power exists there. Therefore the notion of inductance impedance is stated:

$$X_L = \omega L, \quad (3.20)$$

but the power periodically changing in the network is a reactive power:

$$Q_L = U_L \cdot I_L, \quad (3.21)$$

where  $U_L = \frac{U_m}{\sqrt{2}}$  - is RMS value of the voltage and  $I_L = \frac{I_m}{\sqrt{2}}$  - RMS value of the current

$$I_L = \frac{U_L}{\omega L}. \quad (3.22)$$

Consequently the reactive power could be represented as

$$Q_L = \frac{U_L^2}{\omega L}; Q_L = I_L^2 \cdot \omega L. \quad (3.23)$$

The measurement unit of reactive power is var.

### 3.2.3 AC circuits with capacitors

Capacitor is an electric device with two insulated plates. Direct current can not flow through the plates. In the case of alternative current the polarised biasing current is

$$i_C = C \frac{du_C}{dt}, \quad (3.24)$$

where capacity C is measured in farads (F). In fact, the capacity depends on a distance between the plates, their area and characteristics of an insulator between them.

The current flowing through the circuit with a capacitor is (Fig.3.7):

$$i_C = \frac{d(U_m \sin \omega t)}{dt} = \omega \cdot C \cdot U_m \cdot \cos \omega t = I_m \cdot \cos \omega t. \quad (3.25)$$

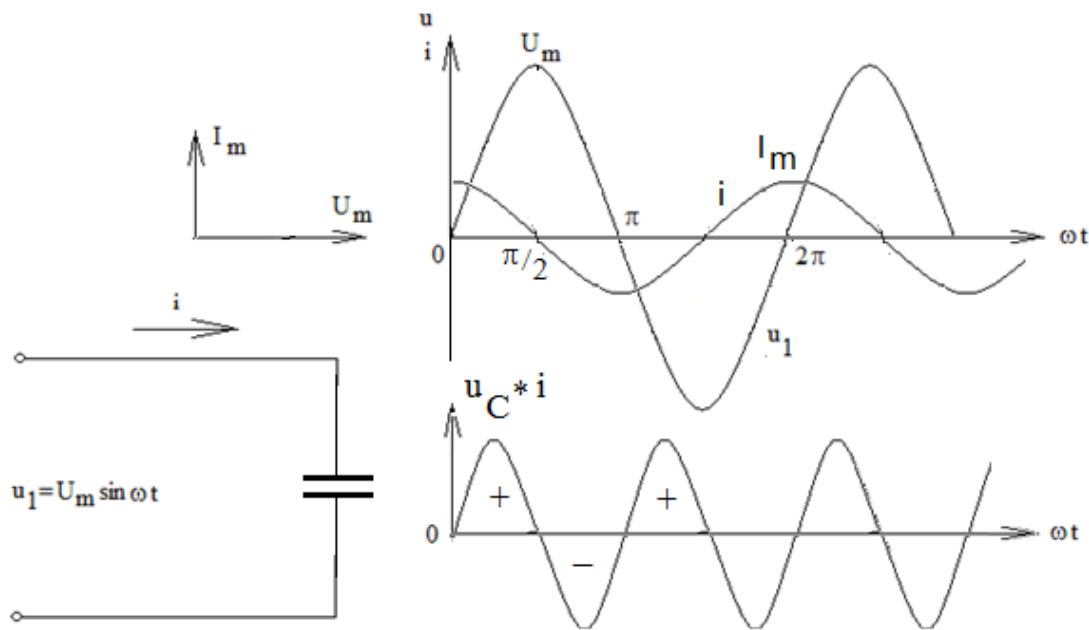


Figure 3.7. Scheme and diagrams of AC circuit with capacitor

When  $t=0$  and  $u_C=0$  the current has a positive maximum value  $I_m$ , but if  $\omega t=\pi/2$  and  $i_C=0$  the voltage has its maximum, i.e. the current  $i_C$  leads voltage by  $90^\circ$  in phase (Fig.3.7.).

The maximum value of the current could be calculated multiplying  $U_m$  by  $\omega C$ . Therefore value  $1/\omega C$  refers to a resistance of capacitor. But in this case like in the previous the multiplication

of  $u_C$  and  $i_C$  is a sine-form signal with double frequency, the average value of which is equal to zero, i.e. no active power generated. Therefore the capacitive impedance is

$$X_C = \frac{1}{\omega C}, \quad (3.26)$$

but the reactive power is

$$Q_C = U_C \cdot I_C = U_C^2 \cdot \omega \cdot C = \frac{I_C^2}{\omega C}, \quad (3.27)$$

where  $U_C = U_m/\sqrt{2}$ , but RMS value of the current is

$$I_C = \frac{U_C}{X_C} = U \cdot \omega \cdot C. \quad (3.28)$$

For both inductive and capacitive cases the phasors of reactive power as well as that of the current are opposite directed. When  $u_L \cdot i_L$  is negative (Fig.3.6.),  $u_C \cdot i_C$  is positive and vice versa.

### 3.2.4 AC circuits with coils and resistors

Conventionally the circuits with resistor R and inductance L are in practical applications (Fig.3.8).

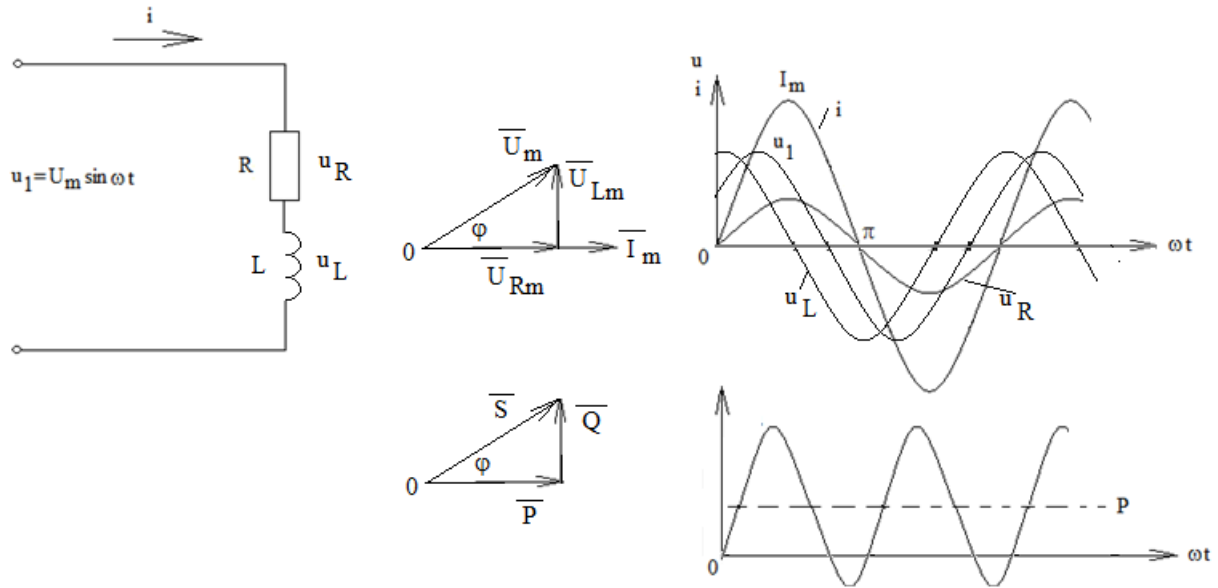


Figure 3.8. AC circuit with coil and resistor

In such circuit the same current  $i$  runs through all the elements. The phasor of the voltage across R coincides in phase with the phasor of current  $I_m$ , but phasor  $U_{Lm}$  leads the current  $I_m$  by  $90^\circ$ . Therefore

$$U_m^2 = U_{Rm}^2 + U_{Lm}^2 = I_m^2 (R^2 + X_L^2). \quad (3.29)$$

In terms of vectors KVL for a series connected circuit is considered as  $\overline{U}_1 = \overline{U}_L + \overline{U}_R$ .

For RMS value

$$I = \frac{U_1}{\sqrt{R^2 + X_L^2}} = \frac{U_1}{Z}. \quad (3.30)$$

where  $Z$  is a total impedance of the circuit.

In this case the product of  $u$  and  $i$  is a periodic curve with double frequency which is asymmetric in accordance to the axis of time, the average value of the signal (active power  $P$ ) is higher than zero level.

The voltage drop across the resistor gives active power (Fig.3.8.)

$$P = U_R \cdot I \text{ [W]}. \quad (3.31)$$

The same voltage gives a reactive power in the coil,  $90^\circ$  leading in phase

$$Q_L = U_L \cdot I = I^2 \cdot \omega L \text{ [var]}, \quad (3.32)$$

but the product of input voltage and current creates so called apparent power

$$S = U_1 \cdot I \text{ [VA]}. \quad (3.33)$$

From the diagram in Fig.3.8. we can see that

$$S^2 = Q_L^2 + P^2 \quad (3.34)$$

or

$$\cos\varphi = \frac{P}{S} = \frac{P}{U_1 \cdot I} = \frac{U_R}{U_1}. \quad (3.35)$$

The least parameter illustrates the relation between useful and total power of the circuit and is called “power factor”. This expression (3.35) can derive another widely used description of active power  $P = U_1 \cdot I \cdot \cos\varphi$ .

The sign of angle  $\varphi$  can be considered from vector diagram of power:  $\tan\varphi = \frac{Q_L}{P} > 0$ ,

i.e. in respect to the formed power for a circuit with inductive impedance and resistor angle  $\varphi$  is signed positive.

### 3.2.5 AC circuits with capacitor and resistor

In such kind of circuits (Fig.3.9.) the same current  $i$  runs through both elements. Vector  $U_{Rm}$  is in phase with  $I_m$ , but vector  $U_{Cm}$  lags that by  $90^\circ$ .

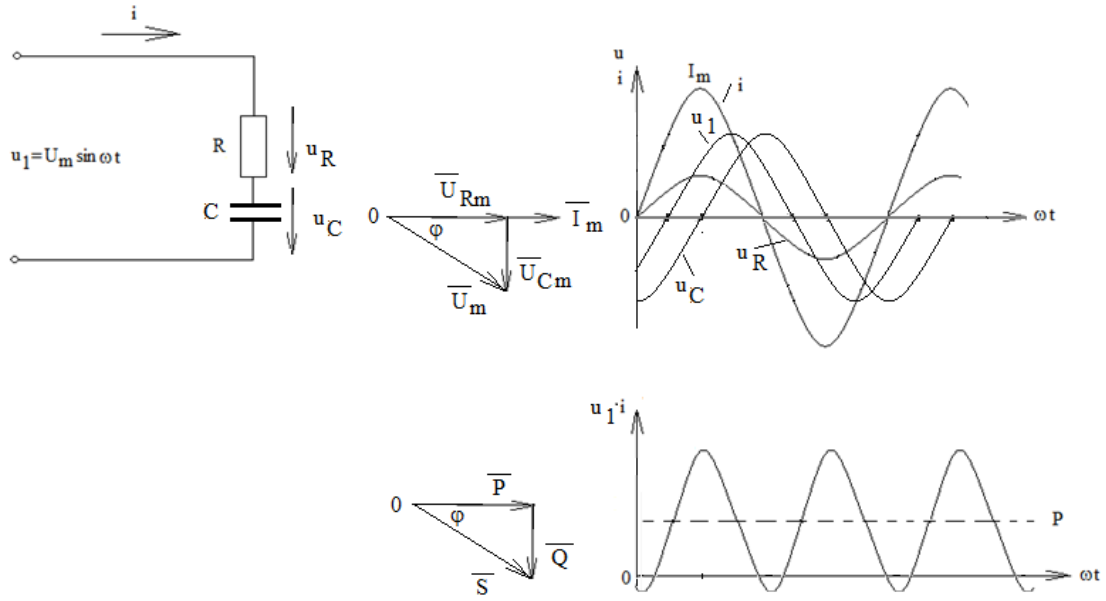


Figure 3.9. Scheme and diagrams of the circuit with capacitance and resistance

Therefore

$$U_m^2 = U_{Rm}^2 + U_{Cm}^2 = I_m^2 (R^2 + X_C^2). \quad (3.36)$$

For RMS values

$$I = \frac{U_1}{\sqrt{R^2 + X_C^2}} = \frac{U_1}{Z}, \quad (3.37)$$

where  $Z$  is a total impedance of the circuit. The product of  $u$  and  $i$  is also a periodic asymmetric to the axis of time curve of double frequency, the average value of which (active power  $P$ ) is higher than zero.

As we can see from Fig.3.9. vector of reactive power  $Q_C$  lags vector  $P$ , but at the same time vector of the apparent power  $S$  lags  $P$  for  $\varphi < 90^\circ$  as well. Power factor in this case is

$$\cos \varphi = \frac{P}{S} = \frac{P}{\sqrt{P^2 + Q_C^2}}, \quad (3.38)$$

where the reactive power of the capacitor is  $Q_C = I \cdot U_C = I^2 \cdot X_C = I^2 / \omega C$ .

From the vector diagram of power  $= \frac{-Q_C}{P} < 0$ .

As  $Q_L$  and  $Q_C$  are mutually opposite directed power  $Q_L$  is assumed to be with “+” sign, but  $Q_C$  – with “-”.

### 3.3 Three-phase AC systems

#### 3.3.1 Generation of three-phase voltage

Like other electric machines three-phase generators also consist of stator and rotor. The rotor in this machine is a constant magnet or electro-magnet, but the stator is represented a three frames of conductors which lag behind each other by  $120^\circ$  (Fig.3.10) with frames A-x, B-y, C-z. If the magnet rotates in a clockwise direction then at the position of the figure the polarity of EMF induced in the frame A-x is positive relatively to the lead x, but the amplitude of EMF is maximum. After rotating the rotor for  $60^\circ$  the lead C of the frame C-z is maximum negative relatively to z. After  $60^\circ$  more the lead B of the frame B-y is maximum positive relatively to y, etc. Thus a system of three sine-form alternative voltages shifted in time and space by  $120^\circ$  and relative rotating vectors  $\bar{U}_{AX}$ ,  $\bar{U}_{BY}$ ,  $\bar{U}_{CZ}$  of this system are formed (Fig.3.10.).

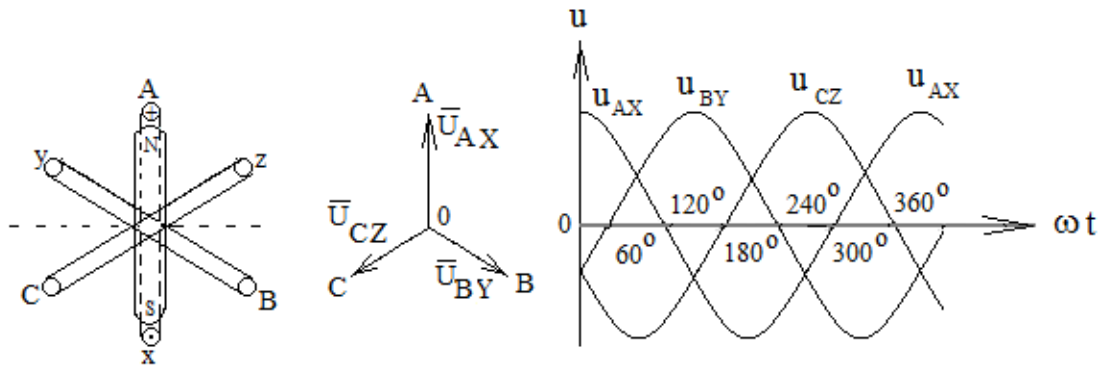


Figure 3.10. Realisation of three-phase sine-form voltages system

If the leads x, y, z are connected together forming so called zero point then the windings have “Y” connection. The voltage induced by each frame is called a phase voltage. Thus there are 3 phase voltages ( $\bar{U}_A$ ,  $\bar{U}_B$ ,  $\bar{U}_C$ ) shifted for  $120^\circ$  each to other and measured between pairs of leads A and 0, B and 0, C and 0 correspondingly. These voltages are called phase-to-zero voltages  $\bar{U}_{p-0}$ .

The load of the generator in its turn could be connected to each of three phase-to-phase voltages

$$\begin{aligned}\bar{U}_{AB} &= \bar{U}_A - \bar{U}_B; \\ \bar{U}_{BC} &= \bar{U}_B - \bar{U}_C; \\ \bar{U}_{CA} &= \bar{U}_C - \bar{U}_A.\end{aligned}\tag{3.39}$$

The correlation between phase-to-phase and phase-to-zero voltages could be defined from the vector diagram (Fig.3.12).

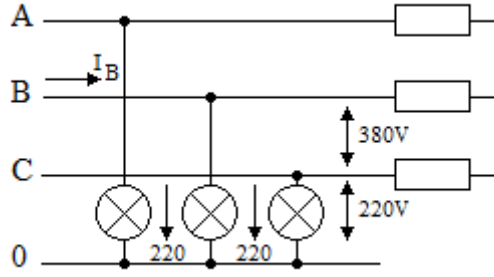


Figure 3.11. Electrical scheme of three-phase power supply

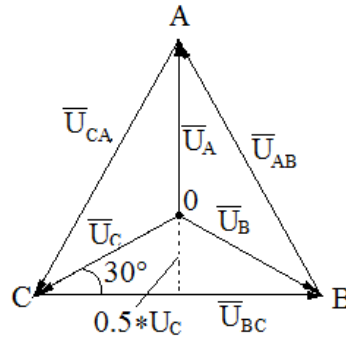


Figure 3.12. Vector diagram of three-phase system

For example from the triangle COB with the side  $0.5 \cdot U_{p-0}$  the following can be determined:

$$U_1 = 2 \sqrt{U_{p-0}^2 - 0.25 U_{p-0}^2} = 2 \cdot U_{p-0} \sqrt{3/4} = U_{p-0} \sqrt{3}. \quad (3.40)$$

The three phases with a zero wire are supplied to the industrial consumers (Fig.3.11.). A voltage of 220V between one phase and zero point, i.e. an RMS value of the phase-to-zero voltage, is used for lighting and household needs of consumers:  $U_1 = \sqrt{3} \cdot 220 = 380V$ .

To balance the supply system all three RMS values of the phase-to-zero voltages should be fully the same and symmetric. So do the three phase-to-phase voltages. Such supply system is symmetric. If all three phase currents of the generator are with the same RMS values and are at same shift angle in respect to the corresponding phase-to-zero voltage then the system of the currents is also symmetric, i.e.  $\bar{I}_A + \bar{I}_B + \bar{I}_C = 0$ . The algebraic sum of the instantaneous values of these three currents at any time is also zero. But if RMS if these currents are not the same or the angle between their vectors is not the same then the sum of all three vectors is not equal to zero and there is a current flowing in the zero wire. Such case is unbalanced or asymmetric.

For the case of symmetric load of a three-phase system the total power is

$$S_3 = 3 \cdot U_{p-0} \cdot I_1 = \sqrt{3} \cdot U_1 \cdot I_1; \quad (3.41)$$

active power

$$P_3 = 3 \cdot U_{p-0} \cdot I_1 \cdot \cos \varphi = \sqrt{3} \cdot U_1 \cdot I_1 \cdot \cos \varphi; \quad (3.42)$$

reactive power

$$Q_3 = 3 \cdot U_{p-0} \cdot I_1 \sin \varphi = \sqrt{3} \cdot U_1 \cdot I_1 \cdot \sin \varphi; \quad (3.43)$$

In the case of asymmetric load each phase has its own RMS value of current and shift angle. Then the total power is the sum of all three phase.

### 3.3.2 Connection schemes with symmetric load

The three-phase load can have wye (Y) or delta ( $\Delta$ ) connection (fig.3.13). In the case of symmetric load with Y connection the RMS values of all three currents are the same. As a total impedance ( $Z$ ) of each phase is the same the phase shift  $\varphi$  of all three currents in respect to a corresponding phase-to-zero voltages are also the same and a vector sum of the currents is

$$\bar{I}_A + \bar{I}_B + \bar{I}_C = 0. \quad (3.44)$$

Zero wire could be disconnected. Voltage  $U_{0-0L}$  between zero points of the generator (0) and that of the load (0L) is equal to 0.

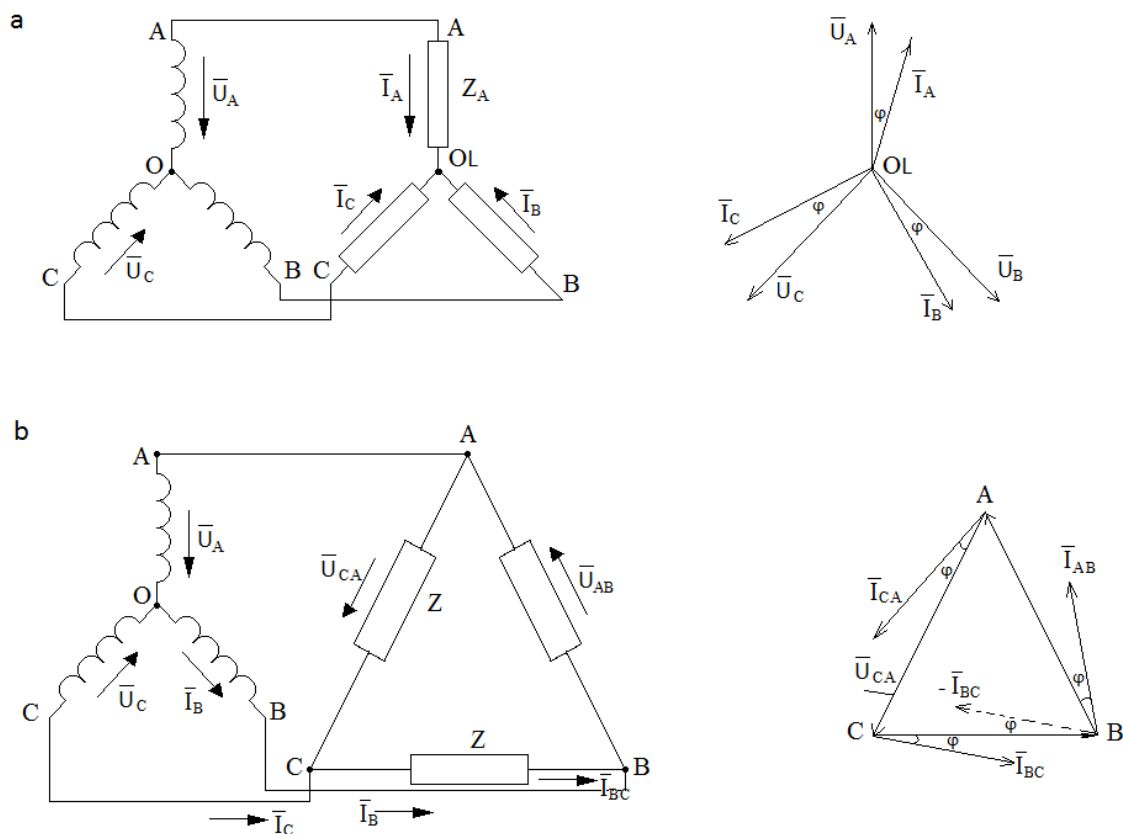


Figure 3.13. Electrical schemes for Y (a) and  $\Delta$  (b) connections of the load

Zero wire does not need  $\Delta$  connection and each of three voltages of the load impedance (legs) is equal to line-to-line voltage. If all three impedances  $Z$  are the same the total power of the load is



$$S_{3A} = 3 \cdot U_1 \cdot I_Z = \sqrt{3} \cdot U_1 \cdot I_1 . \quad (3.45)$$

The angle between  $I_{AB}$  and  $-I_{BC}$  is  $60^\circ$  and the vector sum of them is  $I_B = \sqrt{3} \cdot I_{AB} = \sqrt{3} \cdot I_{BC}$ .

### 3.3.3 Y Connection of asymmetric load

If there is a zero wire in the system then in the case of an asymmetric load

$$\bar{I}_N = \bar{I}_A + \bar{I}_B + \bar{I}_C \neq 0 \quad (3.46)$$

and the current  $I_N$  leaks through the zero wire. The system of phase-to-zero voltages is symmetric because it depends on the symmetric system of supply voltages. If the load is asymmetric, i.e.  $Z_A \neq Z_B \neq Z_C$ , and the system has no zero wire then a voltage exists between zero point of supplying source 0 and load 0L.

This voltage could be defined calculating the following system of equation:

$$\begin{aligned} \bar{U}_A + \bar{U}_{0-0L} - \bar{I}_A \cdot Z_A &= 0 , \\ \bar{U}_B + \bar{U}_{0-0L} - \bar{I}_B \cdot Z_B &= 0 , \\ \bar{U}_C + \bar{U}_{0-0L} - \bar{I}_C \cdot Z_C &= 0 , \\ \bar{I}_A + \bar{I}_B + \bar{I}_C &= 0 . \end{aligned} \quad (3.47)$$

Then the voltage between zero points

$$\bar{U}_{0-0L} = \frac{\bar{U}_A \cdot Z_B \cdot Z_C + \bar{U}_B \cdot Z_A \cdot Z_C + \bar{U}_C \cdot Z_A \cdot Z_B}{Z_B \cdot Z_C + Z_A \cdot Z_C + Z_A \cdot Z_B} . \quad (3.48)$$

The load voltage of phase A

$$\bar{U}_a = \frac{\bar{U}_A (2 \cdot Z_B \cdot Z_C + Z_A \cdot Z_C + Z_A \cdot Z_B) + \bar{U}_B \cdot Z_A \cdot Z_C + \bar{U}_C \cdot Z_A \cdot Z_B}{Z_B \cdot Z_C + Z_A \cdot Z_C + Z_A \cdot Z_B} . \quad (3.49)$$

The load voltage of phase B

$$\bar{U}_b = \frac{\bar{U}_B (2 \cdot Z_A \cdot Z_C + Z_B \cdot Z_C + Z_A \cdot Z_B) + \bar{U}_A \cdot Z_B \cdot Z_C + \bar{U}_C \cdot Z_A \cdot Z_B}{Z_B \cdot Z_C + Z_A \cdot Z_C + Z_A \cdot Z_B} . \quad (3.50)$$

And that of phase C

$$\bar{U}_c = \frac{\bar{U}_C (2 \cdot Z_A \cdot Z_B + Z_B \cdot Z_C + Z_A \cdot Z_C) + \bar{U}_B \cdot Z_A \cdot Z_C + \bar{U}_A \cdot Z_A \cdot Z_B}{Z_B \cdot Z_C + Z_A \cdot Z_C + Z_A \cdot Z_B} . \quad (3.51)$$

Although the system of phase-to-line voltages is out of symmetry, the system of phase-to-phase voltages is symmetric.

## **4 SPECTROSCOPY, MICROSCOPY, MAGNETOMETRY, ELLIPSOMETRY**

## 4.1 Optical spectroscopy of molecules

Molecule is the smallest part of a substance possessing its main chemical and physical properties and capable of independent existence. The molecule consists of different or the same atoms coupled into the single entity by *chemical bonds*. The number of atoms in the molecule may vary from two ( $\text{H}_2$ ,  $\text{N}_2$ ,  $\text{CO}$  etc.) to thousands (e.g. protein molecules). Coulombic attraction of electrons to nuclei of atoms plays the decisive role in chemical bond formation. However, only the quantum theory allows understanding the chemical bond nature as well as physical and chemical properties of molecules.

### 4.1.1 Energy levels and types of molecular spectra

**Energy of molecules. Electronic energy levels.** As it has been mentioned a molecule is a quantum-chemical system and in order to describe it, i.e. to get whole set of possible eigenfunctions (or wavefunctions) and corresponding eigenvalues (energies) for a molecule, one should solve Schrödinger equation, taking into account the motion of electrons and nuclei, interactions between electrons, electrons and nuclei, and also interaction between nuclei within the molecule. The exact analytical solution of this equation, even in the case of a diatomic molecule (except hydrogen ion  $\text{H}_2^+$  [https://en.wikipedia.org/wiki/Dihydrogen\\_cation](https://en.wikipedia.org/wiki/Dihydrogen_cation)) is impossible. However, at some conditions with sufficient accuracy the mentioned problem can be reduced to the system of three new and related problems, which could be solved in a consecutive manner. These problems are as follows: determination of electronic  $E_e$ , vibrational  $E_v$  and rotational  $E_r$  energies of the molecule as well as their corresponding wavefunctions. In this so called Born-Oppenheimer approximation the total energy of the molecule is represented as a sum of the corresponding energies

$$E = E_e + E_v + E_r. \quad (4.1)$$

Initially Schrödinger equation for the electronic wave function  $\psi_e$  should be solved. It describes spatial distribution of electron probability density in the molecule with nuclei fixed at some distances from each other. Only one parameter - distance  $\rho$  between the atoms is needed to describe nuclei configuration for diatomic molecules, while for multi-atom molecules more parameters (or internal coordinates) are required. The solution of Schrödinger equation gives the set of the electron wave functions  $\psi_e(n)$  and corresponding values of electron energies  $E_e(n)$  (where  $n$  – state number). It is noteworthy, that additional approximation - one-electron approximation model is often used while finding the Schrödinger equation solution. The one-electron approximation is based on the assumption that the quantum system can be described as a system of individual electrons moving in an averaged potential field, which takes into account interactions with atomic nuclei and with other electrons. The multi-electron wavefunction  $\psi_e(n)$  of a many-electron system in the one-electron approximation has form of the Slater determinant of a certain set of one-electron wavefunctions. Additionally, electrons in diatomic molecules could be conventionally divided into two types: the electrons remaining at the atomic electron orbitals (called nonbonding  $n$ -electrons) and electrons which wavefunctions participate in construction of molecular  $\sigma$ - and/or  $\pi$ -electron orbitals, providing interatomic chemical bond. For diatomic molecules in a specific

electronic state the internuclear distance  $\rho$  affects values of electronic wavefunction  $\psi_e(n)$  and its energy value  $E_e(n)$ . Thus, solution of Schrödinger equation for the electron wavefunction of the diatomic molecule at different  $\rho$  values ( $0 < \rho < \infty$ ) gives series of  $E_e(\rho)$  functions (so-called *configuration curves*). This is in contrast to the case of atom where the solution gives set of numbers— electronic energy levels (compare Fig. 4.1. and Fig. 4.2.).

Since  $E_e(\rho)$  contains all kinds of the intramolecular energy except the nuclei kinetic energy the function  $E_e(\rho)$  represents the effective potential energy of the fixed molecule. Under this approach the functions  $E_e(\rho)$  are often denoted by  $U_e(\rho)$  and called *potential curves*.

The necessary condition for molecule stability in the given electronic state is the presence of a minimum for  $E_e(\rho)$  function at a certain  $\rho = \rho_{0n}$  value – *equilibrium* internuclear distance (Fig. 4.2). The function  $E_e(\rho)$  increases rapidly if  $\rho$  value becomes smaller than  $\rho_{0n}$  (due to nuclei repulsion) and tends to a limit  $E_e(\infty)$  at larger distances  $\rho$  (at  $\rho \rightarrow \infty$  the mutual interaction between the atoms goes to zero).

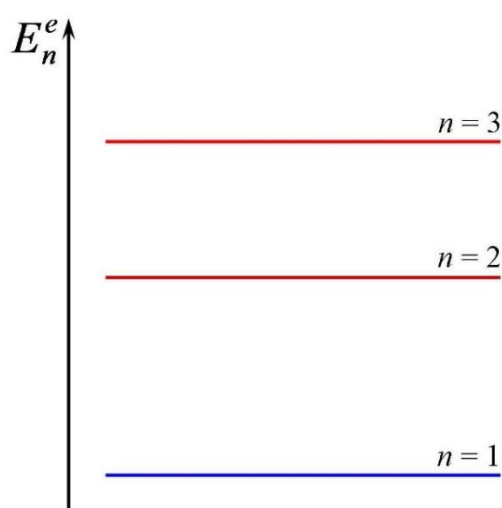


Figure 4.1. Typical electronic energy levels diagram of an atom

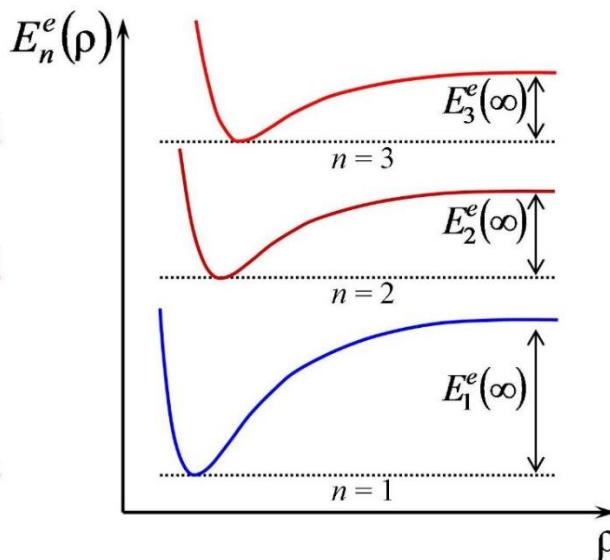


Figure 4.2. Typical potential energy surfaces (configuration curves) for three consecutive electronic states in the molecule

If total energy of a molecule reaches  $E_e(\infty)$  value the molecule will break down i.e. *dissociate* into ions or neutral atoms (depending on nature of chemical bonds). If the number  $n$  increases (i.e. we are going to higher excited electronic states) shape of configuration curves changes: usually, the curve minimum becomes shallow and the equilibrium distance  $\rho_{0n}$  increases (Fig.4.2.). Sometimes, the second excited state ( $n = 3$ ) could not be realized due to too shallow minimum and the molecule becomes unstable.

As mentioned above each electronic state of a molecule is characterized by a specific configuration curve and the corresponding wavefunction  $\psi_e(n)$  which describes spatial density distribution of electrons. As an example the Fig. 4.3. represents the distribution of the electron density for two ground states for  $H_2$  and  $N_2$  molecules.

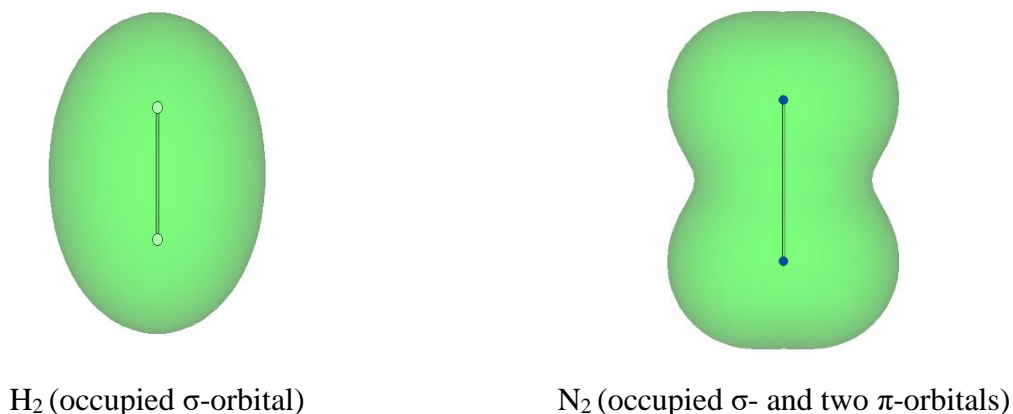


Figure 4.3. Isosurface plot of total charge density in the ground states of  $H_2$  and  $N_2$  molecules. Basis set cc-pVDZ, charge density contour value =0.035

Thus the most important characteristics of electronic state for each molecule are: equilibrium distance  $\rho_{0n}$ , the minimum value of the electron energy  $E_e(\rho_{0n})$  and the energy difference  $E_e(\infty) - E_e(\rho_{0n})$ . This difference (up to a *zero-point vibrations* energy, see below) is equal to molecule dissociation energy  $D_n$  for the corresponding electronic state.

The deeper is the minimum on configuration curve, i.e. when  $D_n$  is larger the chemical bond will be stronger. As a rule, dissociation energy for the ground electronic state ( $n=1$ ) in the molecule is the highest.

The energy difference between neighboring electronic states decreases when number  $n$  increases and for atoms and molecules has the value of 1 – 10 eV.

**The vibrational energy of diatomic molecules.** Diatomic molecule vibrations problem is solved by means of one-dimensional Schrödinger equation for the vibrational wave function  $\psi^v(\rho)$ . In this equation the electron energy  $E_e(\rho)$  should be used as a potential energy  $U(\rho)$  in Schrödinger equation. Configuration curves  $E_e(\rho)$  should be used for each electron in multi-electron molecules to get total potential energy and this results in a set of vibrational problems for the molecule.

The selection of  $U_n(\rho)$  function type depends on the particular model taking into account both theoretical considerations and experimental data.

**Harmonic oscillator model.** The potential function  $U(\rho)$  can be expanded in Taylor series near the equilibrium position  $\rho_0$ :

$$U(\rho) = U(\rho_0) + \left( \frac{dU}{d\rho} \right)_{\rho_0} (\rho - \rho_0) + \frac{1}{2} \left( \frac{d^2U}{d\rho^2} \right)_{\rho_0} (\rho - \rho_0)^2 + \frac{1}{6} \left( \frac{d^3U}{d\rho^3} \right)_{\rho_0} (\rho - \rho_0)^3 + \dots \quad (4.2)$$

Considering that function  $U(\rho)$  has a minimum at  $\rho = \rho_0$  (so that  $\left( \frac{dU}{d\rho} \right)_{\rho_0} = 0$ ) and neglecting in (4.2) cubic and all the subsequent terms one could obtain

$$U(\rho) - U(\rho_0) = \frac{1}{2} \left( \frac{d^2U}{d\rho^2} \right)_{\rho_0} (\rho - \rho_0)^2 = \frac{1}{2} k (\rho - \rho_0)^2 \quad (4.3)$$

where  $k = \left( \frac{d^2U}{d\rho^2} \right)_{\rho_0}$  the *stiffness coefficient* of the interatomic bond within the molecule (its force constant).

The system possessing potential energy (4.3) is called a *linear harmonic oscillator* (LHO). According to quantum mechanics LHO has the following allowed energy levels

$$E_v = h\nu_0 \left( v + \frac{1}{2} \right), \quad (4.4)$$

where  $h$  is the *Planck's constant*,  $v = 0, 1, 2, 3, \dots$  is *vibrational quantum number*. According to (4.4) the minimum energy  $\frac{1}{2} h\nu_0$  is a so-called *zero-point vibrations energy* and corresponds to the ground state ( $v=0$ ) of the harmonic oscillator, and the energy difference between neighboring vibrational levels is constant and equal to  $h\nu_0$  (Fig. 4.4.). The frequency  $\nu_0$  corresponds to the frequency of the classical harmonic oscillator and associated with force constant  $k$  and atomic masses  $m_1$  and  $m_2$  as follows

$$\nu_0 = \sqrt{\frac{k}{M}}, \quad M = \frac{m_1 m_2}{m_1 + m_2}. \quad (4.5)$$

**Anharmonic oscillator model.** This model takes into account that configuration curve  $U(\rho)$  of the real molecule is close to a parabola only near the bottom of the potential well, and in case of large differences between  $\rho$  and  $\rho_0$  the curve has significantly another shape from parabola. Right branch of the "real" potential  $U(\rho)$  is flatter than the same one of the parabola and hence such a potential well expands with increasing energy faster than for the harmonic oscillator (compare Fig. 4.4. and Fig. 4.6). Rigorously in this case the vibrational energy levels will gradually come closer to each

other approaching the dissociation limit (unlike equally spaced LHO levels). Thus the difference between the actual configuration curve  $U(\rho)$  of the real molecule and parabola  $U(\rho_0) + \frac{1}{2}k(\rho - \rho_0)^2$  requires to consider *anharmonic* terms in the expansion (4.2) and leads to deviations of (4.4) when dealing with real molecules. Corrections to (4.4) are usually written as expansion in powers of  $(\nu + \frac{1}{2})$ :

$$E_\nu = h\nu_0\left(\nu + \frac{1}{2}\right) - xh\nu_0\left(\nu + \frac{1}{2}\right)^2 + yh\nu_0\left(\nu + \frac{1}{2}\right)^3 + \dots \quad (4.6)$$

where  $x, y, \dots$ , coefficients called *anharmonicity* constants. Typically these constants are sufficiently small such that  $0 < x \ll 1$ ,  $|y| \ll x$  etc. In contrast to LHO the number of vibrational levels in a diatomic molecule is usually finite: quantum number  $\nu$  takes values  $\nu=0, 1, 2, \dots, \nu_{lim}$  (Fig. 4.5.).

Obviously the sufficient condition for chemical bonding is the presence in the potential well at least one vibrational level and the dissociation energy  $D$  is equal to the well depth with the zero-point vibration energy subtracted (Fig. 4.2. and Fig. 4.5.)

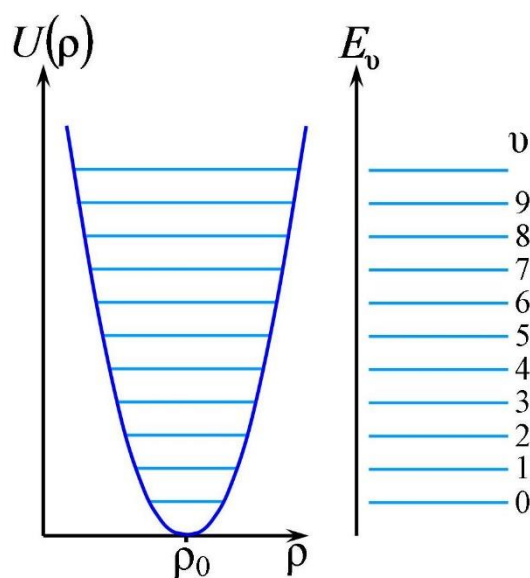


Figure 4.4 The scheme of the vibrational energy levels of a diatomic molecule in the LHO approach

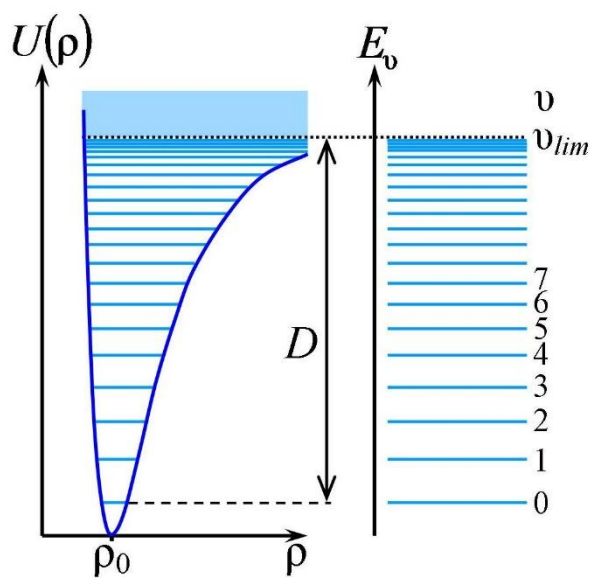


Figure 4.5. The scheme of the vibrational energy levels for a diatomic molecule, assumption of anharmonic vibrations

Providing a molecule with an energy higher than dissociation energy  $D$  results in breaking down of the molecule so that each separate atom (fragment) acquires some translational velocity. Since

the translational kinetic energy of non-bound atoms can take any value (non-quantized), the absorption spectrum of the molecule above the dissociation limit will be continuous (Fig. 4.6).

Hereinafter let us consider the first order of *anharmonicity* correction in (4.6)

$$E_v = h\nu_0 \left( v + \frac{1}{2} \right) - xh\nu_0 \left( v + \frac{1}{2} \right)^2 \quad (4.7)$$

Then the molecule vibrational energy in the state  $v = 0$  (zero-point vibrations energy) is equal to

$$E_v(0) = \frac{1}{2}h\nu_0 - \frac{1}{4}xh\nu_0, \quad (4.8)$$

and the distances between neighboring vibrational levels are *the first energy differences*  $\Delta_1 E_v = E_{v+1} - E_v$  and according to (4.7) do not remain constant (as it would be for LHO approach) and they will decrease linearly as the energy increases:

$$\Delta_1 E_v = E_{v+1} - E_v = h\nu_0 - 2xh\nu_0 \left( v + \frac{1}{2} \right) \quad (4.9)$$

According to (4.9) the number of vibrational levels is finite, and the maximal (*boundary*) vibrational number  $v_{lim}$  is approximately equal to

$$v_{lim} \approx \frac{1}{2x} \quad (4.10)$$

and thus directly related to the anharmonicity constant  $x$ .

Finally, the dissociation energy  $D$  is obviously equal to the sum of the all first differences  $\Delta_1 E_v$ :

$$D = \sum_{v=0}^{v_{lim}} \Delta_1 E_v \quad (4.11)$$

Expression (4.11) containing  $v_{lim}$  of terms according to (4.9) presents the sum of terms of an arithmetic progression decreasing from  $h\nu_0 - 2xh\nu_0$  to zero, and therefore,

$$D = \frac{1}{2}(h\nu_0 - 2xh\nu_0)v_{lim} \quad (4.12)$$

Given the fact  $x \ll 1$ , the following expression can be used instead of (4.12):

$$D = \frac{1}{2}h\nu_0 v_{lim} \quad (4.13)$$



Each electronic state has its own configuration curve and therefore has its own system of vibrational levels. Energy differences between vibrational levels are significantly less than between the electronic ones of about  $10^{-2} - 10^{-1}$  eV.

**Rotational energy of a diatomic molecule.** This part of the molecule energy is determined by solving the Schrödinger equation for the rotational motion. In this approach electronic and vibrational states are assumed to be found already and the distance between the nuclei in the first approximation is assumed to be constant (rigid rotor model). The rotational energy of a diatomic molecule is related to its angular momentum  $L$  as  $E_r = \frac{L^2}{2I}$ , where  $I$  is the molecule moment of inertia relative to the rotation axis passing through the center of mass and perpendicular to the molecular axis. Module of angular momentum vector can take the following quantized values

$$L = \hbar \sqrt{J(J+1)}, \quad (4.14)$$

Where  $J = 0, 1, 2, \dots$  the rotational quantum number. Therefore,

$$E_r = \frac{L^2}{2I} = \frac{\hbar^2}{2I} J(J+1) = B J(J+1), \quad (4.15)$$

Where  $B = \frac{\hbar^2}{2I}$  is the rotational constant. Formally, the (4.15) indicates that the rotational levels form an infinite divergent system. Actually, however, there is a limit rotational energy, where the centrifugal forces "break" the molecule. As (4.15) approaches this limit it is no longer applicable.

Energy differences for the lower rotational levels are two orders less than for vibrational ones and have values of about  $10^{-4} - 10^{-3}$  eV.

**Diatomic molecule energy levels scheme.** Using the mentioned above Born-Oppenheimer approximation the typical scheme of the energy levels of a diatomic molecule can be represented as follows (Fig. 4.6) Each of its electronic states has a set of possible values for vibrational and rotational energies.

**Three types of molecular spectra.** One can distinguish three types of molecular spectra based on transition energy and what kind of energy levels are involved at the transition (Fig. 4.6.).

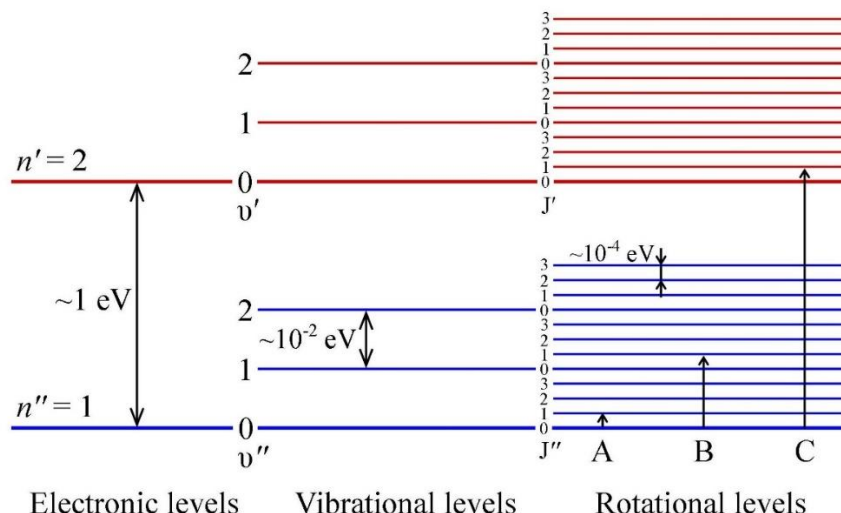


Figure 4.6. Scheme of the electron-vibrational-rotational levels of molecule

**Rotational spectra.** This type of spectra occurs under A type transitions (Fig. 4.6) when only rotational state of the molecule changes. Line spectral frequencies can be derived from the rule of Bohr frequencies:

$$\nu_r = \frac{1}{h}(E_{J'} - E_J) = \frac{B}{h}[J'(J'+1) - J(J+1)] \quad (4.16)$$

Keeping in mind the selection rule for rotational quantum number

$$\Delta J \equiv J' - J = \pm 1, \quad (4.17)$$

the transitions are only possible if quantum of absorption or emission energy of electromagnetic radiation is equal to the difference between neighboring rotational levels

$$\nu_r = \frac{2B}{h}(J+1) \quad (4.18)$$

Thus the rotational spectrum of a diatomic molecule is a series of equidistant lines.

Since  $\Delta E_r = h\nu_r$  has the order of magnitude  $10^{-4} - 10^{-3}$  eV, frequencies and wavelengths for rotational spectrum are as follows:

$$\nu_r = 10^{10} - 10^{11} \text{ Hz}; \quad \lambda_r = 1 \text{ cm} - 1 \text{ mm}.$$

These lines belong to the microwave range, so the rotational spectra are studied by radio spectroscopy methods.

Only the molecules possessing permanent electric dipole moment (i.e. non-zero dipole moment even out of the external electric field) have the rotational spectra. Additionally the substance must

be in gas phase. Molecules without permanent electric dipole moment ( $H_2$ ,  $Cl_2$ ,  $I_2$  etc.) do not absorb light and do not have rotational spectrum.

**Vibrational (vibrational-rotational) spectra.** In order to change vibrational state of a molecule it is necessary to supply energy quantum of  $10^{-2} - 10^{-1}$  eV. Since  $\Delta E_v$  is approximately 100 times larger than  $\Delta E_r$ , change of vibrational state of molecules in gas phase is usually accompanied by change in rotational state also. As a result series of type B transitions are possible among vibrational and rotational energy levels (Figure 4.6). Therefore, in contrast to pure rotational spectra considered earlier, upon excitation of a molecule to another vibrational state series of spectral lines (vibrational-rotational spectrum) can be observed. Expression for the frequencies in vibrational-rotational spectrum is:

$$\nu_v = \frac{1}{h} [(E'_v - E_v) + (E'_r - E_r)] \quad (4.19)$$

Since  $\Delta E_v$  has value of  $\sim 10^{-2} - 10^{-1}$  eV, vibrational-rotational spectrum lies in near infrared range:

$$\nu_v = 10^{12} - 10^{13} \text{ Hz}; \quad \lambda_v = 10^{-2} - 10^{-3} \text{ cm.}$$

The character of the spectrum is easy to figure out if contribution related to rotational levels change is not considered. Since the distance between the vibrational levels gradually decreases and the selection rules for the anharmonic oscillator allow transitions between any vibrational states pure vibrational spectrum would be a group of lines that converge toward shorter wavelengths up to a certain limit. Due to the small rotational term (3.10.19) each pure vibrational line splits into a band with so-called rotational structure (Fig. 3.10.7). In order to observe this fine structure one need to use a spectrometer with sufficiently high resolution. Each line in the vibrational-rotational spectrum is associated with a transition between a pair of levels with certain values of  $v$  and  $J$ . However, "allowed" i.e. actually possible transitions are those that satisfy the selection rules. These rules for vibrational-rotational transitions are as follows:

$$\Delta v = \pm 1, \pm 2, \pm 3, \dots \quad \Delta J = \pm 1. \quad (4.20)$$

Transitions with  $\Delta v = \pm 1$  are more probable to occur than with  $\Delta v = \pm 2$ , and so on. Figure 4.7 shows vibrational-rotational spectrum of HCl.

In fact in diatomic molecules with non-zero electronic angular momentum (for example NO molecule) transitions with  $\Delta J = 0$  are also possible.

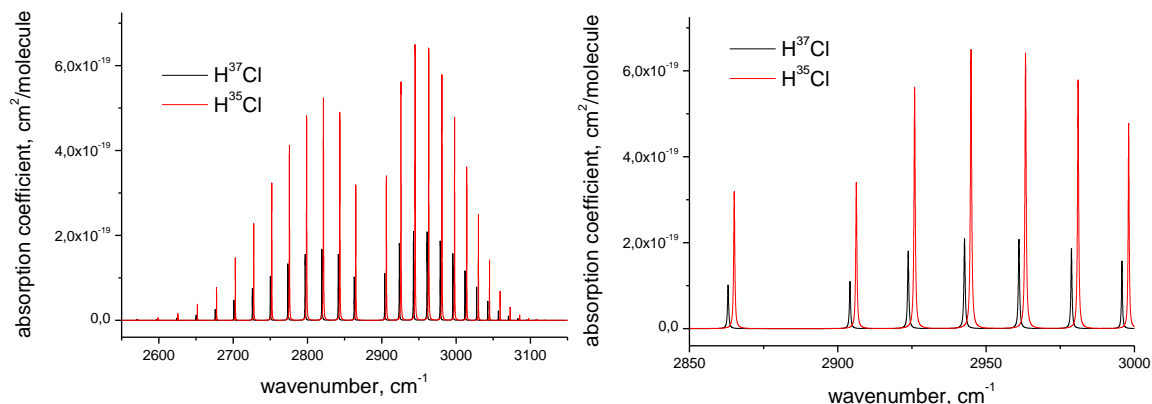


Figure 4.7. IR absorbance spectra of HCl isotopologues simulated using the data of Hitran database.  $T = 296\text{K}$ ,  $p = 1\text{ atm}$

Thus, the frequency  $\nu_v \approx \frac{\Delta E_v}{h}$  as well as other parameters are the specific for a diatomic molecule. Eigenvalues or natural vibration frequencies as well as other constants for the set of diatomic molecules are represented in Table 4.1.

Table 4.1. Constants that characterize the equilibrium configurations of diatomic molecules

Molecule	Internuclear distance $\rho$ , Å	Moment of inertia $I$ , $10^{-47}\text{ kg}\cdot\text{m}^2$	Rotational constant $B$ , meV	Vibration frequency $\nu$ , $10^{14}, \text{s}^{-1}$	Force constant $\chi$ , kN/m
$H_2$	0.741	0.46	7.40	1.318	0.57
$N_2$	1.094	13.99	0.24	0.708	2.30
$O_2$	1.207	19.46	0.18	0.474	1.18
$Cl_2$	1.988	117.15	0.029	1.064	0.34
$HCl$	1.275	2.64	1.29	0.897	0.52
$HF$	0.917	0.80	4.25	1.241	0.97
$CO$	1.128	14.57	0.23	0.651	1.90

**Spectra (electronic-vibrational-rotational spectra).** The nature of these spectra is determined by transition between electronic states of the molecule accompanied by corresponding changes in vibrational and/or rotational energy levels (C type transitions Fig. 4.6.). The corresponding band frequencies in the absorption and emission spectra are determined by the expression:

$$\nu_e = \frac{1}{h} [(E'_e - E_e) + (E'_v - E_v) + (E'_r - E_r)] \quad (4.21)$$

Due to the contribution of quantum energy  $h\nu_e = \Delta E_e$  caused by electronic transition the electronic-vibrational-rotational spectrum falls into the visible or near-ultraviolet electromagnetic range  $\nu_e = 10^{14} - 10^{15}$  Hz;  $\lambda_e = 10^{-4} - 10^{-5}$  cm.

Since many vibrational and rotational sublevels belong to both initial and final electronic states the electronic absorption and emission spectra of gas-like substances have very complex structure and consist of the relatively narrow bands series.

The Boltzmann distribution law (4.22) for molecules across the energy levels and the Franck-Condon principle are the basis for the relative intensity interpretation of the vibrational bands in electronic-vibrational spectra. According to Boltzmann equation the level with the lowest electronic and vibrational energy is the most populated:

$$\frac{N_i}{N_0} = \exp \left[ - \frac{h\nu_{ik}}{kT} \right] \quad (4.22)$$

For example, there are only ~0.1% of the total number of hydrogen molecules at room temperature which are in the first excited vibrational level ( $\nu = 1$ ) while the absolute majority of the molecules remain in the ground vibrational state ( $\nu = 0$ ).

Therefore, the bands corresponding to the transitions from  $n=1, \nu=0$  to  $n=2, \nu=1, 2, 3, \dots$  are much more intense than the bands corresponding to the transitions from  $n=1, \nu=1$  to  $n=2, \nu=2, 3, \dots$  in the electronic spectra of the diatomic molecules.

**Franck-Condon principle.** Due to the fact that the electron mass is much smaller than mass of the nuclei the optical electronic-vibrational transitions can be considered to happen in the molecule with fixed position of nuclei. Graphically the electronic-vibrational transitions can be represented as vertical straight arrows between energy levels of the initial and final states of different electronic potential energy curves. These arrows should be drawn in such a way that their continuations cross the interatomic distances axis at the point corresponding to the most probable internuclear distance typical for the initial electronic state (Fig.4.8). Graphical representation of the vibrational wave functions for the combining states is also shown in Fig. 4.8. It is important to note that probability of the transition between these states depends on overlap of their wave functions.

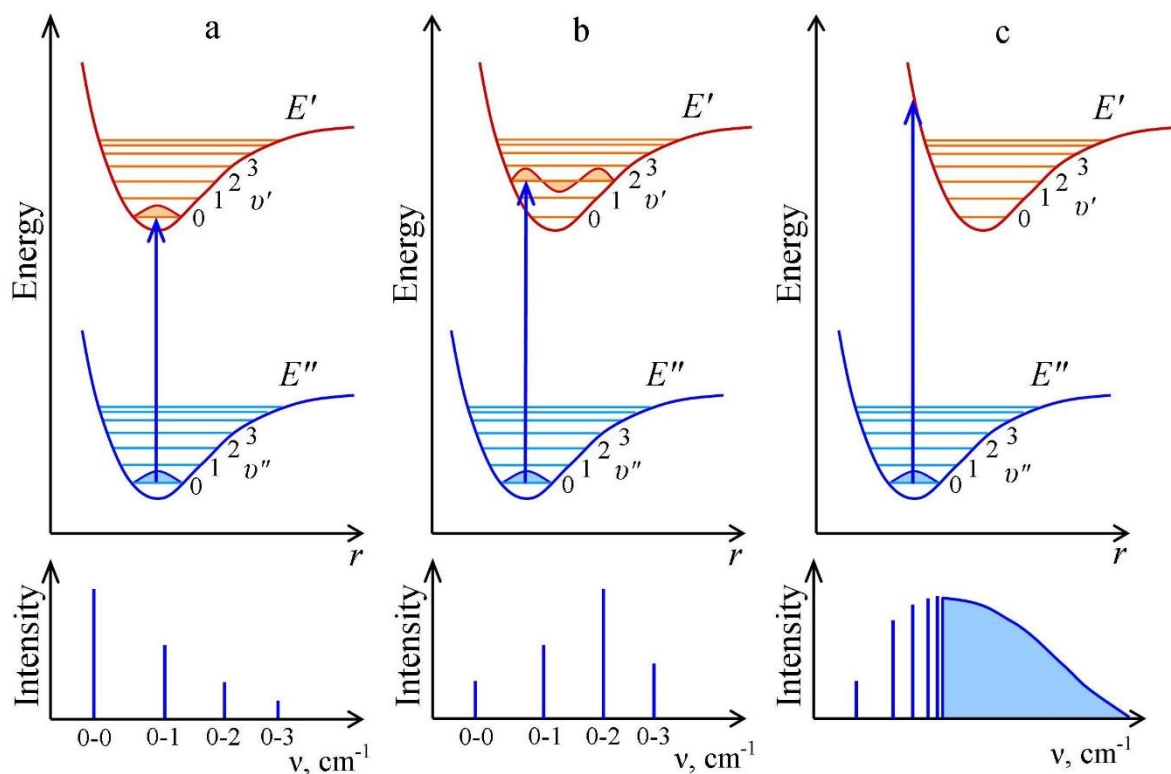


Figure 4.8. Demonstration of the Franck-Condon principle for different situations: a - internuclear distance in the upper and lower states is the same; b - internuclear distance in the upper state is slightly larger than in the lower state; c - internuclear distance in the upper state is significantly larger than in the lower state

#### 4.1.2 Vibrational spectroscopy of polyatomic molecules

**IR (infra-red) absorption spectroscopy of the polyatomic molecules.** The IR spectroscopy method is one of most widely used methods of vibrational spectroscopy due to the simplicity of obtaining spectra and comparatively low cost of spectral setup and that is very important in qualitative and quantitative analysis of substances.

The polyatomic molecules are characterized by complex vibrational motion of atoms. However, any intermolecular vibrations can be represented as the superposition of a certain number of normal vibrational modes. Each of these modes has the characteristic frequency  $\nu_i$  and is quantized similar to the above model of an anharmonic oscillator. The vibrations of molecules are usually considered in the framework of the so-called natural coordinate system, where stretching of chemical bonds  $q_i$  and change of the angles between them  $\gamma_i$  are analyzed. If the vibration demonstrates mainly change in the bond lengths and bond angles are not changed such vibrations are called *stretching* vibrations. If vibrations result mainly in change of bond angles, and bond lengths are almost

unchanged then such vibrations are called *deformation or bending* vibrations. The stretching vibrations are often denoted by the symbol « $\nu$ », deformation ones as « $\delta$ ».

Depending on the symmetry, vibrations can be symmetric or antisymmetric. Stretching symmetric vibrations are denoted as  $\nu_s$  the antisymmetric -  $\nu_{as}$ . The examples of symmetric and antisymmetric stretching vibrations are shown in Fig. 4.9.

#### Stretching vibrations



#### Deformation vibrations

##### Plane deformation vibrations



##### Out of plane deformation vibrations



Figure 4.9. Examples of symmetric and antisymmetric stretching and deformation vibrations. Often the vibrations are denoted using the following symbols:  $\nu$  – stretching;  $\delta$  – in-plane deformation;  $\gamma$  – out-of-plane deformation;  $\rho$  – rocking; s - symmetric; as -antisymmetric

The number of  $Z$  vibration modes is determined by the number of atoms  $N$  in the molecule and its structure. For linear molecules

$$Z = 3N - 5 \quad (4.23)$$

and for non-linear ones

$$Z = 3N - 6 \quad (4.24)$$

The necessary condition for appearance of vibrational absorption or emission spectra is related to whether electric dipole moment of the molecule changes during the vibration. In particular, that is the reason why vibrational absorption spectra could not be observed for diatomic molecules consisting of the identical atoms. In many cases IR spectra of polyatomic molecules do not contain symmetrical vibrations lines. This feature reduces the possibility of IR spectroscopy. It is considered that molecular vibrations that do not change electric dipole moment of the molecule are inactive in IR spectra. However, such vibrations are active in Raman spectra (Raman scattering). The physical nature and application of Raman scattering will be considered separately.

The IR spectroscopy method is a powerful tool for analytical problems of substances containing polyatomic molecules. It is important to obtain the qualitative IR spectrum of the test sample and interpret it using some set of analytical bands. Frequencies and intensities of the spectrum lines represent "fingerprint" of a molecular structure and are specific characteristics of the molecule. According to formula (3.10.5) frequency of the vibrational bands are determined by force constants for each  $q_i$  vibrational coordinate. Thus, they could be determined by quantum chemical calculations or by spectral measurements of reference substances in terms of compound investigation. The analytical spectra interpretation implies establishing correlations between the experimentally observed spectrum bands and certain types of intermolecular vibrations. The vibrations characterizability and the database of the characteristic vibrations frequencies of different chemical bonds of atomic groups should be used for interpretation. Comparison of spectra of numerous different molecules revealed that certain atomic vibrations in different molecules often have the same or close frequency values called *characteristic*. The vibrations of C-H, C-C, C=C, C-O groups and many others are examples of such type vibrations. All atoms in polyatomic molecules are involved into vibrations. However, the vibration amplitude can vary significantly. The presence of characteristic frequencies in the vibrational spectra often allows making conclusion about the molecule structure.

The relative intensity as well as the spectral position of the IR vibrational bands is also an individual characteristic of molecular structure. The molecule vibrational bands intensity in the IR spectrum is proportional to the square of the dipole moment component for the given vibrational coordinate. The absolute band intensity in the IR spectrum depends on the sample molecule concentration. This fact is used for substance quantitative analysis.

Table 4.2 shows approximate positions of the main vibrational bands in the IR spectra for the characteristic vibrations of atomic groups for different classes of chemical compounds. There are tables of characteristic frequencies of the basic functional groups of molecules which show the



frequencies and relative intensities of vibrations of basic functional groups belonging to different classes of substances.

Table 4.2. The characteristic vibrations frequencies of the atomic groups in the IR spectra of polyatomic molecules

Functional group	Vibration frequency, $\text{cm}^{-1}$	Intensity
C–H, stretching	2700-3300	intense
C–H, bending	1300-1500	intense
C–H, bending	800-830	intense
O–H, stretching	3000-3700	intense
O–H, bending	1200-1500	medium
N–H, stretching	3000-3700	medium
N–H, bending	1500-1700	intense
C–C, stretching	800-1200	medium
C=C, stretching	1600-1700	medium
C–O	1050-1300	intense
C=O	1690-1760	intense
NO <sub>2</sub>	1300-1370	intense
$C \equiv C$	2100-2400	weak
$C \equiv N$	2100-2400	medium
C–F	1000-1350	intense
C–Cl	710-770	intense

The stretching vibration frequencies of hydrogen atom bound to carbon, nitrogen or oxygen belong to short-wavelength (high frequency) spectral range 3700 – 3200  $\text{cm}^{-1}$ . Frequencies of X-H groups are in the highest frequency range of the spectrum.

Attaching atom with increased mass to the carbon atom results in appearance of absorption bands in the lower frequency range of the spectrum. Thus, vibration frequency of the C-H-groups is about 3000  $\text{cm}^{-1}$ , C–C vibrations frequencies are in the range of 1100–900  $\text{cm}^{-1}$ , and C–Br about 600  $\text{cm}^{-1}$ . The increase in chemical bond order results in vibrations frequencies increase.

IR spectrum of aromatic rings shows medium intensity peak of stretching vibrations C-H about 3030  $\text{cm}^{-1}$ .

Another characteristic feature is that stretching vibrations of aromatic carbon-carbon bonds are usually observed at 1600 and 1475  $\text{cm}^{-1}$ .

Finally, the aromatic ring reveals intense absorption band in the range of 800-690  $\text{cm}^{-1}$ , caused by deformation C-H vibrations. All of the mentioned features of the aromatic ring are observed in the toluene IR spectrum (Fig. 4.10.).

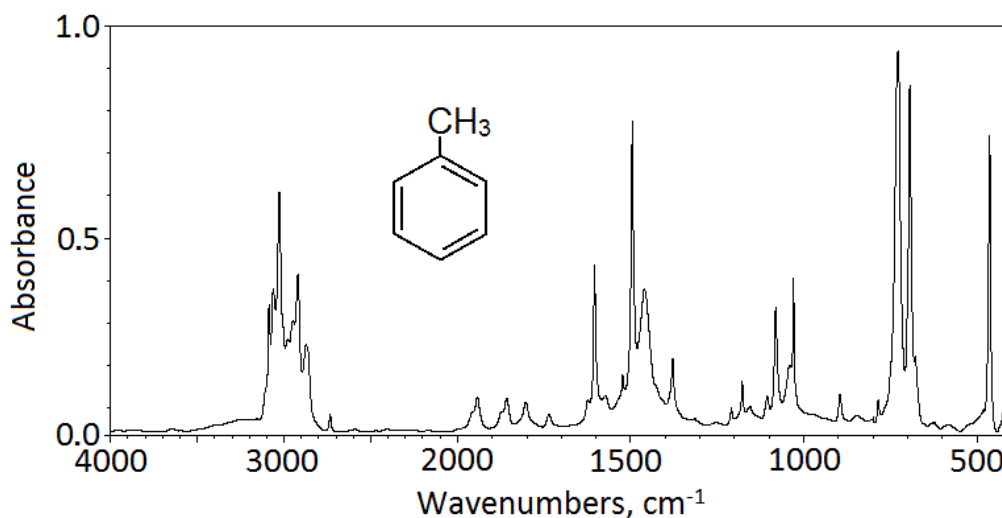


Figure 4.10. IR spectrum of toluene

### The scheme of substance analysis using IR absorption spectra

1. Classification of the organic compounds (saturated, unsaturated, aliphatic, aromatic, heteroaromatic) taking into account the IR absorption spectra is based on the characteristic frequencies of stretching vibrations of C-H, C-C, C = C, C $\equiv$ C bonds and also out of plane deformation vibrations of CH-bonds. The range of frequencies 2500-1500  $\text{cm}^{-1}$  is the most informative for this purpose.

2. After compound classification it is necessary to identify the main functional groups present in the molecule. Absorption bands in the range of 4000-2500  $\text{cm}^{-1}$  should be analyzed. It allows unique identification functional groups such as O-H, N-H, S-H, as well as various types of carbon-hydrogen groups C $_{\text{sp}3}$ -H, C $_{\text{sp}2}$ -H, C $_{\text{sp}}$ -H, (O =) C-H (aldehyde). After the characteristic bands of stretching vibrations for certain types of bonds in the spectrum are detected it is necessary to find out the bonds of respective deformation vibrations in the range of 1500-500  $\text{cm}^{-1}$  for example in the case of O-H, N-H, C-H bonds.

## Raman Spectroscopy.

Raman scattering is an inelastic light scattering related to electric dipole moment oscillation of the molecule induced by electric field  $\vec{E}$  of incident electromagnetic wave. The induced dipole moment of the molecule is determined by its polarizability and the wave electric field

$$\vec{p} = \epsilon_0 \alpha \vec{E}. \quad (4.25)$$

In case of monochromatic wave with frequency  $\nu_0$

$$\vec{E} = \vec{E}_0 \cos 2\pi \nu_0 t \quad (4.26)$$

Value  $\alpha$  is a function of molecular vibrational coordinates and depends on distance between the atoms, i.e.  $\alpha = \alpha(q_i)$ , where  $q_i$  is the generalized vibrational coordinate of atoms in the molecule. In the simplest approximation

$$\alpha(q_i) = \alpha_0 + \left( \frac{\partial \alpha_i}{\partial q_i} \right)_{q=0} q_i, \quad (4.27)$$

where  $q_i = q_{i0} \cos(2\pi \nu_i t)$ . As a result one can obtain

$$\begin{aligned} \vec{p} = & \alpha_0 \vec{E}_0 \cos 2\pi \nu_0 t + \frac{1}{2} \left( \frac{\partial \alpha_i}{\partial q_i} \right) q_i \vec{E}_0 \cos 2\pi(\nu_0 - \nu_i)t + \\ & + \frac{1}{2} \left( \frac{\partial \alpha_i}{\partial q_i} \right) q_i \vec{E}_0 \cos 2\pi(\nu_0 + \nu_i)t. \end{aligned} \quad (4.28)$$

Thus, under the field of propagating monochromatic electromagnetic wave, an induced electrical dipole moment of the molecule will oscillate with frequencies  $\nu_0$ ,  $\nu_0 \pm \nu_i$ . As a result the secondary emission appears as Raman scattering and bands with satellite frequencies  $\nu_0 + \nu_i$  and  $\nu_0 - \nu_i$  will be observed in addition to the fundamental frequency  $\nu_0$  (Fig.4.11).

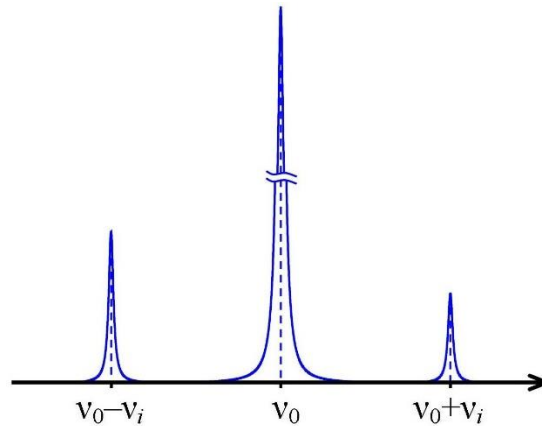


Figure 4.11. Raman spectrum

Occurring of combination frequencies is the consequence of the incident wave field modulation by intermolecular vibrations.

The following features are typical for the Raman scattering.

1. Besides the main line of Rayleigh scattering with  $\nu_0$  frequency in the spectrum there are lines of  $\nu_0 \pm \nu_v$  frequencies representing combinations of the main frequency and the frequencies of the molecule normal vibrations. The intensities of these lines by 6-8 orders of magnitude are lower than the Rayleigh line intensity. Lines with the frequencies of  $\nu_0 - \nu_v$  are called *Stokes lines*; lines with the frequency of  $\nu_0 + \nu_v$  – *anti-Stokes lines*.
2. The difference between the Rayleigh line frequency and the frequency of Raman line does not depend on the incident light frequency and is equal to one of the natural molecule vibration frequencies.
3. The intensity of anti-Stokes lines is much lower than the intensity of the Stokes lines and increases with temperature growth.

Classical theory explains observed combination frequencies in scattering spectra of molecules and crystals but does not allow determination of the individual vibrational bands intensity as well as the difference of the Stokes and anti-Stokes lines intensities. The most complete description of Raman scattering spectra is given by quantum theory.

According to the quantum model the radiation of frequency  $\nu$  is considered as a flow of photons with energy  $h\nu$ . The photons scatter in collisions with molecules. In the case of the elastic scattering photons deflect from the direction of their movement without energy change (Rayleigh scattering). There are also cases when collision causes energy exchange between the photon and the molecule. At this kind of collision the molecule can get or lose some energy according to quantization rules – its energy can change by  $\Delta E$  value corresponding to the energy difference of its two allowed vibrational states. In other words,  $\Delta E$  value must be equal to the molecule vibrational and / or rotational energy change. When the molecule acquires energy  $\Delta E$  at the scattering event the photon changes its energy to  $E_0 - \Delta E$  or frequency to  $\nu - \Delta E/h$  respectively. When the molecule loses energy  $\Delta E$  at the scattering the frequency of the scattered photon will be equal to  $\nu + \Delta E/h$ .

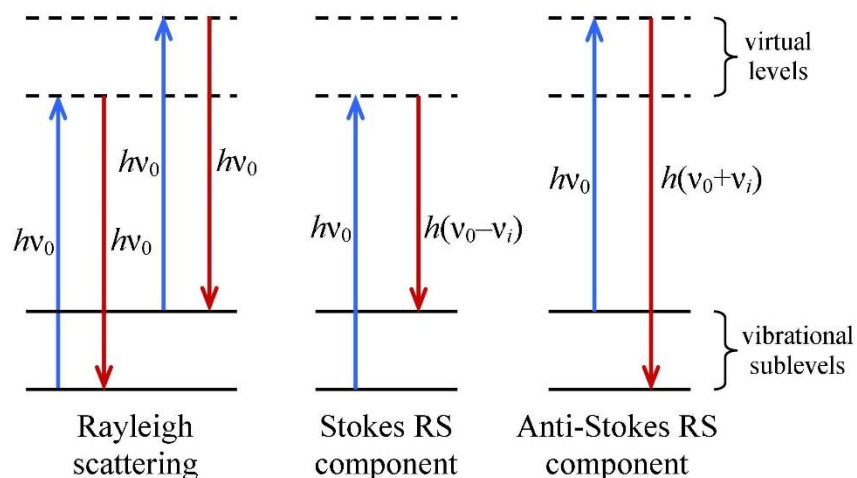


Figure 4.12. The Scheme demonstrating the Raman Scattering quantum model

According to the quantum transitions theory the spectral line intensity is proportional to the number of particles participating in the transition, i.e. number of particles at the initial energy level (level population). Therefore, under thermodynamic equilibrium conditions the intensities ratio of Stokes and anti-Stokes lines should be equal to the population ratio of the corresponding vibrational levels and can be determined by Boltzmann formula (3.10.22). Expression (3.10.22) shows that intensity of the Stokes lines will always be higher than the anti-Stokes lines intensity, and the intensity ratio depends on the temperature and vibrations frequency.

At low temperatures population of the first vibrational level is low. At room temperature for vibrational frequency of  $1000\text{ cm}^{-1}$  there are only 0.7% of the molecules on the first vibrational level, so the anti-Stokes scattering intensity is low. With temperature growth population of the excited vibrational level increases and the intensity of anti-Stokes lines increases.

*Raman spectroscopy* is a section of molecular spectroscopy studying Raman spectra and is widely used for structural analysis and chemical compounds identification.

Raman spectroscopy has the significant disadvantage due to low scattering cross-section, i.e. the relatively low intensity of the individual vibrational bands. Therefore it is necessary to use intensive light sources for excitation and relatively high concentrations of molecules in solutions (usually  $\sim 10^{-2}$  mole/liter or higher).

**Surface-enhanced Raman spectroscopy (SERS).** Study of vibrational spectra of molecules adsorbed on metal surfaces relies on phenomenon of *surface-enhanced Raman scattering* (SERS) discovered in early 80's of the last century. The essence of this phenomenon is significant increase of Raman signal intensity when molecules are adsorbed on roughened metal surfaces (Au, Ag, Cu) or on small (100-1000 Å) particles of noble metals. The intensity enhancement in some cases reaches  $\sim 10^6$  times. The theory of SERS phenomenon is still incomplete; however, its basic regularities are clear today.

It is assumed, as in the case of Raman scattering, the intensity of SERS spectral lines is proportional to the product of molecule polarizability and electric field (see the expression (3.10.25)). Therefore, intensity enhancement also depends on two factors. One of them is associated with increasing in the electric field acting on the molecule near the metal surface and provides increase of Raman cross-section by  $\sim 10^2$ – $10^3$ . This enhancement is determined by resonance of incident or scattered electromagnetic radiation with oscillating electric field of electrons localized near the roughness of metal surface or in a separate metal particle at specific eigenfrequencies. The existence of such a resonance effect also gives qualitative explanation why highest enhancement is achieved close to surfaces of noble metals with high reflectivity in the visible range of spectrum.

Another factor of Raman signal enhancement is related to change of molecule polarizability due to coupling with metal electrons. This kind of interaction has chemical origin. The magnitude of this "chemical" enhancement depends on the type of interactions between the adsorbed molecule and metal surface. Due to chemical interactions with the surface the distribution of bands intensities within SERS spectrum of complex molecules differs significantly from that in Raman spectrum and depends on metal substrate nature.

Roughened surface of silver electrode in electrochemical cell is often used as substrate for SERS spectra. The surface roughening occurs using so called oxidation-reduction cycles. Another substrate type is metal colloid (sols) particles with sizes ranging from 5 to 200 nm. One more type of substrates is island (granular) metal films obtained by metal vacuum deposition on dielectric surface. Such films are composed of metal nanoparticles two-dimensionally distributed on the plane of substrate. It is possible to obtain quasi-periodic film structures with metal islands having lateral dimensions of 5-250 nm and heights of  $\sim 100$  nm or more by varying vacuum deposition conditions (deposition rate, substrate temperature).

The metal island films and lattices of regular structure produced by microlithography methods are also widely used.

One of advantages of SERS spectroscopy is the possibility to detect extremely small amounts of substances (concentration of the solution down to  $10^{-15}$  mole / liter or even less). Using this SERS possibility is extremely important for the study of objects available in very small quantities as well as analysis of real systems with low concentration of analyte molecules. The progress in SERS spectroscopy methods has resulted in the possibility of single molecule detection.

Besides the use of SERS method for study of surface properties, SERS spectroscopy has been widely used in medicine and biophysics studying processes of structure of drug-target complexes, pharmacokinetics of drugs in living cells, etc. It allows investigation of mechanism of drugs action under conditions close to physiological. Such systems are also widely used in the products of chemical reactions analysis, control of the reactants purity as well as the control of environmental pollution by organic compounds.

The schematic representation of a typical experimental system used in SERS spectroscopy is shown in Fig. 4.13.

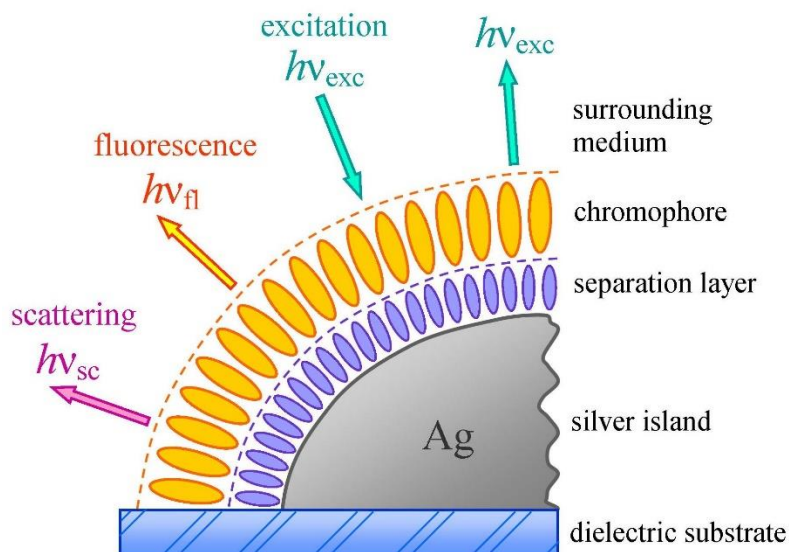


Figure 4.13. The model of SERS experimental system

To provide specific adhesion properties silver island film placed atop the dielectric substrate surface (Ag, Fig. 4.13) can be modified (coated) with a thin layer of surface-active compounds (e.g. stearic acid). Upon irradiation with frequency  $\nu_{ex}$ , the light partially undergoes reflection and elastic scattering without change of photon energy  $h\nu_{ex}$  or is inelastically scattered as SERS signal ( $h\nu_{sc}$ , Raman scattering) or emitted as fluorescence  $h\nu_{fl}$ . In some specific cases not only the Raman signal enhancement occurs but luminescence enhancement of the adsorbed molecules is also observed.

Fig.4.14 shows the Raman spectra of known fluorescent dye thioflavin T under different registration conditions.

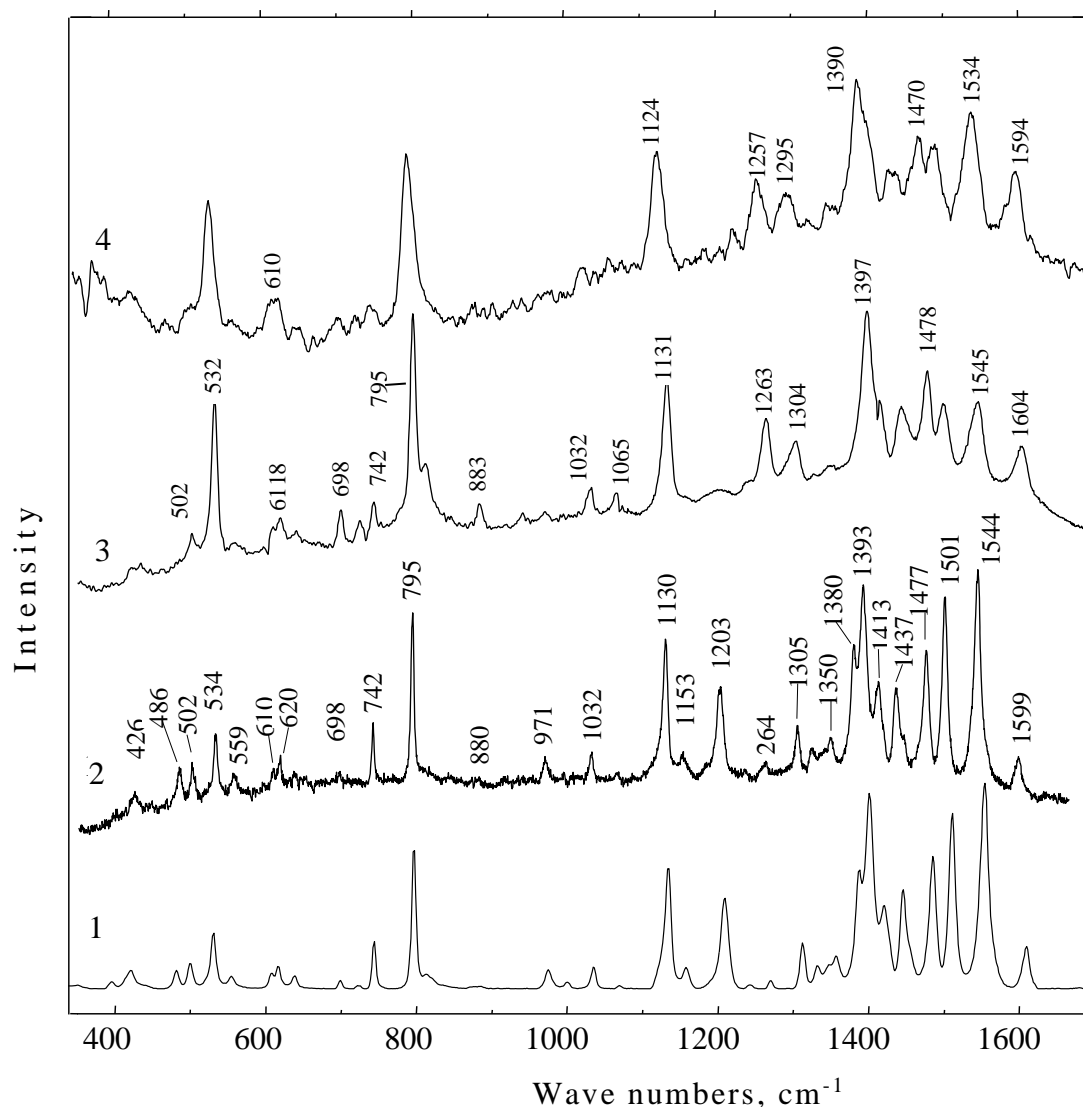


Figure 4.14. Calculated (1), experimental RS spectra of the powder of thioflavin T (2), and SERS spectra of thioflavin T at the SERS-active substrates: citrate sol (3) and thin silver film (4). Excitation wavelength - 633 nm (2) and 473 nm (3, 4)

It is important to note significant difference between RS and SERS spectra of the thioflavin T dye. The probability of vibrational transitions within the molecule changes as the result of its interaction with the metal surface, i.e. the amplification is observed not for all vibrational bands. One can see that the most active vibrational bands in SERS spectrum are at 532, 795, 1139 and 1390  $\text{cm}^{-1}$ . The main contribution to the vibrational bands is made by stretching vibrations of atoms forming the aromatic rings. Silver films, used as a substrate, provide mainly electromagnetic while silver hydrosols provide both electromagnetic and chemical mechanisms for signal amplification. The difference in the number of bands intensity (Fig. 4.14.) is due to the difference in the contribution of chemical and electromagnetic mechanisms of enhancement.



### 4.1.3 Luminescent molecular analysis

#### 1. Background

The luminescence method is one of the most effective methods used for detection and determination of substances, as well as for the study of their structure at molecular and atomic levels.

The prominent feature of this method is the ability to observe the emission signal from extremely low concentrations of luminescent substance ( $10^{-10}$ – $10^{-12}$  g/cm<sup>3</sup>). Use of lasers as excitation sources and highly sensitive photodetectors allowed to significantly improve limit of detection and resulted in capability to register and study even single fluorescing molecules. High sensitivity of luminescence analysis is widely used to solve many scientific and practical problems in geophysics, mineralogy, metallurgy, chemical industry, textile industry, criminalistics, medicine, biology, and even in construction.

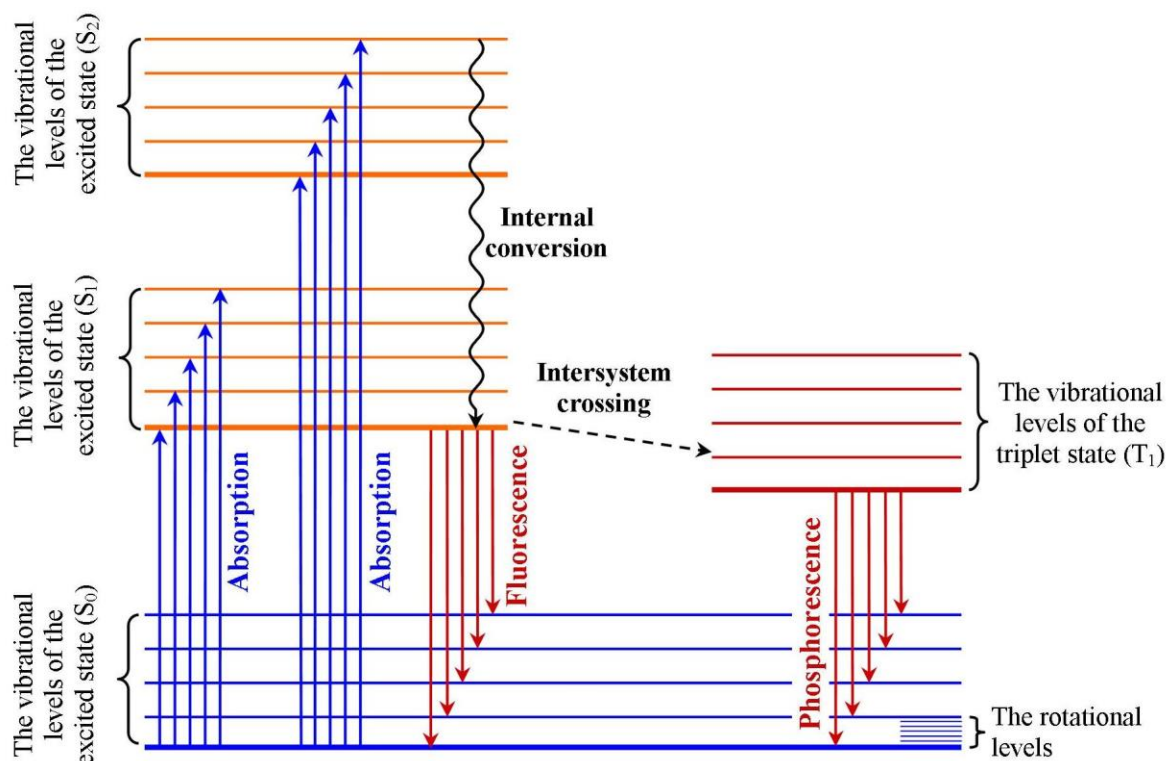


Figure 4.15. Jablonski diagram of electronic, vibrational, and rotational energy levels in a molecule

In absence of excitation (e.g. provided by electromagnetic field) a molecule resides in an electronic state characterized by the lowest possible energy value  $E_0$ . This electronic state is called the *ground state*, and the corresponding energy level is called the electronic ground energy level of the

molecule. Typically, when molecules with closed shell configuration are considered, total spin  $S$  for the ground electronic state of the molecule is equal to zero, i.e. multiplicity is  $2S+1=1$ , and the state is referred to as a singlet state and denoted with symbol  $S_0$  (Fig. 4.15).

When a molecule in the ground state acquires a portion of energy from outside (for example, during absorption of a photon), a transition of electron from one molecular orbital to another occurs and the molecule changes its energy level (i.e. transits) to *the excited electronic state*  $S_i$ , which corresponds to the energy level  $E_i$ . If the transition occurs between the ground  $S_0$  and the first excited  $S_1$  states, this process can be described with the following equation:  $M + h\nu_1 \rightarrow M^*$ ,

where  $M$  and  $M^*$  denote the molecule in the ground and excited electronic states, respectively;  $h\nu_1$  is the amount of the absorbed photon energy,  $h\nu_1 = E(S_1) - E(S_0) = \Delta E_{1e}$

If the molecule absorbs light of a different frequency, such as  $h\nu_2 = E_2 - E_0 = \Delta E_{2e}$ , then transition takes place to the second excited electronic state  $S_2$  (Fig. 4.15).

Excited electronic state has non-equilibrium character, therefore molecules can not stay in this state for a long time. Deactivation of the excited molecule, i.e. loss of the energy gained during the excitation, takes place as a result of several competing processes: spontaneous transition to the ground state accompanied by photon emission (so-called radiative transition); excitation energy transfer (or migration) to another molecule; activation of photochemical reaction (photoisomerization, photodissociation, photoionization, etc.); intersystem crossing – non-radiative process resulting in spin conversion from singlet to triplet states (for example conversion from  $S_1$  to  $T_1$  state); internal conversion process, when the molecule non-radiatively transits to the highly excited vibrational sublevel of low-lying electronic state and the excitation energy is lost (converted to heat) due to energy exchange during collisions with surrounding molecules.

As it was mentioned, one of the routes of the excited state energy deactivation – radiative transition to the ground state – is accompanied by photon emission, and the process of light emission by molecules or atoms during this transition is called *luminescence*. It is essential that luminescence represents the radiation which is not in thermodynamic equilibrium with the source (i.e. the emitting molecule) in contrast to thermal radiation. Luminescence emitted by molecules shows quantum nature of light absorption and emission processes. In general, not only electronic but also vibrational and rotational energy states of the molecule can be changed at the radiative transition (Fig. 4.15). However, it is important to note that electronic transition in a molecule is considered to be the usual condition for luminescence emission to occur.

Depending on a way how the emission is excited, luminescence can be divided into several types: *photoluminescence* (excitation by light), *electroluminescence* (excitation by electrical discharge or electrons), *chemiluminescence* (excitation as a result of chemical reactions), etc.

Almost all atoms, majority of molecules in gas phase, as well as ions of rare earth metals, and aromatic organic molecules in solutions exhibit luminescent properties.

Study of luminescence properties of molecules in solutions provides important information on structure of the molecules and their microenvironment, nature of intra- and intermolecular interactions and kinetics.

Depending on multiplicity of electronic states, between which the radiative transition occurs, and emission decay duration ( $\tau$ ), luminescence is divided into *fluorescence* (transition between two singlet states  $S_1 \rightarrow S_0$  with typical  $\tau = 10^{-6} \div 10^{-10}$  s) and *phosphorescence* (transition between triplet and singlet states  $T_1 \rightarrow S_0$  with  $\tau = 10^{-5} \div 10^2$  s). It is noteworthy that singlet and triplet energy states have significantly different quantum nature and transitions between states of different multiplicity are forbidden by the selection rules of electric dipole transitions. However, the selection rules do not hold strictly and transition between triplet and singlet states is possible but at drastically lower rate. Singlet states have multiplicity equal to 1 and they do not split when magnetic field is applied in contrast to triplet states (multiplicity is 3) which are split in three sublevels in magnetic field.

Molecules containing heavy atoms such as bromine and iodine are frequently phosphorescent. The heavy atoms facilitate intersystem crossing and thus enhance phosphorescence efficiency.

**2. The main parameters of luminescence.** *Luminescence spectrum* is dependence of the radiation intensity distribution, emitted by a molecule or other luminophore, on wavelength or frequency of light. The spectrum represents energetic characteristic of the luminescence. Its position in the frequency scale is determined by difference of energies between the excited and ground states ( $h\nu_{lum} = E_1 - E_0$ ). While luminescence spectra of molecules in gas phase are comprised of series (bands) of narrow spectral lines, the spectra of complex molecules in solutions are generally appeared as a single broad band with full width at half maximum (FWHM) of tens of nanometers. The considerable width of the luminescence band is related to the fact that each electronic energy level of the molecule has a corresponding set of vibrational and rotational sub-levels, which are significantly broadened due to intermolecular interactions with the environment.

Emission of light is a statistical process, relevant for ensemble of many luminescing particles. If the ensemble is comprised of the same type of molecules then each excited molecule of the ensemble is characterized by a certain probability of emission (or rate of emission)  $A$ , that does not depend on how much time it spent in the excited state. In this case, after the preliminary excitation, the number of molecules in the excited state should decay exponentially with time

$$N = N_0 \exp(-At), \quad (4.29)$$

where  $N_0$  is the number of excited molecules at the initial time corresponding to  $t = 0$ . The time (or lifetime) spent in the excited state for individual luminophores of the total ensemble of molecules is different, but the ensemble averaged time spent in the excited state has specific value  $\tau = 1/A$ . This magnitude is used to characterize *lifetime of the excited state*. Since the intensity of emission is proportional to the number of excited luminophores, therefore the emission intensity should follow the same rule:

$$I = I_0 \exp\left(-\frac{t}{\tau}\right). \quad (4.30)$$

Therefore, luminescence decay lifetime  $\tau$  has meaning of time during which the intensity of emission decreases by  $e \approx 2.72$  folds. It is noteworthy that the same value  $\tau$  is used to characterize duration of the emission process (emission decay) by ensemble of molecules and to characterize lifetime of the excited state.

The luminescence efficiency is described using special terms - quantum and energy yields. Quantum yield of luminescence ( $B_{lum}$ ) is defined as ratio of number of luminescence photons  $N_{lum}$  to number of photons absorbed by molecules of the luminescing substance  $N_{abs}$  for the same period of time:

$$B_{lum} = \frac{N_{lum}}{N_{abs}}. \quad (4.31)$$

*Fluorescence quantum yield*  $B_{fl}$  is defined as the ratio of the number of photons emitted through fluorescence to the number of photons absorbed:

$$B_{fl} = \frac{N_{fl}}{N_{abs}} \quad (4.32)$$

The phosphorescence quantum yield is defined in similar way.

The luminescence quantum yield can not exceed 1, and in most cases it is less than 1. Thus, quantum yield represents emission efficiency of a given luminophore and can be considered as probability for the molecule, that absorbed a photon, to emit a luminescence photon.

*Luminescence energy yield* ( $\chi$ ) is the ratio of the total light energy emitted during luminescence to the energy of absorbed light:

$$\chi_{lum} = \frac{\int I_{lum}^{\nu} d\nu}{\int I_{abs}^{\nu} d\nu} \quad (4.33)$$

Since part of the photon energy absorbed by the molecule is converted to vibrational motion (and dissipated in heat), the energy yield is usually lower than the quantum yield and luminescence spectrum position on frequency scale is shifted to lower frequencies in comparison to absorption spectrum.

**3. Principle laws of luminescence. Independence of fluorescence spectrum on excitation light wavelength.** As can be seen from Fig. 4.15 excitation of a molecule in the ground state with photons of different energy will result in transition to different electronic, vibrational, and rotational levels of the excited states. Therefore, one can expect that luminescence spectrum will depend on excitation wavelength. This is indeed true for molecules in gas phase.

However, in the case of substances in condensed state (for instance, in solutions) it turned out that the same fluorescence emission spectrum is generally observed irrespective of the excitation wavelength. This behavior for emission spectrum in condensed states is generally known as Kasha's rule, but it should be noted that independence of fluorescence quantum yield on excitation wavelength for molecules in condensed state was first reported by S.I. Vavilov.

Explanation of Kasha's rule is the following. Upon excitation into higher electronic and vibrational levels, the excess energy is quickly dissipated due to interactions with surrounding solvent molecules leaving the fluorophore in the lowest vibrational level of the first excited electronic state  $S_1$  (Fig. 4.15). This nonradiative process of energy dissipation is called internal conversion and its typical timescale is  $\sim 10^{-12}$  s, which is usually much shorter than lifetime  $\tau$  (typical value is  $\sim 10^{-9}$  s) of the first excited  $S_1$  state.

When rapid internal conversion processes are over, the further deactivation processes originate from the lowest vibrational level of the excited state  $S_1$ . Therefore, emission predominantly occurs from the lowest vibrational level of the excited state  $S_1$  and fluorescence spectra are usually independent of the excitation wavelength. From the practical point of view, this allows to use wide spectral ranges for excitation and neglect spectral composition of the exciting light.

**Stokes-Lommel law.** In 1852 G. Stokes formulated the rule stating that the luminescence light always has a longer wavelength in comparison with the light used for excitation. Although usually this statement is valid (see Fig. 4.15.), in many cases the Stokes rule does not work. The reason for this is partial overlap of absorption and luminescence spectra for many substances (Figure 4.16).

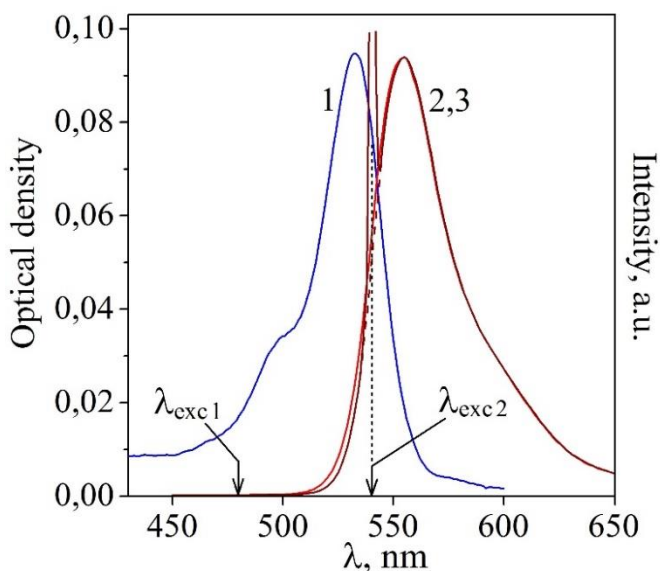


Figure 4.16. Absorption spectrum (1) and fluorescence spectra (2, 3) of rhodamine 6G in ethanol excited at 460 nm ( $\lambda_{\text{exc1}}$ ) and 540 nm ( $\lambda_{\text{exc2}}$ )

According to the Stokes' rule only part of the emission spectrum, which is located at longer wavelengths (or lower frequencies) than the excitation light wavelength  $\lambda_{ex}$ , could be observed. However, in many cases, we can observe the full spectrum of fluorescence, i.e. the Stokes' rule is violated. Part of the fluorescence spectrum within wavelengths range  $\lambda < \lambda_{ex}$  is called *anti-Stokes*.

In Fig. 4.16. *anti-Stokes* fluorescence of rhodamine 6G excited at  $\lambda_{ex2} = 540$  nm is observed in spectral range 500-530 nm. Appearance of anti-stokes fluorescence can be explained by presence of certain vibrational energy in molecules, that absorb photon of excitation light, i.e. absorption takes place from excited vibrational levels of the ground electronic state. Population of these vibrational levels at certain temperature can be estimated using Boltzmann distribution. Sum of vibrational energy of the molecule and energy of the absorbed photon makes possible emission of photons with higher energy than the excitation light, thus explaining anti-stokes part of the fluorescence spectrum.

E. Lommel proposed the correction to the Stokes' rule: *emission spectrum as a whole and its maximum are always shifted in comparison with the absorption and its maximum toward longer wavelengths (or lower frequencies)*.

The shift between maxima of absorption and emission spectra is called Stokes shift. It is important to note that energy losses between excitation light and emission are observed universally for fluorescent molecules in solution. One common cause of the Stokes shift is the rapid decay to the lowest vibrational level of electronic  $S_1$  state due to internal conversion processes. Additionally, upon fluorescence emission fluorophores generally transit not to the lowest vibrational level of the ground  $S_0$  state but to higher vibrational levels, resulting in further loss of excitation energy by thermalization of the excess vibrational energy. In addition to these effects, fluorophores can display additional Stokes shifts due to solvent effects, excited-state reactions, etc.

**4. The mirror-image rule by Levshin.** For a wide range of dye solutions, particularly for aromatic hydrocarbons in solvents of low polarity, one can observe manifestation of so-called mirror-image rule established by V.L. Levshin. According to this rule (Fig. 4.17) absorption and fluorescence spectra, depicted in frequency scale, are symmetric to each other with regards to the line passing through the intersection point of both spectra and perpendicular to the frequency axis, i.e.,

$$\nu_{fl} + \nu_{abs} = 2\nu_{0-0}, \quad (4.34)$$

where  $\nu_{abs}$  is the frequency in the absorption spectrum,  $\nu_{fl}$  is corresponding frequency of the emission spectrum,  $\nu_{0-0}$  is the frequency of the symmetry line, which has meaning of frequency of the purely electronic transition from  $\nu=0$  vibrational level of the ground electronic state  $S_0$  to  $\nu'=0$  vibrational level of the first excited electronic state  $S_1$ .

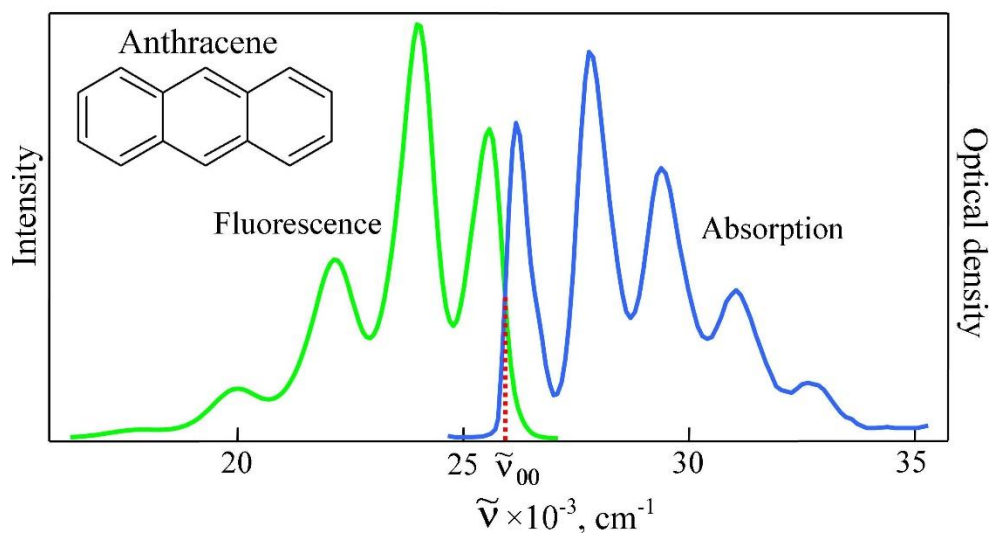


Figure 4.17. An example of the mirror image rule for absorption and fluorescence spectra of anthracene. Spectra of anthracene in cyclohexane are shown in wavenumber scale  $\tilde{\nu} = 1/\lambda$ .

The generally symmetric nature of the absorption and fluorescence spectra is a result of the same transitions being involved in both absorption and emission, and the similar spacings between vibrational energy levels of  $S_0$  and  $S_1$  states. Usually, these energy levels are not significantly altered by the different electronic distributions of  $S_0$  and  $S_1$  states. In absorption spectrum of anthracene (Fig. 4.17.) one can see sharp distinct peaks related to population of vibrational levels in the molecule. These peaks are due to transitions from the lowest vibrational level of the ground  $S_0$  state to higher vibrational levels of the excited  $S_1$  state. Upon the deactivation of the excited electronic state the fluorophore can return to any of the ground state vibrational levels. If these vibrational energy levels have similar spacing in comparison to those in the  $S_1$  state then fluorescence emission spectrum shows the same pattern of peaks as the absorption spectrum.

**5. Luminescence depolarization.** Luminescence depolarization can provide information on orientation of the emitting centers and their mobility during the timescale of luminescence decay. This method is more appropriate to study fluorescent molecules since timescales of fluorescence decay and rotational diffusion of these molecules in solutions are usually comparable ( $\sim$ ns range). Origin of *anisotropy or polarization of fluorescence* emitted by a molecule is related to a well-defined spatial orientation of transition dipole moments (or quadrupole moments if dipole transition is forbidden) within molecular framework which are responsible for radiative transition. Therefore, excitation of dye molecules using polarized light can result in emission of the polarized fluorescence.

Quantitatively fluorescence polarization is characterized by *degree of polarization* or *anisotropy*. Let's consider that incident excitation light, linearly polarized in YZ plane, is directed along Y axis (Fig. 4.18). If we conduct measurements of fluorescence intensity polarized along XZ and XY planes at right angle to the direction of the exciting light propagation (i.e. along X axis, Fig. 4.18)), the degree of polarization  $P$  and anisotropy  $r$  respectively will be equal to

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \cdot r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad (4.35)$$

In equation 4.35  $I_{\parallel}$  and  $I_{\perp}$  are the intensities of fluorescence components linear-polarized in XZ and XY planes, respectively.

Relations between  $P$  and  $r$  values are the following  $r = \frac{2P}{3-P}$  or  $P = \frac{3r}{2+r}$

Let's assume that molecules of the luminescent substance are located in a solvent with high viscosity, which prevents their rotation during the time between light absorption and emission. Additionally, transition dipole moments  $\mu_{ab}$  between electronic states **A** and **B** for ensemble of the dye molecules are randomly oriented in space.

Therefore, only molecules with non-zero value of the transition dipole moment  $\mu_{ab}$  component along Z axis will be able to absorb the linearly polarized light (Fig. 4.18), become excited, and then emit the fluorescence photon.

In general, transition dipole moments for absorption and emission processes can be different and have non-identical orientation, for instance when electronic  $S_0 \rightarrow S_2$  transition takes place during the absorption and  $S_1 \rightarrow S_0$  – during the emission.

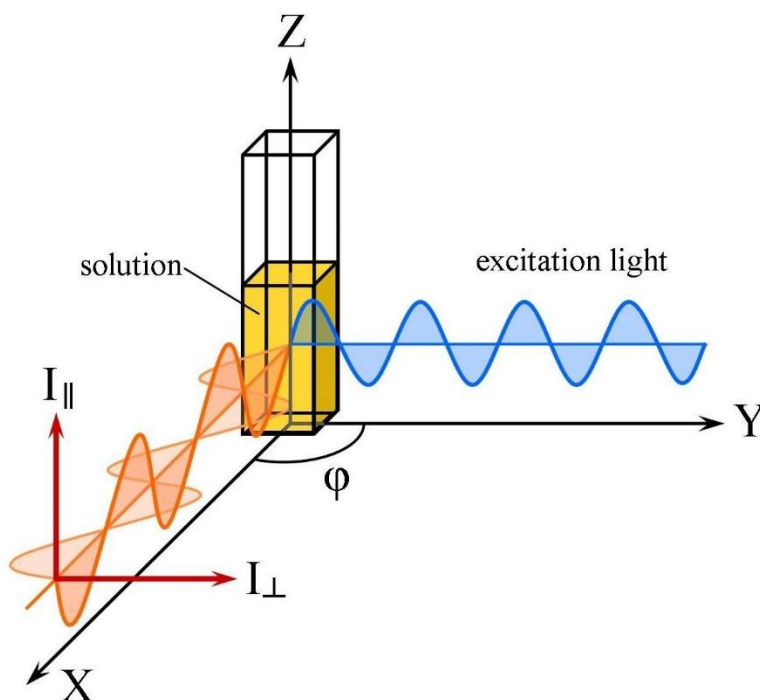


Figure 4.18. To the determination of the luminescence polarization degree

Degree of luminescence polarization depends on angle  $\alpha$  between transition dipole moments for absorption and emission. One can show that



$$P = \frac{3 \cos^2 \alpha - 1}{\cos^2 \alpha + 3}. \quad (4.36)$$

Since the angle  $\alpha$  can be in the range from  $0^\circ$  to  $90^\circ$ , the corresponding value of  $P$  is enclosed within range from  $1/2$  to  $-1/3$ . Similar limiting values for fluorescence anisotropy  $r$  are equal to 0.4 and -0.2, correspondingly. However, in fact these limiting values are rarely achieved due to various factors leading to depolarization of emission: rotational Brownian motion of molecules, energy transfer between molecules, intramolecular motions, etc.

Equation (4.32) allows to explain why fluorescence anisotropy or polarization degree can be dependent on frequency of the exciting light. By changing frequency or wavelength of the excitation one may induce transitions to different excited electronic states ( $S_1$ ,  $S_2$ , ...), while emission due to Kasha's rule occurs predominantly from the excited  $S_1$  state. Usually orientations of the transition dipoles for  $S_0 \rightarrow S_1$ ,  $S_0 \rightarrow S_2$ , ... transitions are not collinear, i.e. angles  $\alpha$  could have different values resulting in different magnitudes of polarization degree.

Besides the dependence on exciting light frequency, degree of polarization depends on excited state lifetime  $\tau$ , solvent viscosity  $\eta$ , temperature  $T$  and hydrodynamic volume  $V$  of the dye molecule. The following equation proposed by F. Perrin and V.L. Levshin relates polarization degree (or anisotropy) of fluorescence with abovementioned parameters:

$$\frac{1}{P} = \frac{1}{P_0} - \left( \frac{1}{P_0} - \frac{1}{3} \right) \tau \frac{k_B T}{V \eta}, \quad (4.37)$$

where  $k_B$  is the Boltzmann constant,  $P_0$  is the limiting value of the polarization degree determined by eq.(4.36) for the case of infinitely high viscosity  $\eta \rightarrow \infty$ .

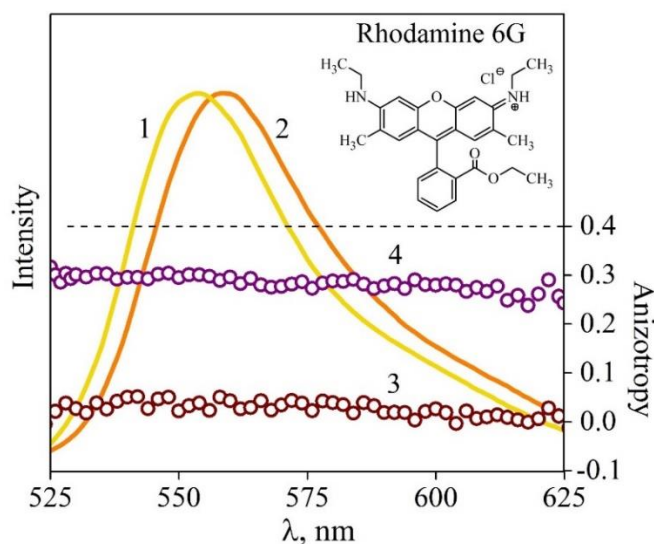


Figure 4.19. Spectra of Rhodamine 6G fluorescence in glycerol (1) and ethanol (2) as well as fluorescence anisotropies in glycerol (4) and ethanol (3) at room temperature

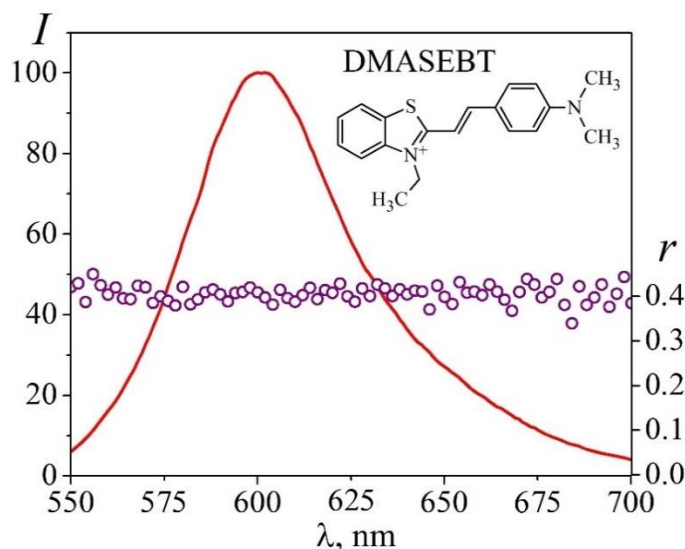


Figure 4.20. Fluorescence spectrum and anisotropy for DMASEBT in ethanol at room temperature

Figure 4.19. shows dependence of fluorescence anisotropy on emission wavelength for the popular dye rhodamine 6G dissolved in ethanol and glycerol. For glycerol with viscosity of  $\sim 1500$  mPa·s anisotropy of rhodamine 6G fluorescence has high value ( $r \approx 0.3$ ) within the whole fluorescence spectrum; in contrast, in the low-viscous ethanol with  $\eta \sim 1$  mPa·s the anisotropy is close to zero. It means that rotational diffusion of the excited Rhodamine 6G in ethanol due to Brownian motion occurs much faster than in glycerol and it is able to completely reorient ensemble of the photoexcited molecules by the time of photon emission and resulting in loss of fluorescence anisotropy.

It should be mentioned that the anisotropy strongly depends on duration of fluorescence decay. Fig. 4.20. shows fluorescence anisotropy for trans-2-[4-(dimethylamino) styryl]-3-ethyl-1,3-benzothiazolium perchlorate (DMASEBT) in ethanol. Despite the low viscosity of the solvent, fluorescence anisotropy is close to the limiting value  $r \approx 0.4$  for DMASEBT molecule. This phenomenon can be easily understood if we take into account the fact that its fluorescence decay lifetime has very short value of  $\sim 50$  ps. Indeed, using effective radius  $a = 4.6$  Å for DMASEBT molecule in accordance with Perrin-Levshin equation one can get  $r \approx 0.38$ , which agrees well with experimental data.

### Time-resolved fluorescence spectroscopy

The most important characteristics of fluorescence are the spectrum, the duration of its decay, quantum efficiency and anisotropy. Most practical applications of fluorescence are connected with the registration of emission spectra (the so-called steady-state fluorescence). However, the most

important progress in fluorescence spectroscopy is associated with using methods based on the measurement of decay duration (time-resolved fluorescent spectroscopy). The fluorescence (or in general photoluminescence) lifetime is characteristic property for each fluorescent or phosphorescent molecule and thus can be used to characterize a sample. It is, however, also influenced by the environment and its chemical composition. Additional processes like Förster Resonance Energy Transfer (FRET), quenching, charge transfer, solvation dynamics, or molecular rotation due to diffusion also have an effect on the decay kinetics. Lifetime changes can therefore be used to gain information about the local chemical environment or to follow reaction mechanisms.

### **Methods of measurement of fluorescence decay duration**

There are two most common ways of measuring the fluorescence decay duration (Berezin and Achilefu 2010): 1) phase-modulation techniques using modulated laser radiation for the excitation (frequency-domain); 2) methods based on the study of the temporal response of the system after the excitation by a short light pulse (time-domain or pulse fluorometry).

In the first case, one can use monochromatic radiation of laser sources, subjected to sinusoidal modulation (Gratton and Barbieri 1986; Hedstrom, Sedarous et al. 1988; Sipior, Carter et al. 1996; Vos, Strobbe et al. 1997). As a result of the fact that the fluorescence has a finite duration, the phase of the recorded signal response is shifted relative to the excitation one by a certain angle. Furthermore, the emission has a smaller modulation depth compared with the excitation light. As a result, by measuring the phase shift and the modulation depth (demodulation index) we calculate the fluorescence decay duration. The limit of time resolution ( $\sim 0.1$  ns) is determined by the maximum modulation frequency that usually does not exceed 10 GHz. The method of phase modulation fluorometry allows to register the fluorescence decay kinetics described by one or two exponentials. Using this technique to study more complicated cases of decay is difficult.

Depending on operation principles the following methods are distinguished for time-domain fluorometry: time-correlated single-photon counting (TCSPC), stroboscopic (James, Siemiarz et al. 1992; Matthews, Summers et al. 2006), streak camera (Graf, Buehler et al. 1994; Becker 2012), non-linear up-conversion (Mataga, Shibata et al. 2000; Peon and Zewail 2001).

The TCSPC (O'Connor and Phillips 1984; Birch and Imhof 2002; Becker, Bergmann et al. 2004) method is the most widely used due to the large dynamic range of measurements and the possibility of studying the kinetics of luminescence of samples with low fluorescence quantum yield. In accordance with TCSPC method, in order to excite fluorescence we use a pulsed light source emitting short light pulses that are directed to the object under study. Due to the absorption, some fluorophore molecules transit to an excited state and then emit fluorescence. If the sample consists of molecules of the same type, in the simplest case, the function of fluorescence intensity decay may be represented by a decaying exponential

$$f(t) = \alpha_0 e^{-\frac{t}{\tau}}, \quad (4.38)$$

where the pre-exponential factor  $\alpha_0$  – the intensity at the moment of time  $t = 0$ ,  $\tau$  – the duration of the fluorescence decay. If fluorescence is emitted by several types of non-interacting emitters, the number of which is  $n$ , then the function  $f(t)$  is described by the sum of the  $n$  exponents

$$f(t) = \sum_{i=1}^n \alpha_{0i} e^{-\frac{t}{\tau_i}} \quad (4.39)$$

Sometimes, when fluorescent systems exhibit structural and microenvironment heterogeneity (for example, fluorescent probes embedded in biopolymers, fluorescent semiconductor nanocrystals, etc.), it is more suitable and convenient to use continuous distribution of the emitting centers on decay lifetime  $\alpha(\tau)$  (typically we use linear combination of distributions such as Gaussian, Lorentz, etc.). In this case function  $f(t)$  can be represented by the following equation:

$$f(t) = \int \alpha(\tau) e^{-\frac{t}{\tau}} d\tau \quad (4.40)$$

Then the expression will correspond to discrete representation of the distribution of emitting centers on lifetimes  $\tau$ .

The main purpose of the pulse fluorometer is to measure the fluorescence decay curve and determine the function  $f(t)$ .

### **The operating principle of TCSPC fluorometer**

Let's consider operating principle of a pulse fluorometer, based on the time-correlated single photon counting method (Maskevich, Stsiapura et al. 2013). Typical scheme is shown in Fig. 4.21. It consists of such elements as: a pulsed light source (Pulsed source), monochromators, sample compartment (Sample), where the sample (usually a cell with solution of the studied compound) is placed, photodetectors working in single photon counting mode (photomultiplier tubes, PMT or single-photon avalanche diodes, SPADs).

Pulsed diode lasers delivering light pulses with picosecond duration are widely used in TCSPC instruments for excitation of the sample. They provide light pulses with highly stable temporal properties and variable repetition, however, at single wavelength only. Therefore, in order to be able to excite at different wavelengths one needs a set of diode lasers. Sometimes gas-discharge lamps are used as light sources although typical widths of generated light pulses are longer (subns- and ns- ranges).

Simultaneously with the excitation light produced by the Pulsed Source a short electrical pulse (denoted as START pulse) is generated, which triggers time-to-amplitude converter (TAC). The excitation light is absorbed by sample molecules and fluorescence photons of a specific spectral range are collected by photodetector (usually PMT working in single photons counting mode). When single photon of fluorescence reaches the photodetector, this event is registered and an electrical signal is generated – so-called STOP pulse. Intensity of the excitation light is chosen in such a way that only single photon of fluorescence is registered per single excitation pulse. TCSPC

method is based on repetitive measurements of time delays between START and STOP pulses (Fig. 4.21), i.e. between the excitation time and the time of fluorescence photon arrival. One can show that histogram of these time delays corresponds to temporal kinetics of fluorescence decay for the studied dye compound.

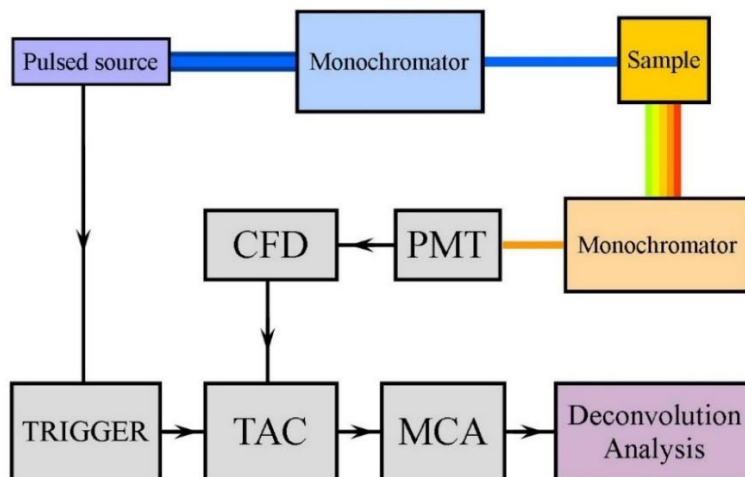


Figure 4.21. Scheme of typical TCSPC fluorometer

Due to statistical nature of signal amplification in photodetector the electrical STOP pulses could have significant amplitude variations (sometimes amplitude of START pulses can be also affected). This leads to some uncertainty in time when the pulse arrives to TAC if triggering by constant discrimination threshold is employed (jitter effect, i.e. pulse arrival time is not measured correctly). To reduce this uncertainty, so-called constant fraction discriminators (CFD) are used, which produce standard pulses with sharp front edges, synchronized with the time when STOP (or START) pulse reaches specific fraction of the amplitude.

Thus, signal from the PMT output goes to CFD and, after the conversion, to the input of time-to-amplitude converter (TAC). TAC produces voltage of certain amplitude, proportional to the time interval between START and STOP pulses, which is analyzed by multichannel analyzer MCA. Usually CFD, TAC, MCA are integrated in a single circuit board that can be installed into extension slot of a computer.

Time-to-amplitude converter (TAC, Fig.4.22.) is a critically important unit of TCSPC setup which is responsible for accurate measurement of time interval between START and STOP pulses. Simultaneously with the excitation light from laser (or other light source) the START pulse is generated which initiates charging of TAC capacitor (Fig. 4.22). Voltage on the TAC capacitor increases linearly in time until STOP signal is received indicating that fluorescence photon reached the photodetector. The voltage on the capacitor is directly proportional to the time interval between START and STOP signals and can be measured with an analog-to-digital converter (ADC). The resolution of such a system is in the range of 1 to 10 ps (Józef 2004). Value from ADC is read by multi-channel analyzer (MCA) and count number for the channel, corresponding to voltage interval

of the data point, is increased by one. Then TAC capacitor is discharged during some time which is called “dead time” and cannot be used for measurements. If TAC does not receive STOP pulse during some specified time after the START pulse – so-called TAC conversion time– the measurement is ceased and the capacitor is discharged.

Not more than one pulse from PMT can be registered during the TAC conversion time. Only the first photon arriving to the detector can be registered. Therefore one needs to ensure that not more than one photon of fluorescence could reach the detector per excitation pulse. This requirements can be satisfied if counting rate for STOP pulses is much less than for START pulses; usually it is supposed that STOP/START ratio of <5% is enough.

If measurements of time intervals between START and STOP pulses are repeated many times one can obtain histogram of these time intervals which corresponds to fluorescence decay kinetics. Parameters of fluorescence decay law  $f(t)$  can be determined by analysis of this histogram (see. The data analysis)

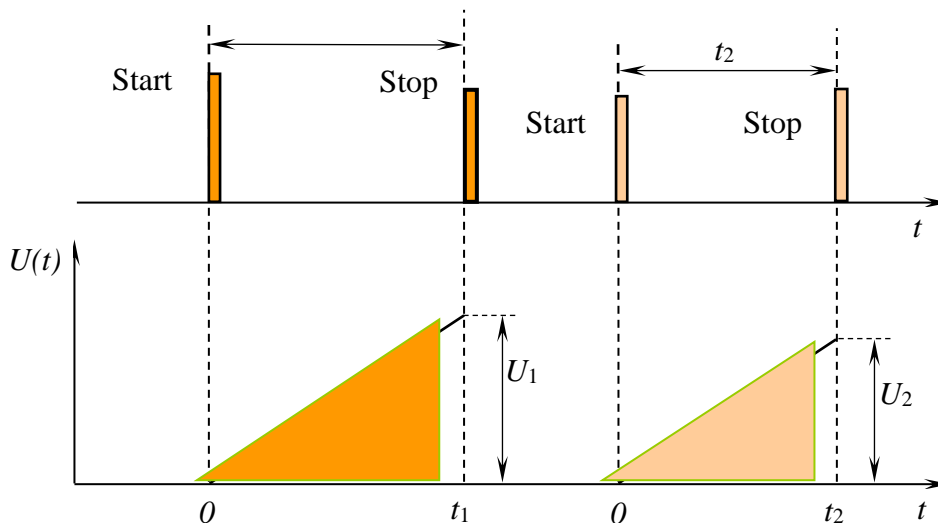


Figure 4.22. Principle of TAC operation

Compounds with monoexponential fluorescence decay, stable and known decay lifetimes so-called fluorescence lifetime standards can be used to determine and calibrate temporal resolution of TCSPC fluorometers. Parameters of some popular standards are given in Table 4.3.

Table 4.3. Fluorescence Lifetime Standards

Nanosecond Lifetime Standards	Lifetime [ns]	Conditions for Lifetime Measurement	Excitation [nm]	Emission [nm]	Ref.
NADH	0.40	0.1 M PB pH 7.4, 20 °C	330-370	400-600	(Lakowicz 2006)

NATA	3.0	0.1 M PB pH 7.4, 20 °C	275	310-400	
p-Terphenyl	1.05	Ethanol	280-320	310-412	
PPD	1.20	Ethanol	240-340	310-440	
PPO	1.40	Ethanol	280-350	330-480	
POPOP	1.35	Ethanol Abs.	280-390	370-540	
Dimethyl- POPOP	1.45	Ethanol	300-400	390-560	
2-Aminopurine	11.34	Water	290	380	
L-Tyrosine	3.27	Water	285	300	
Anthranilic Acid	8.9	Water	290	400	
Indole	4.49	Water	290	360	
Fluorescein, dianion	4.10	NaOH/Water	400	490-520	(Magde, Rojas et al. 1999)

PB = phosphate buffer

NATA = N-Acetyl-L-tryptophanamide

PPD = 1,5-diphenyl-1,3,4-oxadiazole

PPO = 2,5-diphenyl-oxazole

POPOP = 1, 4-bis(5-phenyloxazole-2-yl)benzene

### *The data analysis*

It should be noted that the experimentally detected signal  $F(t)$  is a convolution of the time profile of excitation pulse  $L(t)$ , the response function of registration system  $H(t)$ , and the fluorescence decay function  $f(t)$  of the sample molecule:

$$F(t) = f(t) \otimes L(t) \otimes H(t). \quad (4.41)$$

Equation (4.36) is usually written in integral form:

$$F(t) = f(t) \otimes P(t) = \int_0^t P(t') f(t - t') dt', \quad (4.42)$$

where  $P(t)$  is convolution of the excitation pulse and the response function of registration system, the so-called instrument response function (IRF) of fluorometer. When full width at half-maximum (FWHM) of IRF function  $P(t)$  is much less than typical duration of the decay law function, one can neglect contribution from IRF and consider that  $F(t) \approx f(t)$ . However in many situations this is not the case, temporal durations of  $P(t)$  and  $f(t)$  are comparable and it is necessary to apply deconvolution operation to determine kinetics parameters of the fluorescence decay law  $f(t)$ .

Thus, in order to get  $f(t)$  function one needs to register temporal dependence of experimental fluorescence decay curve  $F(t)$  and IRF function  $P(t)$ . Recovery of  $f(t)$  function from convolution (4.37) in the presence of noise is regarded as ill-posed mathematical problem (Bakushinsky and Goncharsky 1989; Tikhonov, Leonov et al. 1995) and requires special methods for its treatment.

The important feature of TCSPC method is the known noise statistics (number of counts in each channel has Poissonian distribution) of the collected experimental data  $F(t_i)$  and  $P(t_i)$ . Information about the noise statistics can be explicitly taken into account in data analysis and while testing statistical hypotheses about functional form of  $f(t)$ .

As it was mentioned, fluorescence decay law  $f(t)$  is often parametrized and represented as a linear combination of exponentials (Eq 4.37). In this case deconvolution of (eq. 4.42) and determination of parameters  $\alpha_{0i}$  and  $\tau_i$  in Eq. 4.37 are usually carried out using iterative reconvolution method, variant of the nonlinear least squares method, (Johnson and Frasier 1985; Bevington and Robinson 2003).

Goodness of fit is estimated by  $\chi^2$  criterion, standard deviation of weighted residuals, plots of weighted residuals and their autocorrelation function. Fig. 4.23. shows plots with fluorescence decay kinetics of PPO and tryptophan fitted by single and biexponential decay functions.

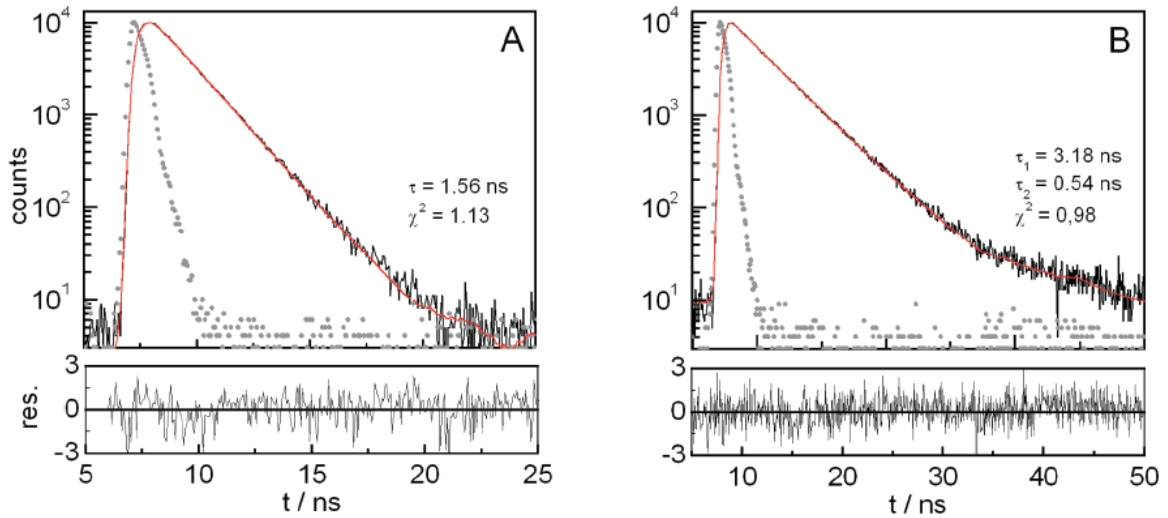


Figure 4.23. Fluorescence decay (black curves) of (A) PPO in degassed ethanol and (B) Trp in phosphate buffer at pH 7.0 and the corresponding IRFs (dotted curves). The fitted decays (red



curves) and the weighted residuals are for (A) a single-exponential and (B) a 2 double-exponential fit. The best-fit values for the decay times and  $\chi^2$  –values for each fit are noted. All measurements were done at 20°C with a time resolution of 64 ps.

([https://www.picoquant.com/images/uploads/page/files/7354/appnote\\_tr\\_proteins.pdf](https://www.picoquant.com/images/uploads/page/files/7354/appnote_tr_proteins.pdf))

Usually not just one fluorescence decay curve is measured during an experiment but series of fluorescence decay curves at slightly different experimental conditions (e.g. at various wavelengths) are recorded. There is an approach referred to as global analysis (GA) (Beechem, Knutson et al. 1983; Beechem, Ameloot et al. 1985) that allows to analyze and fit all the recorded fluorescence decay curves simultaneously in the framework of the same model using the common parameters. Selection of a model that describes behavior of the fluorescing molecular system is equivalent to definition of relationships among the kinetics parameters for the individual decay curves. By treating entire set of data simultaneously, one can reduce total number of the fitting parameters, and thus significantly increase the reliability and statistical accuracy of results.

Moreover using the GA method, one can test hypotheses about applicability of specific physical models to describe kinetics and behavior of the system under study. On the other hand, *a priori* information about the system under study, obtained by other methods (e.g. molecular modeling), can be explicitly used to select a suitable physical model and establish links and relationship among the parameters of the fluorescence kinetics. Thus, GA method allows to perform comprehensive analysis of time-resolved fluorescence, its spectral and kinetic properties while taking into account *a priori* information. This approach is particularly useful for studying objects with structural and dynamical heterogeneity like biological macromolecules and their complexes.

To demonstrate capabilities of GA method, let's consider results of fluorescence decay kinetics for pyrene in liposomes (spherical vesicles with a membrane composed of lipid bilayer). Pyrene is a widely used fluorescent probe to study biological membranes, in particular their viscosity and heterogeneity. Emission spectrum of pyrene in liposomes from egg lecithin is shown in Fig. 4.24.

The fluorescence spectrum of pyrene monomer, located in the range of 360-420 nm, is characterized by the pronounced vibrational structure. Solubility of pyrene (hydrocarbon molecule) in water is very low and the dye molecules are practically confined within the lipid bilayer of liposomes. Therefore, one can see a new structureless band with maximum at 460 nm, which normally appears in the emission spectrum at high concentrations of pyrene molecules in solution (above  $10^{-4}$  M). Emergence of this band is attributed to formation of pyrene excimers. Excimer (or excited dimer) is a short-living dimeric molecule which is stable only in the excited state and is formed due to interaction of excited and non-excited monomer molecules. Since excimerization is a diffusion-controlled process, the ratio monomer / excimer, that can be measured for pyrene molecules, is sensitive to microviscosity and diffusion rate in the immediate

environment of the probe, thus giving information about membrane viscosity and heterogeneity.

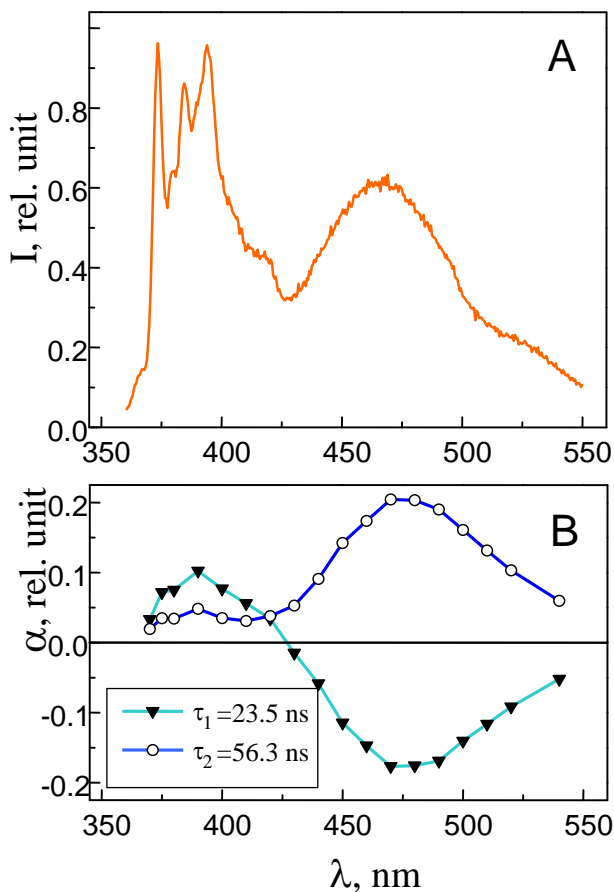
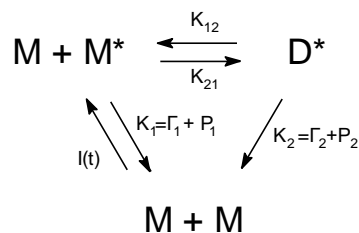


Figure 4.24. Spectral and kinetic data of pyrene fluorescence in liposomes. A – steady-state spectrum of pyrene fluorescence. B - kinetics parameters of pyrene fluorescence obtained by global analysis. Parameters  $\tau_1$  and  $\tau_2$  are considered as common for all decay curves at different emission wavelengths.  $\lambda_{\text{exc}} = 337$  nm

Photophysical processes that are induced by photoexcitation of pyrene molecule can be approximately described using the following kinetic scheme:



where  $D^*$ ,  $M$ ,  $M^*$  - pyrene excimer, pyrene monomer in the ground (non-excited) and the excited states, respectively.  $K_1$ ,  $K_2$  – deactivation rate constants ( $\Gamma_i$  and  $P_i$  – radiative and non-radiative rate

constants, respectively) for the excited states of pyrene monomer and excimer,  $K_{2I}$ ,  $K_{I2}$  – rate constants for the formation and dissociation of excimer.

Set of kinetic parameters  $K_I$ ,  $K_2$ ,  $K_{I2}$ , and  $K_{2I}$  can be determined by analyzing fluorescence decay curves measured at different emission wavelengths. Assuming that fluorescence decay lifetimes  $\tau_1$  and  $\tau_2$  are constant and common for all wavelengths within the emission spectrum range, other parameters of decay kinetics ( $\alpha_i(\lambda_i)$ ,  $\alpha_2(\lambda_i)$ ) can be obtained (Fig. 4.24B).

Moreover, we can reformulate task of the global analysis in such a way that the model parameters ( $K_I$ ,  $K_2$ ,  $K_{I2}$  and  $K_{2I}$ ) would be used as adjustable parameters, and not the parameters describing kinetics of fluorescence decay ( $\alpha_i$ ,  $\tau_i$ ). Using GA we can directly obtain values of the rate constants:  $K_I = 1.2 \cdot 10^7 \text{ s}^{-1}$ ,  $K_2 = 2.0 \cdot 10^7 \text{ s}^{-1}$ ,  $K_{2I} = 2.3 \cdot 10^7 \text{ s}^{-1}$ ,  $K_{I2} = 5 \cdot 10^6 \text{ s}^{-1}$ .

Often, when complex fluorescing systems are studied, e.g. biopolymers, fluorescent semiconductor nanocrystals, etc., it is difficult to select an appropriate model to describe fluorescence decay law  $f(t)$ . Therefore, it is not always possible to set *a priori* functional form for  $f(t)$  or  $\alpha(\tau)$ . In this case, one can use mathematical methods to estimate the most probable solution for  $\alpha(\tau)$  distribution of the emitting centers on decay lifetimes (or decay rate constants), without any prior assumption about the functional form. One of these methods is based on maximum entropy method (MEM) (Brochon 1994; Steinbach, Ionescu et al. 2002), where the recovered distribution  $\alpha(\tau)$  is determined in process of maximization of functional  $\Psi$  (Livesey and Brochon 1987; Skilling 1989; Maskevich, Stsiapura et al. 2010):

$$\Psi = S - \mu(\chi^2 - 1), \quad (4.43)$$

where  $S$  - entropy function,  $\mu$  - additional variable parameter,  $\chi^2$  – reduced chi-square functional.

Parameter  $\mu$  adjusts contributions of entropy  $S$  and  $\chi^2$  in functional  $\Psi$  value. For small  $\mu \ll 1$  entropy gives the main contribution, and the recovered distribution  $\alpha(\tau)$  approaches the prior model  $m(\tau)$  (usually uniform lifetime distribution is used as the prior model). However, in this case one can see significant discrepancy between the calculated and experimental data. For large  $\mu \gg 1$  there is good correspondence of calculated and experimental data, however, the recovered solution for  $\alpha(\tau)$  distribution is over-structured and often does not have physical meaning. The goal is to find the optimal value for  $\mu$  which allows you to simultaneously achieve two effects: to get the most "smooth" (i.e. the most probable) solution for  $\alpha(\tau)$  distribution with good correspondence of the calculated and experimental data.

Fig. 4.25. represents results of MEM-analysis of fluorescence decay for thioflavin T complex with amyloid fibrils – filamentous insoluble aggregates of proteins, which formation is related to several neurodegenerative disorders. Thioflavin T is a fluorescent probe widely used for detection of amyloid fibrils due to dramatic enhancement of its fluorescence intensity (and decay lifetime) upon binding to these aggregates. Lifetime distributions for Thioflavin T bound to amyloid fibrils at

various stages of their formation are shown. Lifetime distributions have bimodal character and ratio of amplitudes for the short- and the long-lived modes in  $\alpha(\tau)$  is increased by  $\sim 5$  times during fibrils maturation. This ratio is a good indicator of the fibrils "maturity" and can be used to determine the stage of fibrillogenesis.

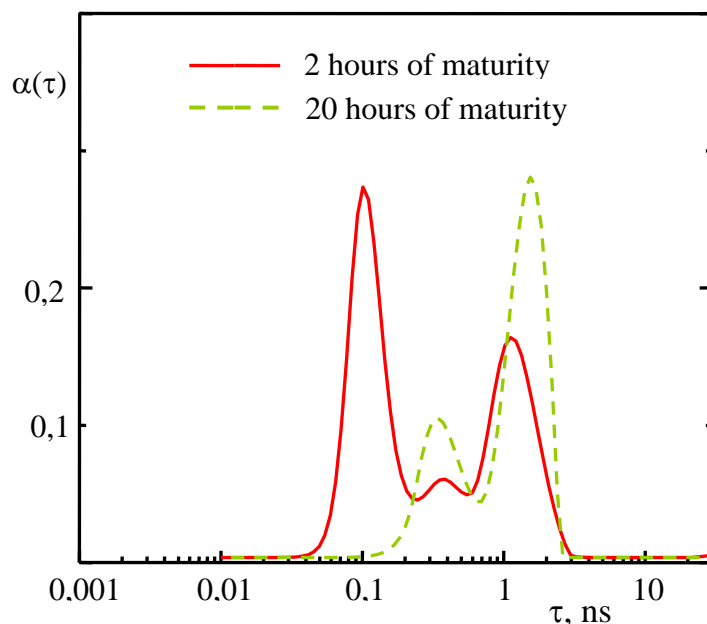


Figure 4.25. Fluorescence lifetime distribution for thioflavin T bound to amyloid fibrils from insulin at different stages of fibrillogenesis

#### 4.1.4 References and bibliography

##### References

- <https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/Spectrpy/InfraRed/infrared.htm>.  
[https://www.researchgate.net/file.PostFileLoader.html?id=4fbf619be39d5e1817000000&assetKey=AS%3A271735000109069%401441798081243](https://www.researchgate.net/file.PostFileLoader.html?id=4fbf619be39d5e1817000000&assetKey=AS%3A271735000109069%401441798081243;);  
<http://www2.ups.edu/faculty/hanson/Spectroscopy/IR/IRfrequencies.html>;  
[https://en.wikipedia.org/wiki/Infrared\\_spectroscopy\\_correlation\\_table](https://en.wikipedia.org/wiki/Infrared_spectroscopy_correlation_table)

1. Bakushinsky, A. and A. Goncharsky (1989). "Ill-posed Problems: Numerical Methods and Applications [in Russian], Izd." MGU, Moscow.
2. Becker, W. (2012). "Fluorescence lifetime imaging—techniques and applications." *Journal of microscopy* 247(2): 119-136.

3. Becker, W., A. Bergmann, et al. (2004). Advanced time-correlated single photon counting techniques for spectroscopy and imaging in biomedical systems. *Lasers and Applications in Science and Engineering*, International Society for Optics and Photonics.
4. Beechem, J. M., M. Ameloot, et al. (1985). "Global analysis of fluorescence decay surfaces: excited-state reactions." *Chemical physics letters* 120(4-5): 466-472.
5. Beechem, J. M., J. R. Knutson, et al. (1983). "Global resolution of heterogeneous decay by phase/modulation fluorometry: mixtures and proteins." *Biochemistry* 22(26): 6054-6058.
6. Berezin, M. Y. and S. Achilefu (2010). "Fluorescence lifetime measurements and biological imaging." *Chemical reviews* 110(5): 2641.
7. Bevington, P. R. and D. K. Robinson (2003). "Data reduction and error analysis." McGraw-Hill.
8. Birch, D. J. and R. E. Imhof (2002). Time-domain fluorescence spectroscopy using time-correlated single-photon counting. *Topics in fluorescence spectroscopy*, Springer: 1-95.
9. Brochon, J.-C. (1994). "[13] Maximum entropy method of data analysis in time-resolved spectroscopy." *Methods in enzymology* 240: 262-311.
10. Graf, U., C. Buehler, et al. (1994). Optimized streak-camera system: Wide excitation range and extended time scale for fluorescence lifetime measurement. *OE/LASE'94*, International Society for Optics and Photonics.
11. Gratton, E. and B. Barbieri (1986). "Multifrequency phase fluorometry using pulsed sources: theory and applications." *Spectroscopy* 1(6): 28-36.
12. Hedstrom, J., S. Sedarous, et al. (1988). "Measurements of fluorescence lifetimes by use of a hybrid time-correlated and multifrequency phase fluorometer." *Biochemistry* 27(17): 6203-6208.
13. James, D. R., A. Siemiarzuk, et al. (1992). "Stroboscopic optical boxcar technique for the determination of fluorescence lifetimes." *Review of scientific instruments* 63(2): 1710-1716.
14. Johnson, M. L. and S. G. Frasier (1985). "[16] Nonlinear least-squares analysis." *Methods in enzymology* 117: 301-342.
15. Józef, K. (2004). "Review of methods for time interval measurements with picosecond resolution." *Metrologia* 41(1): 17.
16. Lakowicz, J. R. (2006). *Principles of Fluorescence Spectroscopy*. NY, Springer.
17. Livesey, A. and J. Brochon (1987). "Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method." *Biophysical journal* 52(5): 693-706.
18. Magde, D., G. E. Rojas, et al. (1999). "Solvent dependence of the fluorescence lifetimes of xanthene dyes." *Photochemistry and Photobiology* 70(5): 737-744.
19. Maskevich, A., V. Stsiapura, et al. (2010). "Analysis of fluorescence decay kinetics of thioflavin t by a maximum entropy method." *Journal of Applied Spectroscopy* 77(2): 194-201.

20. Maskevich, A. A., V. I. Stsiapura, et al. (2013). "Hardware and software complex for fluorescence decay studies." *Vesnik GrSU* 3(159): 107-119.
21. Mataga, N., Y. Shibata, et al. (2000). "Internal conversion and vibronic relaxation from higher excited electronic state of porphyrins: Femtosecond fluorescence dynamics studies." *The Journal of Physical Chemistry B* 104(17): 4001-4004.
22. Matthews, D. R., H. D. Summers, et al. (2006). "Technique for measurement of fluorescence lifetime by use of stroboscopic excitation and continuous-wave detection." *Applied optics* 45(9): 2115-2123.
23. O'Connor, D. and D. Phillips (1984). *Time-correlated single photon counting*, London [etc.]: Academic Press.
24. Peon, J. and A. H. Zewail (2001). "DNA/RNA nucleotides and nucleosides: direct measurement of excited-state lifetimes by femtosecond fluorescence up-conversion." *Chemical physics letters* 348(3): 255-262.
25. Sipior, J., G. M. Carter, et al. (1996). "Single quantum well light emitting diodes demonstrated as excitation sources for nanosecond phase-modulation fluorescence lifetime measurements." *Review of scientific instruments* 67(11): 3795-3798.
26. Skilling, J. (1989). *Classic maximum entropy. Maximum entropy and Bayesian methods*, Springer: 45-52.
27. Steinbach, P. J., R. Ionescu, et al. (2002). "Analysis of kinetics using a hybrid maximum-entropy/nonlinear-least-squares method: application to protein folding." *Biophysical journal* 82(4): 2244-2255.
28. Tikhonov, A. N., A. S. Leonov, et al. (1995). *Nonlinear ill-posed problems*. Moscow, Nauka.
29. Vos, R., R. Strobbe, et al. (1997). "Gigahertz Phase Fluoremetry Using a Fast High-Gain Photomultiplier." *Journal of Fluorescence* 7(1): 33S.

## Bibliography

1. *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*. 3rd edition. 127 pages.

2. <http://www.bipm.org/en/publications/si-brochure/>

### 4.1.5 List of acronyms

BIPM - International Bureau of Weights and Measures

CCQM - Consultative Committee for Amount of Substance – Metrology in Chemistry

CGPM - General Conference on Weights and Measures

CODATA - Committee on Data for Science and Technology

GUM - Guide to the Expression of Uncertainty in Measurement

IAEA - International Atomic Energy Agency

ICSU - International Council of Scientific Unions

IEC - International Electrotechnical Commission

IFCC - International Federation of Clinical Chemistry and Laboratory Medicine

ILAC - International Laboratory Accreditation Cooperation

ISO - International Organization for Standardization

IUPAC - International Union of Pure and Applied Chemistry

IUPAC-CIAAW - International Union of Pure and Applied Chemistry – Commission on Isotopic Abundances and Atomic Weights

IUPAP - International Union of Pure and Applied Physics

JCGM - Joint Committee for Guides in Metrology

JCGM/WG 1 - Working Group 1 of Joint Committee for Guides in Metrology

JCGM/WG 2 - Working Group 2 of Joint Committee for Guides in Metrology

OIML - International Organization of Legal Metrology

VIM - International Vocabulary of Metrology -- Basic and General Concepts and Associated Terms

WHO - World Health Organization

## 4.2 Principles of optical microscopy

### 4.2.1 Introduction

Term “optical microscopy” refers to the set of methods employing light (optical) microscopes for observation and study of small objects, which are invisible for the naked eye.

Human eye represents the natural optical system able to perceive light, distinguish colors in so-called visible range of electromagnetic waves spectrum (400–800 nm) and which is characterized by specific spatial resolution, i.e. minimal distance that can be resolved between two point-like objects that are being observed.

The eye receptors in the retina used for sensing color are the cone cells; another kind of photoreceptor cells – the rod cells – are used for distinguishing levels of intensity (but not the color). These cells are located on the retina inside of the eye. The front of the eye (see Figure 4.26), including the iris, the curved cornea, and the lens are respectively the mechanisms for admitting light and focusing it on the retina.

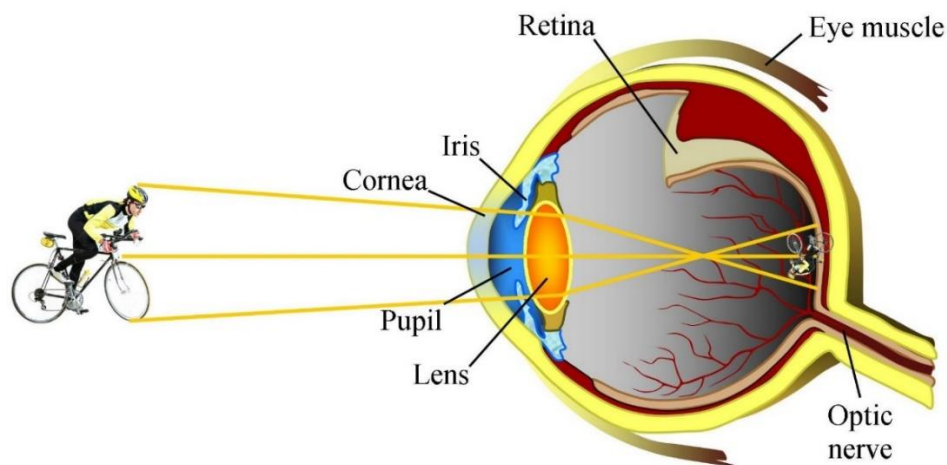


Figure 4.26. Image formation in eye. (the figure has been modified from the source <http://micro.magnet.fsu.edu/primer/anatomy/introduction.html>)

Increasing the solid angle under which an object is seen we increase image of the object on the retina. Therefore, moving the object closer to the eye results in increase of the image magnitude on the retina, which allows to see and resolve smaller details. However, range of the eye accommodation, i.e. adjustment of the crystalline lens element of the eye to alter the refractive power and bring objects under observation into sharp focus, is limited and normally ~25 cm is the minimal distance from the eye where vision is still clear. For an image to be seen clearly, it must spread on the retina at a sufficient visual angle. Unless the light falls on non-adjacent rows of retinal cells, we are unable to distinguish closely-lying details as being separate (resolved). Human eye can resolve two objects if they are located at angular distance  $\sim 1'$  (so-called angular resolving power), which corresponds to minimal separation of two neighboring cone cells in retina  $\sim 2.5 \mu\text{m}$ .



Therefore spatial resolution of the naked eye at ~25 cm distance, where two closely-lying objects are perceived as separate, is ~ 80–100  $\mu\text{m}$ .

#### 4.2.2 Optical magnifications and microscopes

In order to increase angle of the image projection on the retina and resolve smaller details one needs to use optical instruments like magnifying glass or microscope.

Magnifying glass is the simplest optical system from a convex lens having focal length of 10÷100 mm. Linear magnification that can be achieved using magnifying glass is in the range of 2.5÷25. One can significantly increase magnification of the object image using microscope.

The first optical microscope was constructed in Netherlands by the 1620s. Spectacle-makers Zacharias Janssen, Cornelis Drebbel, and Hans Lippershey were mentioned among the likely inventors. The first successful application of optical microscope in scientific research was done by English scientist R. Hook (1665 r.) who discovered cellular structure of animal and plant tissues. The important contribution came from Antonie von Leeuwenhoek who achieved up to 300 times magnification using a microscope of his construction (Figure 4.27) and discovered micro-organisms.

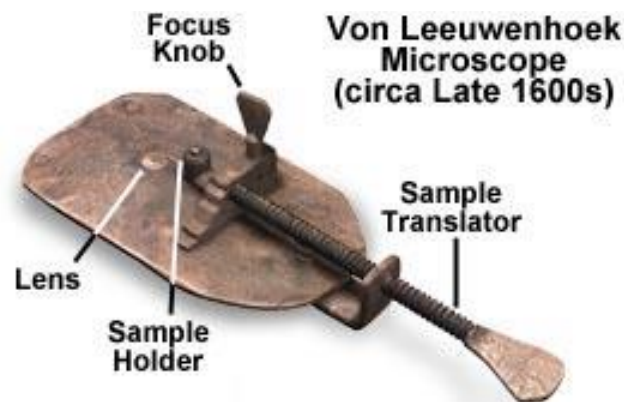


Figure 4.27. Construction of von Leeuwenhoek Microscope. (the image was taken from <http://micro.magnet.fsu.edu/primer/anatomy/introduction.html>)

A microscope is a combination of two optical elements – objective and ocular (or eyepiece) mounted at opposite ends of a tube, known as the body tube. Schematics of light path in typical microscope is shown in Fig. 4.28.

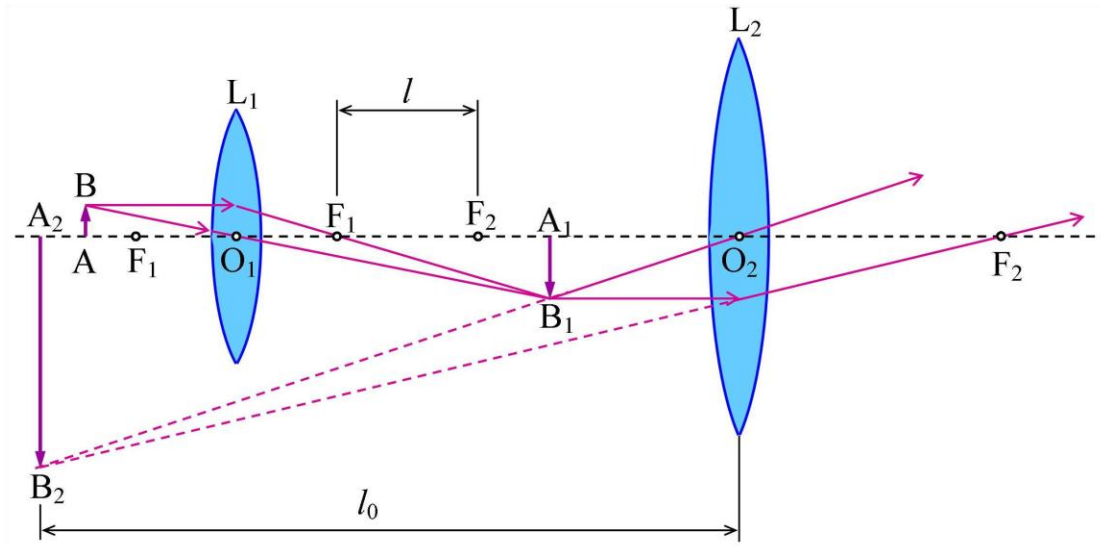


Figure 4.28. Schematics of light ray path in objective  $L_1$  and eyepiece  $L_2$  lenses of typical optical microscope.  $O_1F_1$  and  $O_2F_2$  - focal lengths of the objective and eyepiece lenses respectively. Object  $AB$ , real intermediate image  $A_1B_1$  formed by the objective lens, and virtual image  $A_2B_2$  formed by the eyepiece lens are shown.

The investigated object  $AB$  is usually located at a distance little bit larger than focal length  $F_1O_1$  and objective  $L_1$  produces an intermediate image  $A_1B_1$  which is real, inverted and with size larger than of the initial object  $AB$ . The intermediate image  $A_1B_1$  is further magnified by the eyepiece to produce an image on the retina or camera. For visual inspection the eyepiece lens should be located from  $A_1B_1$  position on a distance smaller than  $O_2F_2$ . When the microscope eyepiece is used for projection on screen or camera, it is necessary to position  $A_1B_1$  on a distance from the eyepiece lens  $L_2$  larger than focal length  $O_2F_2$  to get real (not virtual) image  $A_2B_2$ .

#### 4.2.3 Magnification of a microscope

One of the most important parameters of microscope is linear magnification which represents the ratio of the image size and the object size

$$M = \frac{A_2B_2}{AB} \approx \frac{l_0 l}{OF_1 \cdot OF_2} \quad (4.44)$$

where  $l$  – distance between rear focal point of the objective lens and front focal point of the eyepiece lens. Both objectives and eyepieces are classified in terms of their magnification. Objectives typically have magnifying powers that range from 1:1 ( $1\times$ ) to 100:1 ( $100\times$ ), with the most common powers being  $10\times$ ,  $20\times$ ,  $40\times$ , and  $100\times$ . In order to achieve high magnification values, microscope objectives should have small focal lengths of  $\sim 2\text{--}5$  mm. To manufacture such objectives and decrease effect of optical aberrations compound lenses, i.e. combination of lenses, are usually used (Fig. 4.29).

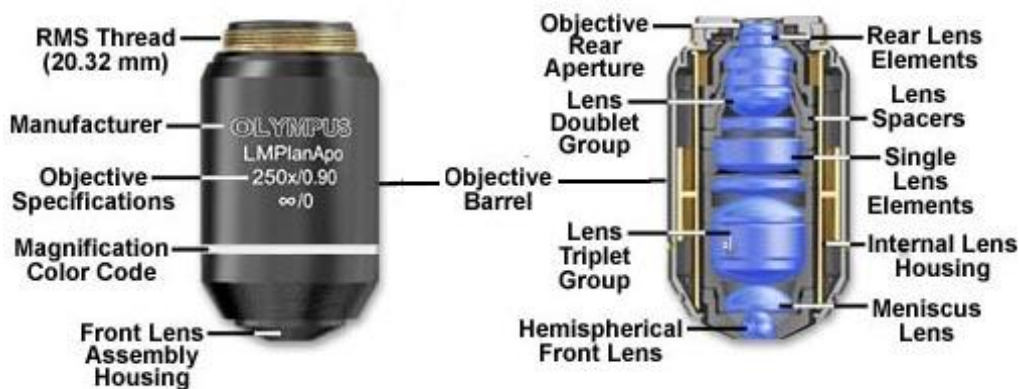


Figure 4.29. Objective and its internal structure (the image is taken from <http://micro.magnet.fsu.edu/primer/anatomy/objectives.html>)

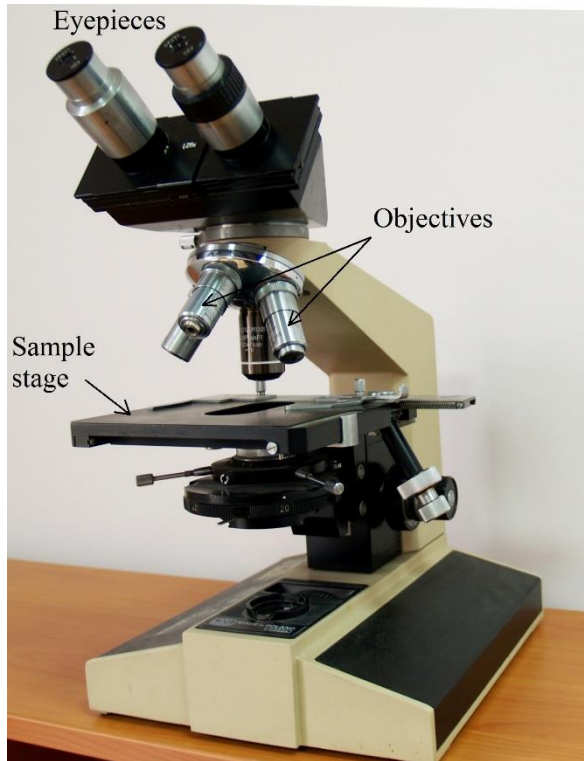
Most commonly used eyepieces have magnification values of 10×-15×. Total visual magnification of the microscope is derived by multiplying the magnification values of the objective and the eyepiece.

The range of useful magnification for an objective/eyepiece combination is defined by the numerical aperture of the microscope optical system. There is a minimum magnification necessary for the detail present in an image to be resolved, and this value is usually rather arbitrarily set as 500 times the numerical aperture (500×NA) and defined by the equation

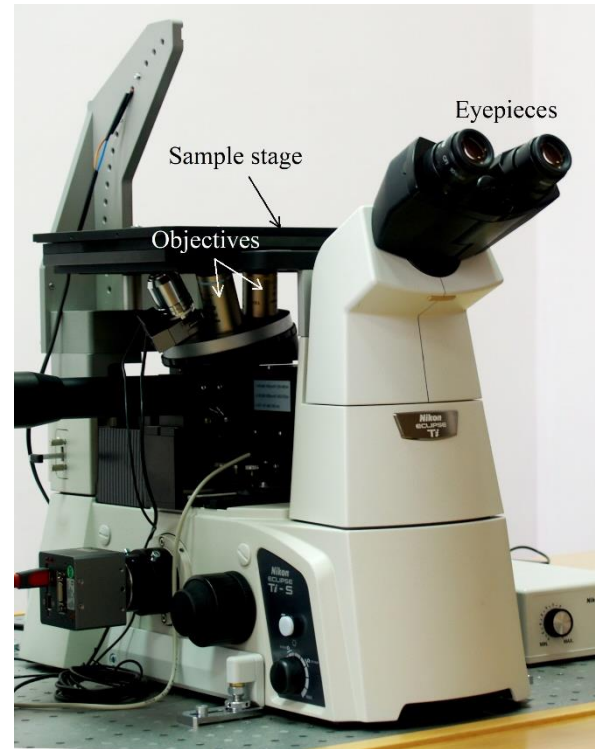
$$\text{Useful Magnification } M = 500 \div 1000 \times \text{NA}$$

where NA is numerical aperture of the objective. Magnifications higher than this value will yield no further useful information or finer resolution of image detail, and will usually lead to image quality degradation.

Depending on design and mutual arrangement of the objective and the sample stage all optical microscopes can be divided into two groups: 1) up-right microscopes and 2) inverted microscopes (Fig. 4.30). In upright microscopes objectives are located above the stage with a sample, while in inverted microscopes objectives are below the stage.



Upright microscope



Inverted microscope

Figure 4.30. Examples of upright and inverted microscopes

#### 4.2.4 Resolution of optical microscope

Resolution of any optical system including microscope can be limited either by aberration or by diffraction effects causing blurring of the image, i.e. the produced image is not anastigmatic. In the case of aberration-free optical system diffraction of light becomes the only factor that limits resolution, i.e. ability to distinguish two objects in a specimen as separate entities in an image. Diffraction comes from the wave nature of light and its effect is determined by the finite apertures of the optical elements. Therefore each point in the object plane does not correspond to a point in the image plane but to a ring-shape diffraction pattern, so-called Airy pattern.

Let us consider imaging of a point light source by optical system and diffraction pattern produced by the objective aperture. About 86% of light intensity in this pattern is concentrated within the borders of the central maximum known as Airy disc and we may consider intensity distribution in Airy disc to be a so-called Point Spread Function (PSF), corresponding to representation of a point object in the image space. For the case of Fraunhofer diffraction by a circular aperture the angular size of Airy disk  $\varphi_1$  can be determined using the equation

$$\varphi_1 = \frac{0.61\lambda}{r}, \quad (4.45)$$

where  $r$  – aperture radius.

Illustrations of diffraction patterns produced from two closely-lying point light sources are shown in Figure 4.31. Ability for separate observation of these objects via optical system can be estimated using Rayleigh criterion.

Two point light sources of equal strength are regarded as just resolved when the principal diffraction maximum of one image coincides with the first minimum of the other. If the distance is greater, the two points are well resolved and if it is smaller, they are regarded as not resolved.

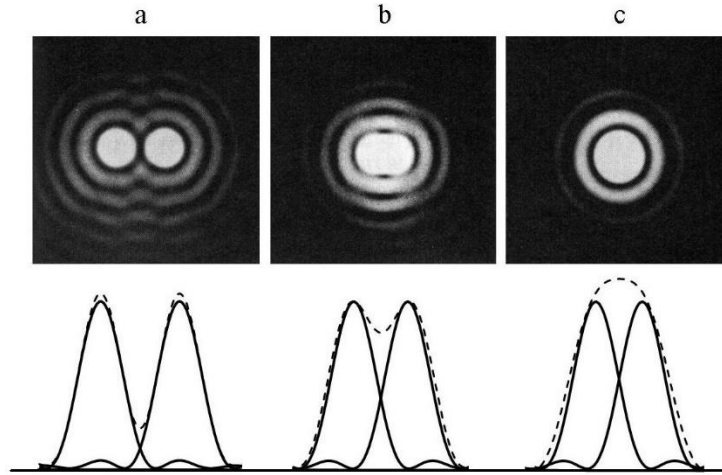


Figure 4.31. Illustrations of diffraction patterns from two closely-lying point sources. Figures were modified from the image in *M. Cagnet, M. Francon and J. C. Thierr: Atlas of Optical Phenomena (Springer-Verlag, Berlin, Göttingen, Heidelberg, 1962*

Airy diffraction patterns generated by light from two points passing through a circular aperture. Points far apart (a) or meeting the Rayleigh criterion (b) can be distinguished. Points closer than Rayleigh criterion (c) are difficult to distinguish.

Since optical microscope is used for observation and study of small objects, its resolving power is characterized by minimal distance  $y_1$  between distinguishable objects in an image. For determination of microscope spatial resolution it is important to take into account two factors. First of all, since object is located close to the microscope objective, the light going from the object and focused in the image cannot be considered as a plane wave. Secondly, it is important to take into account coherence of light used in image formation.

When two objects under observation are independent self-emitting light sources their radiation will not be coherent. If we observe objects irradiated by an external source of light, the light coming from the objects can be at least partially coherent depending on spatial and temporal coherence of the external light source.

The shortest distance  $y_1$  between two self-emitting point objects on a specimen that can still be distinguished by the observer or camera system as separate entities (so-called lateral resolution) can be estimated with the following equation

$$y_1 = \frac{0.61\lambda}{n_1 \sin u_1}, \quad (4.46)$$

where  $u_1$  – one-half angular aperture of light going through the objective (Figure 4.32). Sometimes under resolving power of microscope the magnitude reciprocal to  $y_1$  is used

$$Y_0 = \frac{1}{y_1}. \quad (4.47)$$

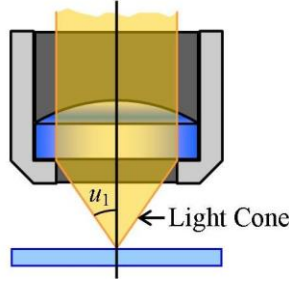


Figure 4.32. Angular aperture of the objective lens

The angle  $u_1$  is related to the numerical aperture NA via the following equation

$$NA = n_1 \sin u_1. \quad (4.48)$$

Thus, combining eq. (4.47) and (4.48) one can define lateral resolution in the specimen plane  $R_{Lateral}$  according to Rayleigh criterion:

$$R_{Lateral} = \frac{0.61\lambda}{NA}. \quad (4.49)$$

Usually external light source is used in microscopes for specimen illumination and therefore light scattered by two point objects in the sample can be partially coherent. Treatment of this case and account for coherent properties of the scattered light on spatial resolution of microscope was first done by Abbe. Usually for practical reasons a full width at half maximum (FWHM) of the central spot of Airy disk (which is easier to measure) is used to estimate lateral resolution instead of Airy disc radius:

$$R_{Lateral} = \frac{0.5\lambda}{NA}. \quad (4.50)$$

Therefore, resolving power of microscope both for illuminated and self-emitting objects grows with decrease of light wavelength used for observation and increase of numerical aperture of objective.

#### 4.2.5 Bright –field and darkfield microscopy

##### *Bright-field illumination method*

Bright-field microscopy is the simplest method of a range of techniques used for illumination and visualization of samples. **Bright-field illumination method in transmitted light** can be used to study transparent specimen that contains absorbing pigments. Bright-field observation is widely used for stained or naturally pigmented highly contrasted specimens (mineral sections, stained fixed sections of biological tissues, etc.) mounted on a glass microscope slide. In this case light passing through the specimen and its surroundings is collected to form an image against a bright background and contrast in the image is caused by absorbance (or scattering) of some part of the light transmitted by the specimen.

**Bright-field method in reflected light** is used to visualize and study non-transparent reflective objects (metal surfaces, slabs of different materials) and widely applied in metallography, inspection of semiconductor wafers and liquid crystal boards. Specimen illumination in this case is performed from above through the objective which simultaneously plays the role of illumination condenser. Due to non-equality of reflective and light scattering properties at different points of the specimen surface one can get structural information from the image produced by the microscope.

**Darkfield microscopy** is a special contrast technique where only the light diffracted or scattered from the specimen is used to form the image. **Transmitted darkfield illumination** can be used to increase the visibility of transparent (non-absorbing) specimens that lack sufficient contrast for satisfactory observation and imaging by ordinary brightfield microscopy method.

Usually this technique is used to create contrast in unstained transparent biological objects such as living cells or microorganisms. Darkfield illumination requires blocking out of the central light which ordinarily passes through and around the specimen, allowing only oblique rays to "strike" the specimen mounted on the microscope slide, and specimen illumination in this case is performed using a condenser of special construction – so-called dark-field condenser. Several constructions of darkfield condensers (for observation in reflected and transmitted light) are shown in Fig. 4.33.

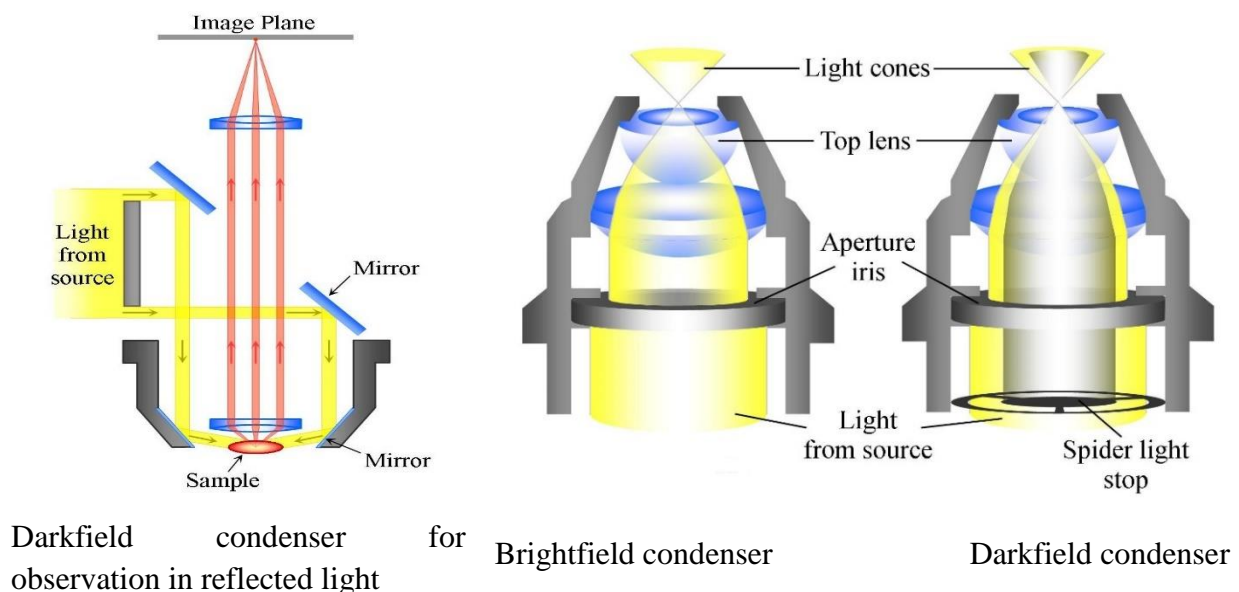




Figure 4.33. Constructions of darkfield condenser for specimens illumination and observation in reflected (left) and transmitted (right) light

Light coming from the condenser forms a hollow cone and if rays do not change their direction while passing through a transparent specimen they cannot reach the objective and participate in image formation. Image is produced only by small fraction of rays that undergo scattering, reflection or diffraction in the specimen and therefore are going through the microscope objective. The specimen can then be seen bright on a black background (Figure 4.34), however it should be noted that only elements of specimen structure where light is diffracted, reflected, and/or refracted by optical discontinuities (such as the cell membrane, nucleus, and internal organelles) are visualized.

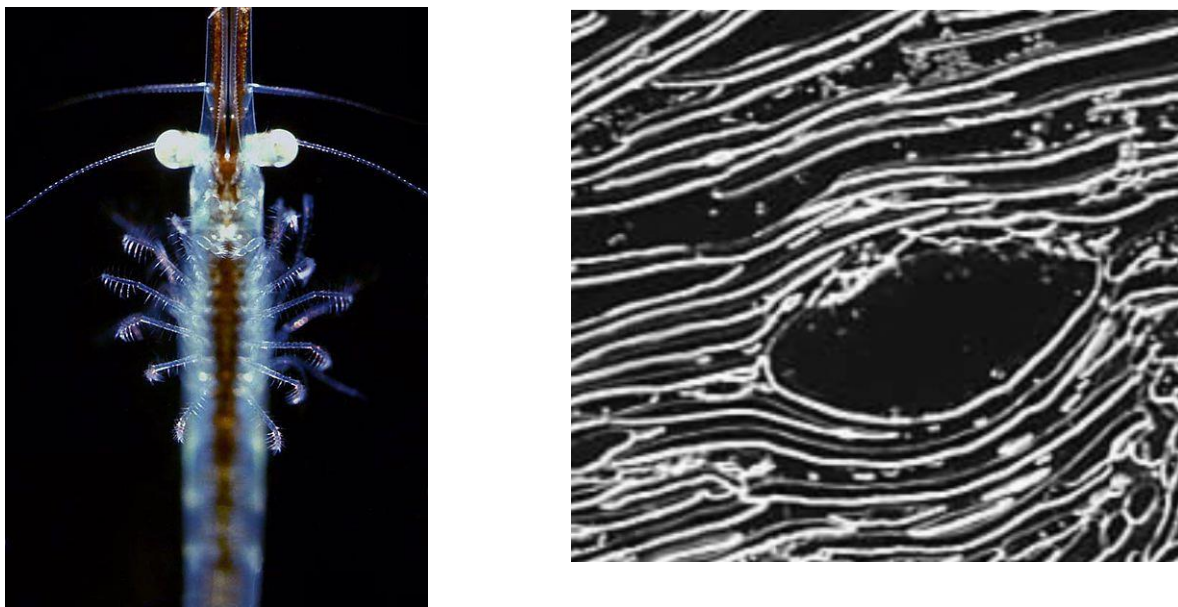


Figure 4.34. Example of microscopic images of Mysis from crustacean group (a) and liquid crystalline DNA (b) with darkfield illumination. (Images were taken from <https://upload.wikimedia.org/wikipedia/commons/thumb/c/c5/Mysis2kils.jpg/440px-Mysis2kils.jpg> and <http://olympus.magnet.fsu.edu/primer/techniques/darkfieldgallery.html>)

#### 4.2.6 Polarised light microscopy

**Polarized light microscopy** is employed to observe specimens containing (or comprised of) optically anisotropic elements using polarized light illumination. This technique is widely used in geology to study minerals in rock thin sections and can also be used to study many other specimens including both natural and industrial materials, composites, such as ceramics and polymers, and crystalline or highly ordered biological molecules such as DNA, polysaccharides, etc.



Optical properties of anisotropic objects are dependent on the object orientation with respect to the incident light. Image contrast arises from the interaction of plane-polarized light with a birefringent specimen to produce two individual wave components (ordinary and extraordinary light waves) which propagate through the sample and then recombine after passing through the analyzer. Observation can be performed both in transmitted and reflected light. Polarized light microscopy is a contrast-enhancing technique that improves the quality of the image obtained with birefringent materials when compared to other methods such as darkfield and brightfield illumination (Figure 4.35).

This technique provides all the benefits of brightfield microscopy and offers additional information, which can be used to reveal data about the structure and composition of materials and properties of anisotropic micro-objects, which are invaluable for identification and diagnostic purposes.

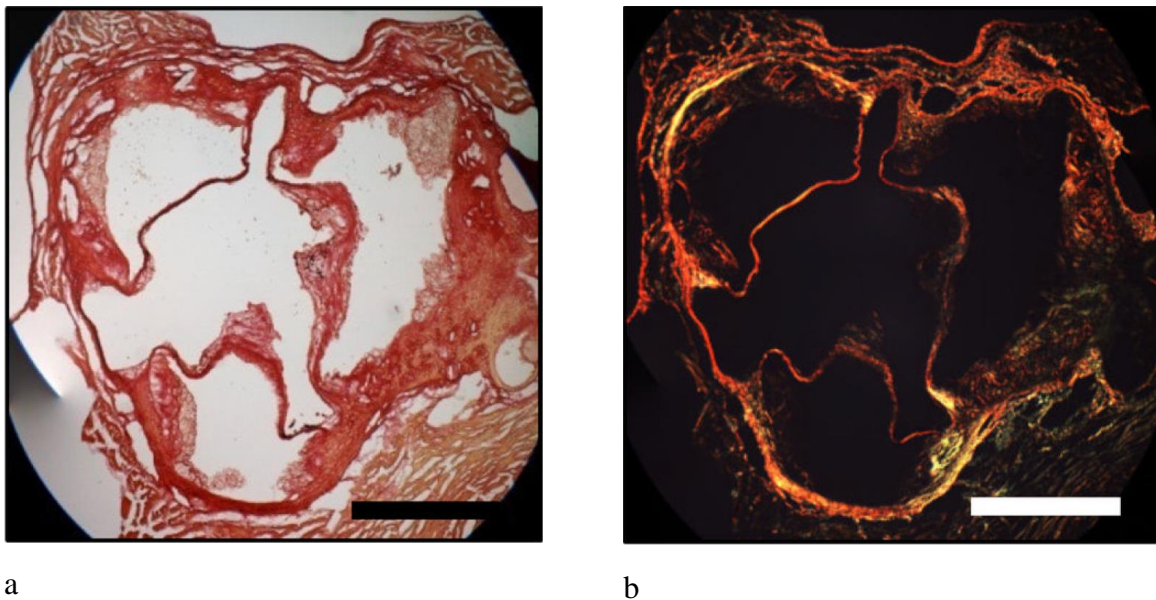


Figure 4.35. Representative images of aortic root sections stained for collagen using picrosirius red and visualized under bright field (a) or polarized light (b) illumination. Bright field images were used to define the area of plaque and collagen was quantified using the images taken under polarized light. Scale bars 500  $\mu\text{m}$ . (Hiebert, Boivin et al. 2013)

#### 4.2.7 Phase contrast microscopy

Depending on optical properties of specimens studied under the optical microscope one can consider two limiting cases. In first case, specimens exhibit significant dependence of their transparent or reflective properties across XY-plane that is imaged. This kind of specimens –so-called amplitude objects - can be easily visualized in transmitted/reflected light and image contrast is achieved due to variation of absorbance or reflective properties at different points within the specimen.

From the other hand, a large spectrum of unstained biological specimens (e.g., live organisms or cell cultures) are virtually transparent when observed in the optical microscope under brightfield illumination and variation in their absorption and reflective properties within the imaging plane is too little to provide appreciable contrast. Propagation of light through this kind of specimens, referred to as phase objects, does not lead to change in light wave intensity but results in alteration of the light phase. Phase of the light diffracted by the specimen is usually retarded by approximately  $\pi/2$  as compared to the undeviated direct light passing through or around the specimen unaffected. Therefore, due to variation of refractive properties (thickness, index of refraction, i.e. optical path length) of the specimen across the imaging plane one may expect that distribution of phases of the light wave passed through the specimen will be produced.

In 1934 Dutch physicist Fritz Zernike proposed the method describing how this phase distribution can be optically translated into corresponding changes in light intensity and therefore utilized to produce image contrast.

In phase contrast microscope configuration, the condenser aperture diaphragm is replaced by an annular ring that forms illumination in shape of a hollow cone similar to darkfield illumination technique (Fig. 4.36). These wavefronts enter the objective and an image of the annular ring is created in the rear focal plane where a phase ring is positioned. The phase ring adds a constant phase shift (usually  $\pi/2$  or  $3\pi/2$ ) to the bright, direct light originating from the annular ring in the condenser and additionally attenuates its intensity. Wavefronts of the light diffracted by the specimen will not pass through the phase ring, meaning that they will not be attenuated or retarded. Then wavefronts of the undeviated (i.e. zero order diffraction light) and the diffracted light are recombined allowing them to interfere constructively or destructively and the image is formed by the eyepiece lens. Intensity attenuation of the direct (undeviated) light is needed to make amplitudes of the diffracted and the undeviated light comparable, thus improving the contrast.

Application of phase contrast method dramatically enhances visibility of structural features of live unstained microorganisms or cell cultures and they are observed as dark objects in bright background (positive phase contrast) or light objects on dark background (negative phase contrast).

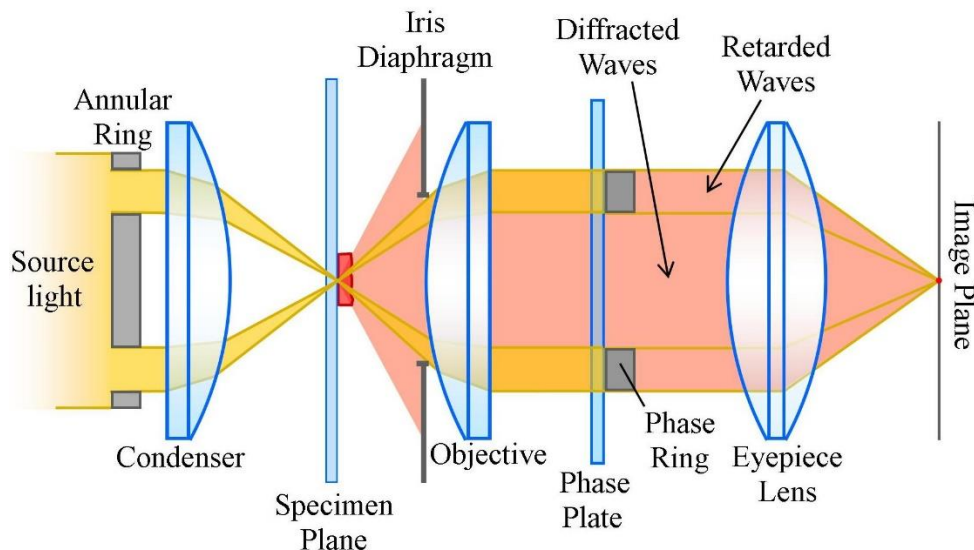


Figure 4.36. Phase contrast scheme

#### 4.2.8 Interference microscopy

Differential interference contrast (DIC) microscopy proposed by G. Nomarski has similarities with phase contrast technique and is widely employed for visualization and study of transparent specimens. Image contrast is produced by a beam-shearing interference system where the illumination light is split into two components which are spatially displaced (sheared) at the sample plane by a small distance comparable with Airy disk diameter, and recombined before observation. Polarizer and two Nomarski-modified Wollaston prisms are needed for implementation of this technique. In DIC microscopy gradients of optical path length in lateral direction of sample plane are primarily responsible for contrast.

There is also another kind of interference microscopy referred to as quantitative interference microscopy which uses two separate light beams produced by a beamsplitter and having much greater lateral separation than the ones used in differential interference microscopy (DIC). One of the beams is sent through a specimen and the other goes aside and is used as a reference beam. Both beams travel through the same optical components and then are recombined to interfere. Magnitude of optical path length difference for the beams can be measured using a compensator with high accuracy.

#### 4.2.9 Fluorescence microscopy

**Fluorescence microscopy** is a special kind of optical microscopy which uses emission of certain compounds – fluorophores in order to enhance image contrast. Term “fluorophore” denotes a substance (usually an organic dye molecule) that absorbs light of specific wavelength and effectively emits fluorescence at longer wavelengths in accordance with Stokes rule. Fluorescence microscopy uses fluorescence light emitted by a specimen for image formation and this technique is very popular to study materials and biological objects of different kind.

Normally biological objects have very weak fluorescence, therefore they are often stained using different highly-fluorescing organic dyes or inorganic nanoparticles (e.g. colloidal quantum dots CdSe/ZnS). High fluorescence quantum yield, photostability, and selectivity/specificity of binding to a biological object that is investigated are the most important criteria of fluorescent labels selection. Alexa Fluor dyes are frequently used as cell and tissue labels in fluorescence microscopy.

Fluorescence properties of some organic dyes exhibit strong dependence on polarity, viscosity, pH, concentration of specific ions in their immediate microenvironment, and these fluorophores can be used to selectively visualize compartments or structural features of biological specimens such as membranes, organelles in fixed or live cells (Slavík 2013). For example, fluorescent probes DAPI (4',6-diamidino-2-phenylindole) and Hoechst dyes are extensively used to stain DNA and RNA molecules (Fig. 4.37). Antibodies labeled with fluorescein isothiocyanate (FITC) and tetramethylrhodamine isothiocyanate (TRITC) play an important role in selective labeling of biological objects. Actin fibers in mammalian cells are usually stained with phalloidin. Popular pH sensitive dyes are based on fluorescein (carboxyfluorescein, SNARF, BCECF) and pyranine (DHPDS).

Fluorescent molecular rotors (Auramin O, thioflavin T, DCVJ, etc.) with strong dependence of their emission intensity on viscosity/rigidity of their microenvironment represent another example of environment sensitive dyes (Valeur 2001; Haidekker and Theodorakis 2007). In low viscous media their fluorescence intensity is extremely low, however binding to macromolecules or increase of solution viscosity result in significant growth of fluorescence quantum yield which makes them useful for detection of specific macromolecules or aggregates such as amyloid fibrils (Maskevich, Stsiapura et al. 2007; Stsiapura, Maskevich et al. 2008; Amdursky, Erez et al. 2012).

Fluorescent proteins constitute an important class of fluorescent probes suitable for biological in vivo studies to monitor cellular processes in living systems using optical microscopy. Discovery of green fluorescent protein (GFP) from the *Aequorea Victoria* jellyfish in the early 1960s – a natural protein that is able to fluoresce in visible range of spectrum - and rapid development of techniques to modify DNA sequences allowed scientists to genetically modify proteins and embed a fluorescent reporter in their structure. Application of fluorescent proteins has a tremendous potential for tracking of tagged biomolecules in living cells and development of biosensors.

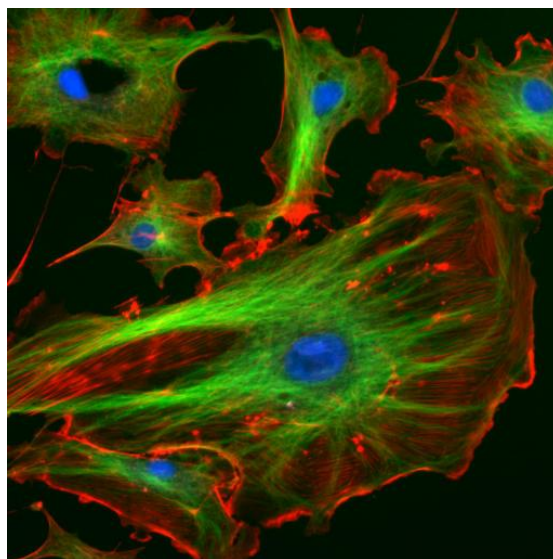


Figure 4.37. Representative image of endothelial cells under the fluorescence microscope. Nuclei are stained blue with DAPI, microtubules are marked green by an antibody bound to FITC and actin filaments are labelled red with phalloidin bound to TRITC. The image was taken from demo slide of ImageJ program (public domain) (<http://rsb.info.nih.gov/ij/images/FluorescentCells.jpg>)

Construction of a fluorescence microscope is shown in Figure 4.38. Usually fluorescence microscopes have epifluorescence design, i.e. the light source is mounted above (epi) the specimen and the excitation light passes through the microscope objective lens on its way towards the specimen. An excitation band-pass filter is located after the light source and selects a spectral range, corresponding to absorption band of the used fluorophores and suitable for specimen excitation. Excitation light is focused on the specimen by the objective lens and the same objective is used to capture fluorescence emission and direct it further to dichroic beam splitter, eyepiece lens and detector. Dichroic beam splitter (able to reflect excitation light and transmit fluorescence emission) and emission filter are used to ensure that scattered excitation light could not reach the detector.

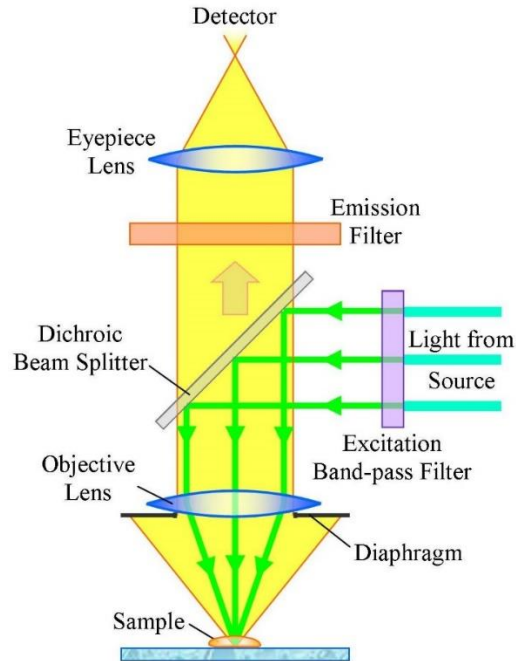


Figure 4.38. Construction of fluorescence microscope

#### 4.2.10 Total internal reflection fluorescence microscopy, (TIRFM)

Total internal reflection fluorescence microscopy provides invaluable tool for researcher to investigate processes on surfaces, interfaces, and in thin films. This technique employs effect of total internal reflection (TIR) to restrict the excitation and detection of fluorophores to a thin region of the specimen.

TIR phenomenon can be observed for a light propagating from optically more dense medium with index of refraction  $n_1$  to a medium with lower refraction index  $n_2$  ( $n_1 > n_2$ ) (Figure 4.39). For the light traveling through the interface between these two media the Snell's law gives the relation between the incident  $\phi$  and refracted  $\psi$  angles. Magnitude of the incident angle is called critical  $\phi = \phi_{cr}$  when the refraction angle reaches its maximal value  $\psi = \pi/2$ .

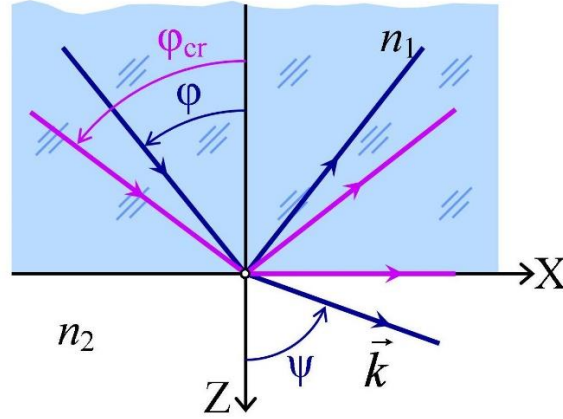


Figure 4.39. Propagation of light through the interface between two media having different indexes of refraction. Condition of total internal reflection

When condition of TIR is satisfied, i.e. at incident angles equal or larger than  $\varphi_{cr}$ , light is reflected entirely back into the first medium, however a highly restricted electromagnetic field (evanescent field) adjacent to the interface is created in the lower-index medium. Its intensity drastically decays by exponential law with increase of penetration depth  $z$  into low refraction medium and depends on incident angle, vacuum wavelength  $\lambda_0$ , relative index of refraction  $n = n_1/n_2$ :

$$I = I_0 e^{-\frac{4\pi}{\lambda_0} Z \sqrt{\sin^2 \varphi - n^2}} \quad (4.51)$$

One can assume effective penetration depth to be  $z = z_0$  where intensity of the evanescent field is lowered down by  $e$  times, i.e.  $z_0 = \frac{\lambda_0}{4\pi \sqrt{\sin^2 \varphi - n^2}}$ . Therefore, when TIR condition is satisfied

the light can penetrate approximately the distance of quarter wavelength ( $z_0 \approx \frac{\lambda_0}{4}$ ) into the second medium. When light is reflected on glass-air interface ( $n = 0.67$ ) at  $\varphi = 70^\circ$  the penetration depth of the evanescent field is  $z_0 \approx 86$  nm.

In a typical experimental setup, fluorophores located in the vicinity of the glass-liquid surface can be excited by the evanescent field, and the excitation of fluorophores is confined to a region that is about 100 nm in thickness. Because excitation of fluorophores in the bulk of the specimen is avoided, a much higher signal-to-noise ratio can be achieved compared to conventional schemes of illumination. This enhanced signal level makes it possible to detect single-molecule fluorescence by the TIRFM method. It is noteworthy, that optical section thickness provided by TIRFM method is  $\sim 10$  times smaller than the one produced by scanning confocal fluorescence microscopy technique.



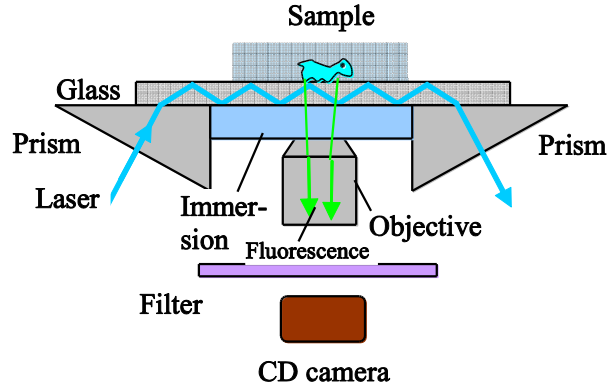


Figure 4.40. Scheme of TIRFM method

#### 4.2.11 Confocal microscope

Confocal microscopy is the optical imaging technique employing spatial filtering of light coming from a specimen (e.g. fluorescence emission or reflected light) to increase optical resolution and contrast of the image produced (Prasad, Semwogerere et al. 2007). It enables imaging at different depths along Z-coordinate within a thick specimen (a process known as optical sectioning) with following 3D- reconstruction of the specimen structure.

Laser scanning confocal microscope employs two principal ideas: point by point illumination of the sample and rejection of out-of-the-focus light.

The image of the extended specimen is generated by scanning the laser beam across the specimen in X- and Y- directions. XY-scanning can be performed either by moving the specimen relative to the laser beam or by tilting pair of galvanic mirrors which direct the laser beam position with regard to the specimen surface.

The main feature of the confocal microscope is the confocal aperture (pinhole) positioned in the plane conjugated with the imaging plane within the specimen. As a result, the detector registers only the light passing through the pinhole. The principle of a confocal microscope operation is shown in Fig. 4.41. The pinhole rejects the light originated from  $Z_2$  plane, and only the light coming from the focal plane  $Z_1$  of the specimen can be registered by the detector. Thus, all out-of-focus and off-axis light beams arriving from the sample are effectively blocked by the pinhole.

Size of the confocal pinhole is usually adjusted to match size of the Airy disc. Selection of a smaller size for the pinhole will result in decrease of the signal-to-noise ratio, while at larger pinhole sizes out-of-focus light will be registered affecting spatial resolution. Airy unit (AU) is usually used for quantification of the pinhole diameter (compare with eq.(5))

$$1AU = \frac{1.22\lambda}{NA}, \quad (4.52)$$



since AU can be considered as an “optical coordinate” unit for microscopic images in lateral direction. Pinhole sizes (PH) in  $\mu\text{m}$  and in AU are related by total magnification factor (M) of the optical system including objective lens

$$PH(\mu\text{m}) = 1\text{AU} \cdot M = \frac{1.22\lambda}{NA} M . \quad (4.53)$$

By moving the focused laser beam along a sample surface (XY-scan), an image of a thin optical layer, located in the focal plane  $Z_1$  within the specimen, can be obtained. By changing the position of the focal plane relative to the sample surface (Z - scan) one can get series of images (Z-stack) that can be used to reconstruct spatial 3D image of the sample (Figure 4.42).

The resolution along Z axis is described by axial point spread function, which depends on the wavelength of light, refractive index  $n$  between the objective lens and the specimen, and numerical aperture NA. When pinhole size is set to 1AU, axial resolution is determined by

$$R_{axial} = \frac{0.88\lambda}{n - \sqrt{n^2 - (NA)^2}} . \quad (4.54)$$

Sometimes confocal microscope can be combined with a spectrometer, which allows to analyze intensity and spectral properties of emitted fluorescence or Raman scattering with high spatial resolution (Figure 4.43).

The scanning confocal microscopy technique has gained popularity due to relative ease with which high-quality images can be obtained from specimens prepared for conventional fluorescence microscopy.

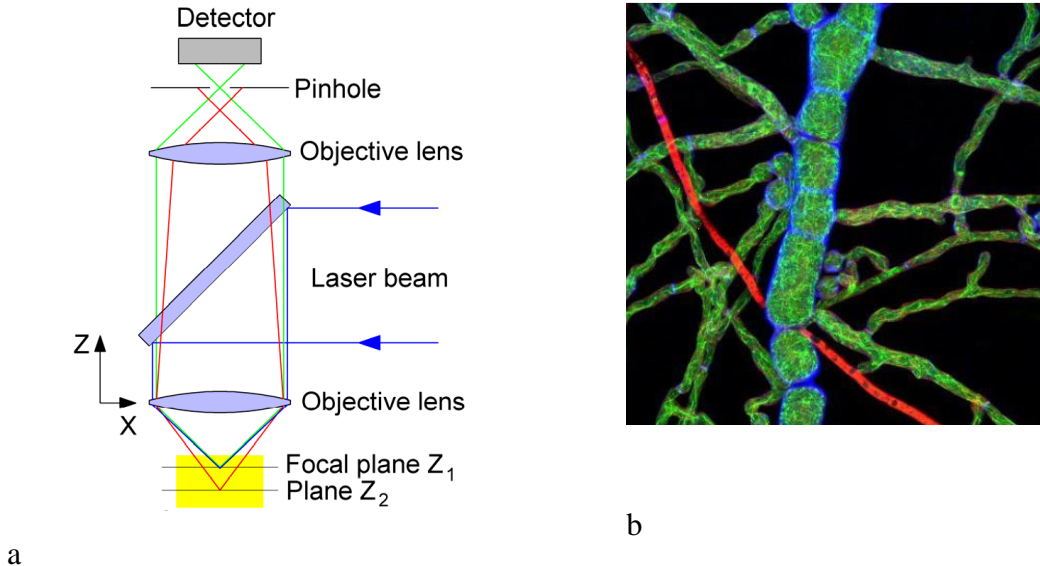


Figure 4.41. Schematics of scanning confocal microscope (a); image of fungi spore (*Neurospora crassa*) by confocal microscope (b). Cell wall was stained with Calcofluor white (blue), cytoskeleton in green (GFP) and endomembranes in red (FM-64). Image was taken from <http://www.selectscience.net/SSUserPhotos.aspx?imgID=22>.

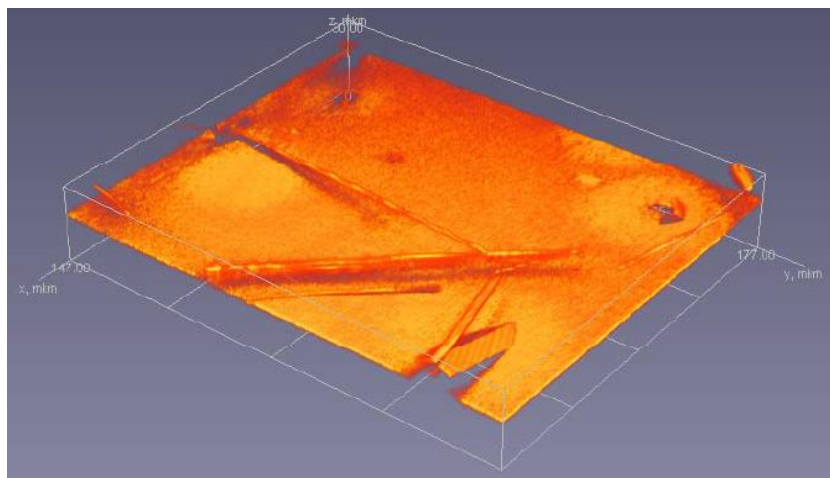


Figure 4.42. 3D image reconstruction of ZnO needles on silicon substrate using confocal microscope. Scale is  $110 \times 132 \times 30 \text{ } \mu\text{m}$ . (the image is taken from [http://www.dragon.lv/exafs/confocal\\_microscopy.htm](http://www.dragon.lv/exafs/confocal_microscopy.htm))

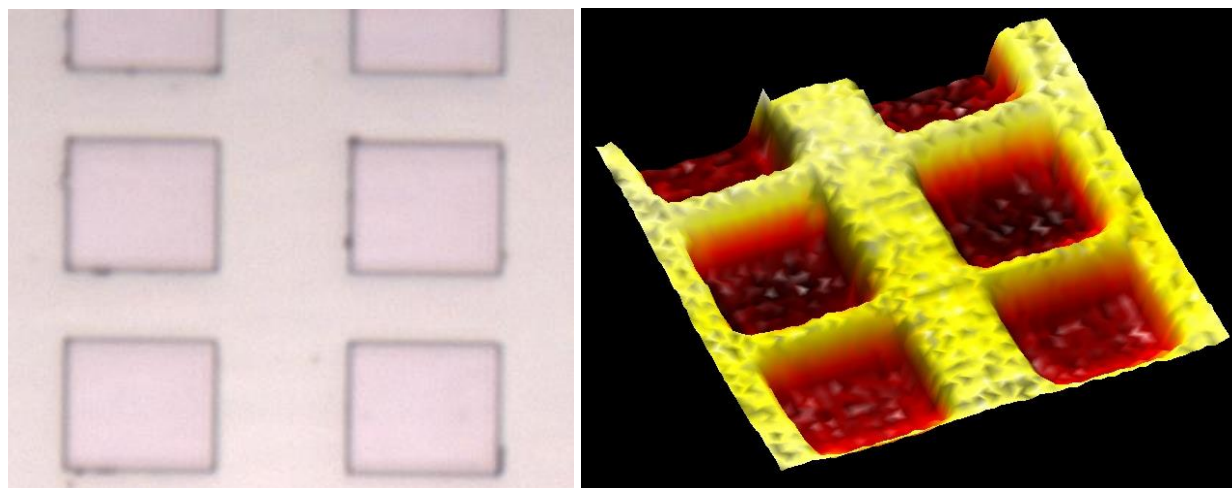


Figure 4.43. Image of poly-Si squares with  $5 \times 5 \text{ } \mu\text{m}$  sizes on silicon plate in reflected light with bright-field illumination (left) and spatial distribution of Raman peak intensity at  $520 \text{ cm}^{-1}$  (Si line) for the same sample measured by confocal microscope with spectrometer Nanofinder-S. (image is taken from <http://solinstruments.com>)

#### 4.2.12 Fluorescence lifetime imaging microscopy

Earlier we considered the microscopy methods where imaging relied on differences in absorption or reflection of light passing through a specimen or scattering from its surface, variation of refractive or fluorescent properties of specimens. An important characteristic of fluorescing objects, that also can be used for imaging, is the duration of the fluorescence decay. Typical

fluorescence lifetime changes in a range from a few picoseconds ( $10^{-12}$  s) to several tens of nanoseconds ( $10^{-9}$  s).

Fluorescence Lifetime Imaging Microscopy (FLIM) produces an image based on the differences in the excited state decay rate from a fluorescent sample. Thus, FLIM is a fluorescence imaging technique where the contrast is based on the lifetime of individual fluorophores rather than their emission spectra. The fluorescence lifetime is defined as the average lifetime of molecules in the excited state. As the fluorescence lifetime does not depend on concentration, absorption by the sample, sample thickness, photo-bleaching and/or excitation intensity it is more reliable parameter than intensity based methods. At the same time, the fluorescence lifetime depends on a wealth of environmental parameters such as pH, ion or oxygen concentration, molecular binding or the proximity of energy acceptors making it the technique of choice for functional imaging of many kinds. For example, FLIM microscopy is widely used in biological research to visualize stained cells (Figure 4.44).

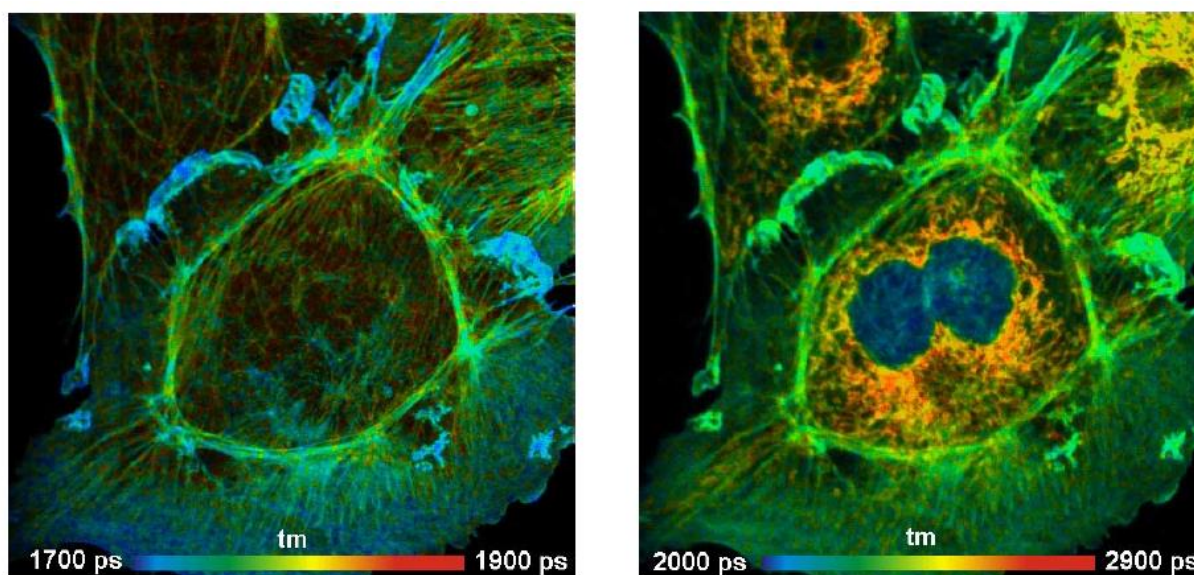


Figure 4.44. FLIM-images of BPAE cells stained with Alexa 488 phalloidin (left) and Mito Tracker Red (right). Changes of fluorescence decay lifetime are represented with color. (image taken from <http://www.becker-hickl.com>)

#### 4.2.13 Ultramicroscopy and super-resolution microscopy

Despite the advantages of traditional optical microscopy, the technique is hampered due to the resolution limit set by the diffraction of light, which restricts the amount of information that can be captured with standard objectives. The condition (5) determines the minimal distance between two point-like objects which can still be visible as separate entities using optical microscope with a certain numerical aperture under illumination with a certain light wavelength. A particle of this or greater size can be observed without distortion of its real shape. Diffraction effects of light limit ability to resolve structural features of smaller particles.

However, as it was shown more than century back, an ordinary optical microscope can be used to detect presence of particles with even smaller size (e.g., colloidal particles) and study their position and movement when special method of observation is utilized. This technique is based on effect of light scattering by small particles (Tindall scattering) and principle of the method introduced by Zsigmondy in 1902 is shown in Figure 4.45.

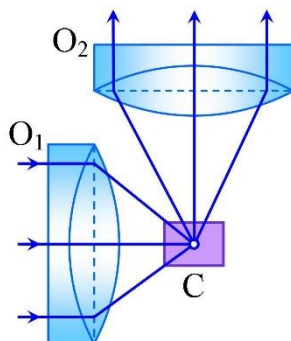


Figure 4.45. Principle of the ultramicroscope setup

Illumination light is focused using objective  $O_1$  on solution inside the cuvette  $C$  with colloidal nanoparticles. The light is scattered by colloids and captured at right angle by the second objective  $O_2$  forming a diffractional pattern. Based on this pattern one can deduce about presence, position and motion of particles, however data about size and shape of the particles obtained with this method are not very reliable.

Similar illumination mode as in the ultramicroscope is employed for **Light sheet fluorescence microscopy technique**, which features good optical sectioning capabilities and high speed. In contrast to epifluorescence microscopy only a thin slice (usually a few hundred nanometers to a few micrometers) of the sample is illuminated perpendicularly to the direction of observation.

In the past ten years, a number of novel approaches have been employed to overcome the diffraction limit, including near-field scanning optical microscopy (**NSOM**), stimulated emission depletion microscopy (**STED**), stochastic optical reconstruction microscopy (**STORM**) and structured illumination microscopy (**SIM**). These techniques have achieved improved lateral (**x-y**) resolution down to tens of nanometers, i.e. about order of magnitude lower than the diffraction limit.

#### 4.2.14 References

<https://www.microscopyu.com/>

<http://micro.magnet.fsu.edu/index.html>

<http://solinstruments.com>

<http://www.becker-hickl.com>

1.Amdursky, N., Y. Erez, et al. (2012). "Molecular rotors: what lies behind the high sensitivity of the thioflavin-T fluorescent marker." *Accounts of Chemical Research* 45(9): 1548-1557.

2. Haidekker, M. A. and E. A. Theodorakis (2007). "Molecular rotors--fluorescent biosensors for viscosity and flow." *Org Biomol Chem* 5(11): 1669-1678.
3. Hiebert, P. R., W. A. Boivin, et al. (2013). "Perforin and granzyme B have separate and distinct roles during atherosclerotic plaque development in apolipoprotein E knockout mice." *PloS one* 8(10): e78939.
4. Maskevich, A. A., V. I. Stsiapura, et al. (2007). "Spectral Properties of Thioflavin T in Solvents with Different Dielectric Properties and in a Fibril-Incorporated Form." *Journal of Proteome Research* 6(4): 1392-1401.
5. Prasad, V., D. Semwogerere, et al. (2007). "Confocal microscopy of colloids." *Journal of Physics: Condensed Matter* 19(11): 113102.
6. Slavík, J. (2013). *Fluorescence Microscopy and Fluorescent Probes*, Springer US.
7. Stsiapura, V. I., A. A. Maskevich, et al. (2008). "Thioflavin T as a Molecular Rotor: Fluorescent Properties of Thioflavin T in Solvents with Different Viscosity." *The Journal of Physical Chemistry B* 112(49): 15893-15902.
8. Valeur, B. (2001). *Molecular Fluorescence: Principles and Applications*. New York, Wiley-VCH Verlag GmbH.

## 4.3 Scanning probe microscopy

### 4.3.1 Introduction

Gerd Binnig and Heinrich Rohrer were awarded the Nobel Prize in physics in 1986, for inventing the scanning tunneling microscope (STM) (Binnig and Rohrer 1987). In contrast to conventional optical microscopes with resolution limited by diffraction effects, resolution of STM is determined by the size of the probe-sample interaction volume which can be very small when atomically sharp tips are utilized as probes. This invention led to the appearance of wide set of different scanning microscopes, defined today as scanning probe microscopy (SPM) that form images of surfaces using a physical probe (usually an ultimately sharp) that raster scans the specimen, i.e the probe is mechanically moved in close proximity ( $\sim$ nm) above the sample surface. A variety of different probes and detection modes can be used to generate a topological map of the surface, to detect and image some other physical characteristics of the specimen, or to modify them with atomic resolution. Conceptually principle of SPM raster scanning is similar to the mechanical stylus profilometer, but introduction of a feedback loop by G.Binnig and H.Rohrer allows to reduce significantly force of probe-sample interaction necessary for stable scanning which has resulted in high lateral resolution. To understand how SPM works we should emphasize its main parts: XYZ-scanner, sample-probe detection system and feedback concept.

### 4.3.2 XYZ – scanner

The requirements for XYZ-scanner are as follows: scan range in three (XYZ) dimensions of the order of few tens of micrometers with resolution down to tenths of angstroms; high speed (resonance frequency up to 1 kHz), low thermal release. Piezoelectric transducer based on inverse piezoelectric effect satisfies all these requirements and typically used in SPM. Principle of its operation is shown in Fig.4.46- 4.48. Inverse piezoelectric effect means the change of the sizes of material after application of external electric field. So let us assume that we have a cube from piezoelectric material with two electrodes (top and bottom) as shown in Fig 4.46a. If voltage  $U$  is applied to these electrodes (Fig 4.46b) the cube is elongated in vertical direction. So changing applied voltage  $U$  we can easily change the vertical size of the cube. It is not necessary to have voltage electrodes placed along direction of elongation: piezoelectric material can be polarized in such a way that application of voltage to side electrodes can produce similar displacement (Fig 4.46c).



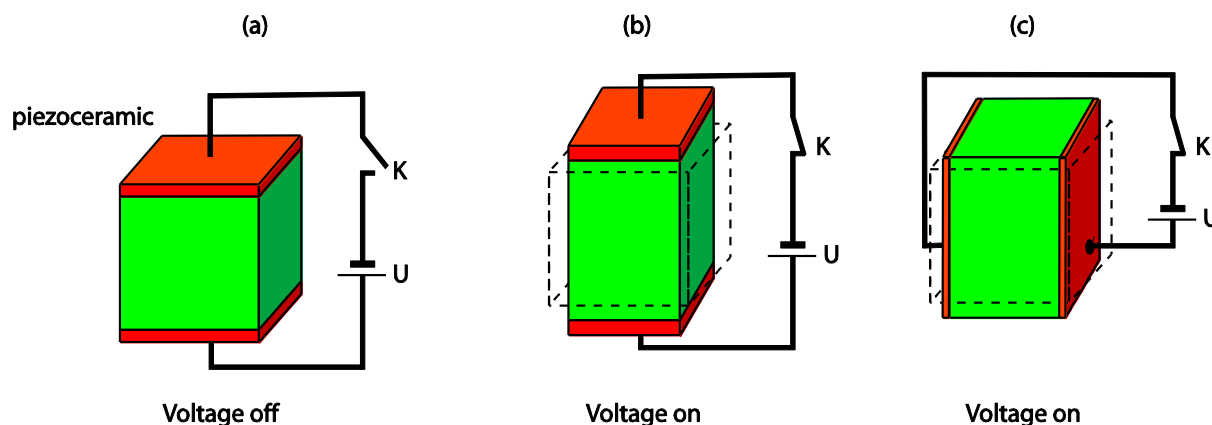


Figure 4.46. Application of inverse piezoceramic effect for displacement. (a) Applied voltage is off, initial state. (b) Voltage is applied and piezoelement is moved in vertical direction. (c)

It is not necessary to have displacement and applied electric field be collinear. When piezoceramics is appropriately polarized, then application of horizontal electric field (side pair of electrodes) can lead to vertical elongation. Dashed lines show initial dimensions before voltage was applied.

To have three dimensional (XYZ) scanner we should use three cubes with orthogonal directions of elongations glued together in piezoelectric tripod, or, which is more convenient, use scanner based on piezoelectric tube (Fig 4.47). External electrode of piezoelectric tube is divided into four electrically isolated pads and voltage proportional to displacement in X, Y directions are applied to the corresponding pairs of electrodes (X, -X) and (Y, -Y). Voltage applied to one pair of electrodes (e.g. (X,-X)) leads to tube bending in that direction, i.e. probe mounted at the bottom of the tube will be moved in X and slightly in Z directions. Motion in Z direction is realized by application of voltage to the internal electrode. Procedure of raster scanning is as follows (Fig 4.48). First saw-like voltage is applied to X electrodes pair and full range displacement in this direction is achieved; then one step voltage is applied to Y electrodes pair and scanner returns back in X-direction. Procedure is repeated until scanner reaches full range X,Y position. In raster scanning the motion along X axis is rather quick and this direction is named as direction of fast scanning, and respectively Y direction is named as direction of slow scanning. Although X and Y directions in the isotropic sample are equivalent, due to raster scanning procedure the directions of slow and fast scanning differ and it should be taken into account while analyzing acquired data.

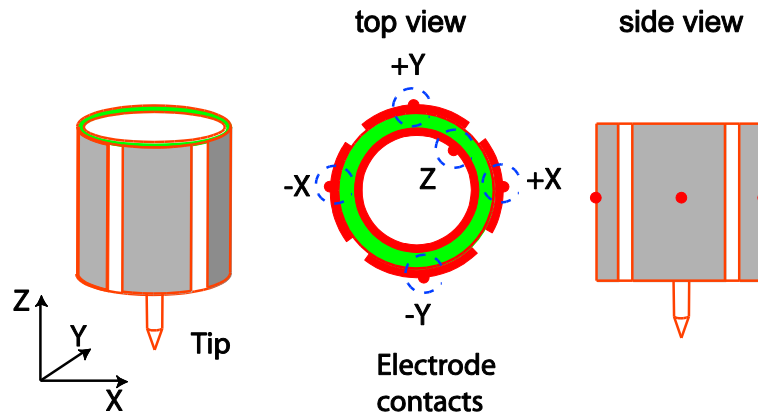


Figure 4.47. Piezoelectric scanner on the base of piezoelectric tube

Outer electrode is divided on four electrically isolated parts, to which voltages proportional to displacements in  $+X$ ,  $-X$ ,  $+Y$ , and  $-Y$  directions. Voltage proportional to displacement in  $Z$  direction is applied to inner electrode. Sharp scanning tip (or cantilever) is connected to the bottom part of the tube. Upper part of the tube is fixed. During the scanning application of the voltage to the opposite side electrodes lead to tube bending in appropriate direction, which is equivalent to the horizontal displacement in  $XY$  plane.  $Z$ -voltage leads to elongation tube in  $Z$ -direction.

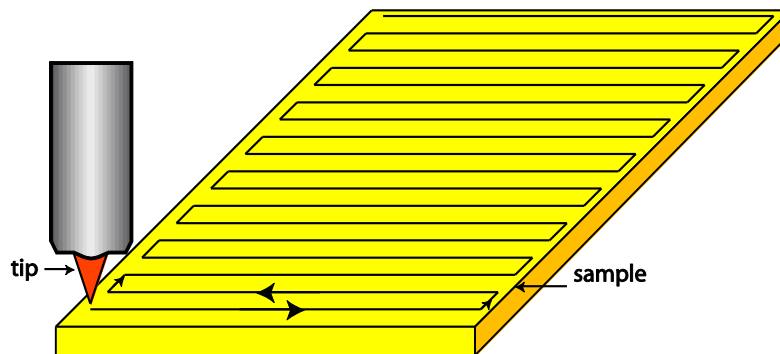


Figure 4.48. Principle of scanning probe imaging – raster scanning

Piezoelectric scanner moves tip along shown line while feedback system control tip-sample distance. Two horizontal directions are defined as fast and slow scanning axis

#### 4.3.3 Sample – probe detection system

The construction of sample-probe detection system will be considered on the example of atomic force microscope (AFM). In AFM the force acting between probe and sample is used for detection probe-sample interaction. The dependence of force between probe and sample on their separation can be approximated as a gradient of Lennard-Jones potential and schematically is shown in Fig 4.49. Initially, when there is no direct contact between the tip and the sample, the decrease of the



separation gives rise to the appearance of an attractive force. When the tip is approached further to the surface repulsive forces become dominant and determine resistance of material to the mutual penetration.

Two sharp parts of this curve denoted as "attractive" and "repulsive" can be used for selection of a set point in AFM operation. Operation with a set point within a range where attractive forces dominate (due to van der Waals force) means that the probe is far from the surface and this operation mode is named as "non-contact". Non-contact mode is very sensitive to the sample and probe surfaces state which results in practical inability of its application in ambient conditions besides ultra high vacuum environment. This non-contact mode should not be confused with "Tapping mode®" explained below. Contact mode is due to sample rigidity and is widely used.

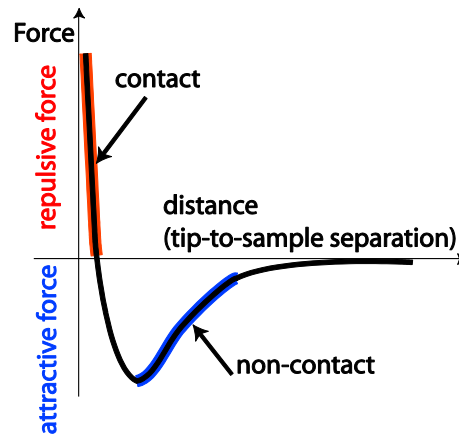


Figure 4.49. Sample – tip interaction. Two parts of dependency (repulsive and attractive) can be used for stabilization of tip-sample position

To have possibility to measure a force between the probe (very sharp tip) and sample surface we should mount this probe on a small plate named cantilever. When the probe tip is brought into proximity of a specimen surface, arising forces due to the interaction with a specimen should lead to cantilever bending and using measured displacement and cantilever spring constant this force can be calculated according to Hooke's law. The cantilever side opposite to the tip is reflective and the bending can be quantified by deflection of the laser beam (Fig.4.50) measured with four quad photo-detector. The construction of the photo-detector allows to measure both vertical and horizontal deflections of the cantilever where the latter deflection appears due to cantilever torsion with friction force.

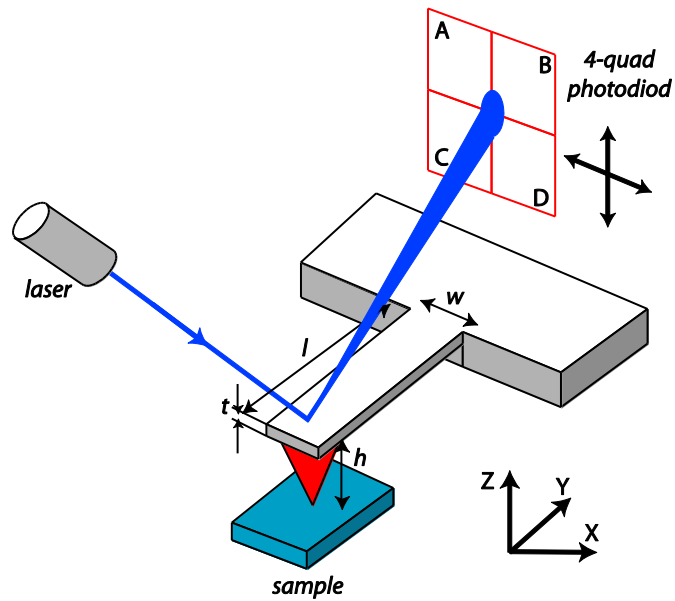


Figure 4.50. Beam deflector principle

The elastic parameters of cantilever are characterized via its geometrical parameters: beam width  $w$ , length  $l$  and thickness  $t$ . Height of tip  $h$  typically is about 10  $\mu\text{m}$ . Deflection of cantilever is measured via four-quadrant photodetector which output signal depends on displacement of laser spot after reflection from bend cantilever.

So the entire construction of AFM should look as follows (Fig. 4.51). A sample is mounted on XYZ-scanner, a force between probe and sample is measured with displacement detector and passed to the control electronics which should stabilize its value during XY-scanning by sample displacement in Z direction.

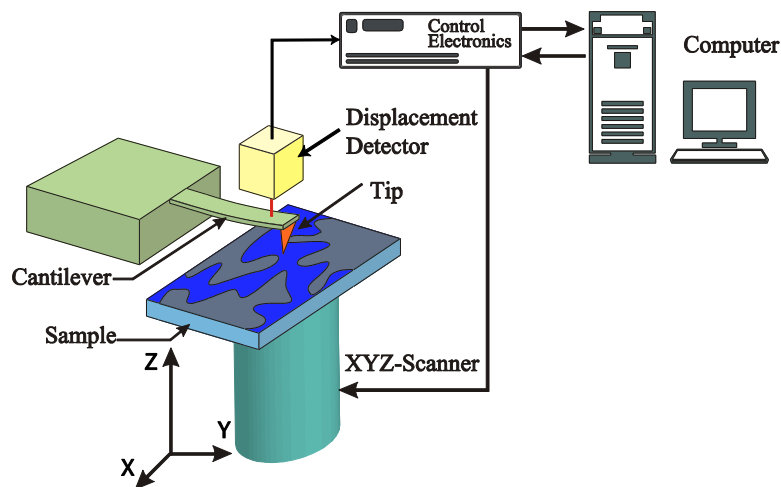


Figure 4.51. Principle of AFM imaging. Sample is mounted on the top of piezoelectric tube XYZ-scanner

Displacement of cantilever under tip-sample interaction is measured with displacement detector. Control electronics stabilize tip-sample interaction and realize communication with computer.

#### 4.3.4 Feedback concept

Now let us consider the function of control electronics in detail (Fig 4.52). Signal from the displacement detector is compared with a reference force value ("set point"). If their difference is zero then everything is OK. But if the difference signal (error "U(t)") is not equal to zero then we apply voltage to Z direction to reduce the difference signal to zero. The required function can be realized using proportional-integral-derivative (PID) regulator according to its transfer function

$$Z(t) = K_P \cdot U(t) + K_I \int U(t)dt + K_D \frac{dU(t)}{dt}, \quad (4.55)$$

where  $K_P$ ,  $K_I$ ,  $K_D$  – are coefficients of proportional, integral and derivative control.  $U(t)$  – input error signal.

The selection of  $K_P$ ,  $K_I$ ,  $K_D$  coefficients should provide the stability of the feedback loop (absence of self generation) together with maximal performance (speed of scanning).

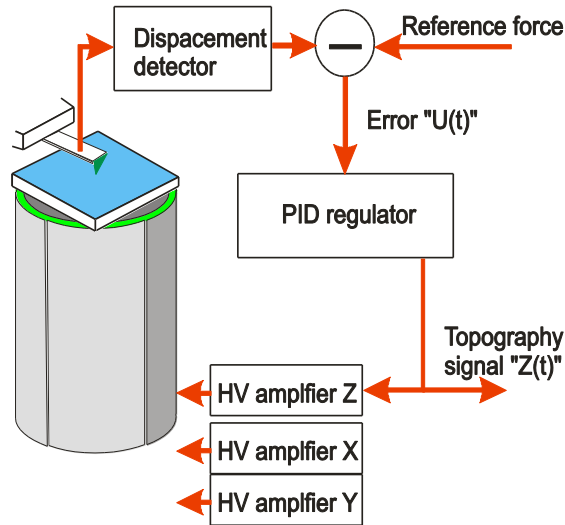


Figure 4.52. Feedback loop for SPM measurements

Signal from displacement sensor is compared with reference force. Their difference ("error") is applied to the input of PID-regulator. Output of PID-regulator is applied to Z-electrode of piezoscanner to stabilize tip-sample interaction and stored in the computer as topography signal "Z(t)". The parameters of PID-regulator should be selected to provide stabilization with maximal feedback performance.

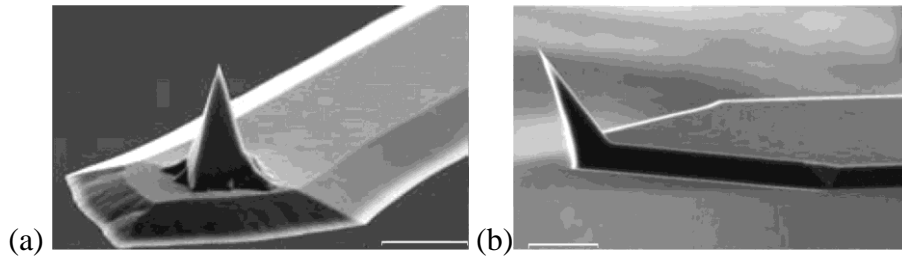


Figure 4.53. a) Standard AFM tip (Nanosensors, PointProbe® Plus) and b) a special AFM tip optimized for top visibility (Nanosensors, Advanced Tip at the End of the Cantilever). The pointed cantilever as well as the angle of the tip apex towards the front increase the exposure of the apex towards the focused laser light and reduce shadowing effects. Scale bar 10  $\mu\text{m}$ . Images from Nanosensors (<http://www.nanosensors.com>).

#### 4.3.5 AFM probes and AFM imaging

The sharpness of the probe is a key factor affecting lateral resolution. In earlier days AFM probes were prepared manually from a specially bent etched tungsten wire. This process required special training and reproducibility of probes was quite low. Nowadays all commercial AFM probes are prepared by silicon etching techniques from semiconductor industry and their images are shown in Fig 4.53. Curvature radius of the tip can be as low as few nm, but typically for ambient conditions of measurements it can be evaluated as 20-50 nm. For special requirements (combined optical measurements) tip can be curved in a special way (Fig 4.53b) to prevent cantilever shadow effect during light illumination.

To enhance sensitivity of AFM (and reduce noise level) the cantilever can be introduced to the forced oscillations with amplitude up to hundred nm. Let us consider this process in detail. Cantilever can be represented as a mass  $m$  mounted to the spring with stiffness  $k$ . Probe –sample distance is  $z$ , the amplitude of cantilever oscillations is  $Z_d(t)$  (Fig 4.54).

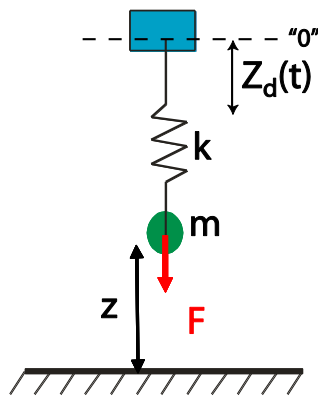


Figure 4.54. Mechanical model of the dynamic tip-sample interaction

The cantilever is represented as mass  $m$  on the spring with stiffness  $k$ . Tip-sample separation distance is  $z$ . An external oscillating force is applied to the cantilever resulting in its oscillation with amplitude  $Z_d(t)$ .

When probe-sample force is  $F$  then spring is elongated and  $z = z_d + F/k$ . In dynamic mode cantilever is introduced to the forced oscillation by additional piezodriver and

$z_d = A_d \cdot \cos \omega t \rightarrow z = A \cos(\omega t + \varphi)$ . So equation of probe motion can be written as  $m \frac{d^2 z}{dt^2} = F - k(z - z_d) - m\gamma \frac{dz}{dt}$ , where  $\omega_0 = \sqrt{k/m}$  and  $\gamma = \omega_0/Q$ ,  $Q$  - quality factor.

In the case  $F=0$  (probe is far from the surface)  $A = \frac{A_d}{\sqrt{(1 - \omega^2/\omega_0^2)^2 + (\omega/Q\omega_0)^2}}$  and  $\tan \varphi = \frac{\omega\omega_0}{Q(\omega_0^2 - \omega^2)}$ .

Their graphic dependencies are shown in Fig 4.55. If probe is approaching the sample interaction force  $F$  can be expanded as a series  $F = F_0 + z \frac{dF}{dz} + \dots$ , which results in the shift of resonance frequency  $\omega_r = \sqrt{\frac{k - dF/dz}{m}}$  (Fig 3.12.9) and according to the equations  $\frac{\Delta f}{f} = -\frac{1}{2k} \frac{dF}{dz}$  and  $\Delta \varphi = \frac{Q}{k} \frac{dF}{dz}$ .

For typical cantilever with  $f_r = 80\text{kHz}$ ,  $k=1\text{ N/m}$ ,  $Q = 100$ , interaction force of  $F = 0.2\text{pN}$  leads to  $\Delta f/f = 1.5 \cdot 10^{-5}$  and  $\Delta \varphi = 0.15^\circ$ .

Dynamic mode of operation was patented by Digital Instruments as "Tapping Mode" (Tapping Mode® is a trademark of Digital Instruments, USA), so many other AFM manufactures cannot use this term and have to introduce other terms as "semicontact", "non-contact", etc.

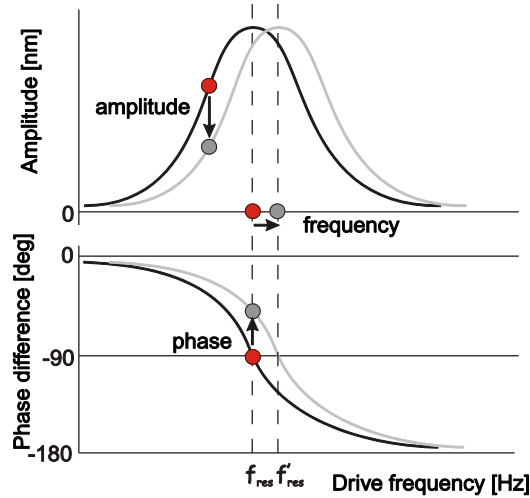


Figure 4.55. Amplitude-frequency and phase-frequency dependences in mechanical model of the dynamic tip-sample interaction. Tip-sample interaction lead to change of resonance frequency, which can be detected as change of amplitude of the forced oscillations or their phase

The sensitivity of static and dynamical modes from thermodynamic parameters was evaluated in (Albrecht, Grütter et al. 1991) as  $dF_{min} = \sqrt{\frac{4kk_BTB}{\omega_0 Q}}$  and  $[\frac{dF}{dz}]_{min} = \frac{1}{A} \sqrt{\frac{4kk_bTB}{\omega_o Q}}$ ,

where B-bandpass  $B = 4\pi \cdot f_{scan} \cdot N_{pixel}$ ,  $f_{scan}$  – frequency of fast scanning,  $N_{pixel}$  – number of pixels per line, T – temperature,  $k_B$  - Boltzman constant

For typical cantilever with  $f_0=100$  kHz,  $Q=100$ ,  $k=1$  N/m,  $f_{scan}=0.1$  Hz,  $N_{pixel}=100$ ,  $A=10$  nm

$$dF_{min} = 1.5 \cdot 10^{-13} N, \left[ \frac{dF}{dz} \right]_{min} = 1.5 \cdot 10^{-5} N/m$$

An example of AFM imaging is given in Fig 4.56 as topographic image (a) and cross-section along the line AB (b). The measurements of high temperature superconductor  $La_{2\delta}Sr_{\delta}CuO_4$  film with 50 nm thickness were done at ambient conditions in static contact mode.

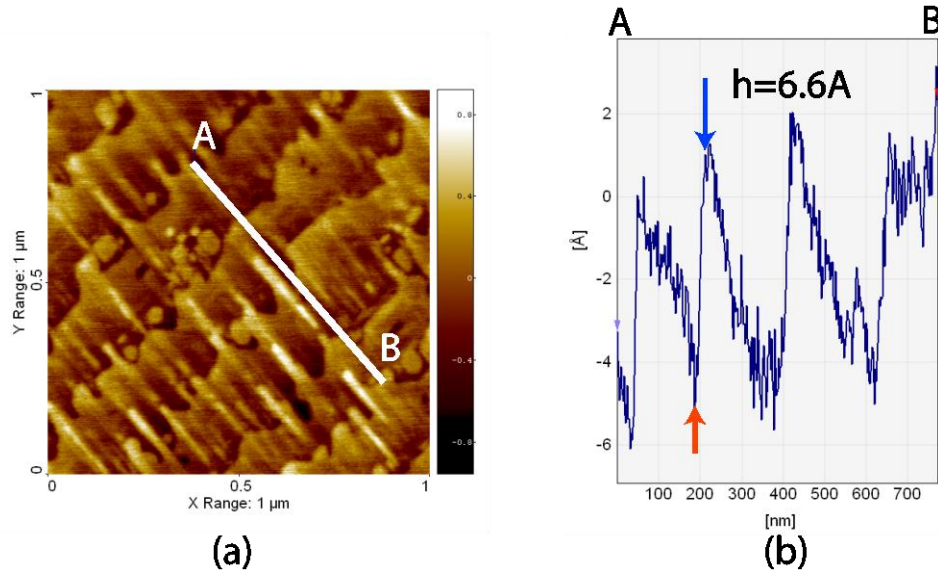


Figure 4.56. (a) AFM image of high temperature superconductor  $La_{2\delta}Sr_{\delta}CuO_4$  film. Half-cell steps (height 6.6Å) area shown on topography cross section along line AB (b)

#### 4.3.6 Magnetic force measurements

Method of interleave scanning was developed to measure long range forces (magnetic, electric, etc). It was named as "Lift Mode®" method and registered by Digital Instruments. Principle of the lift mode is explained in the following example of magnetic force measurements (MFM) (Fig 4.57) – between magnetic sample and magnetic probe.

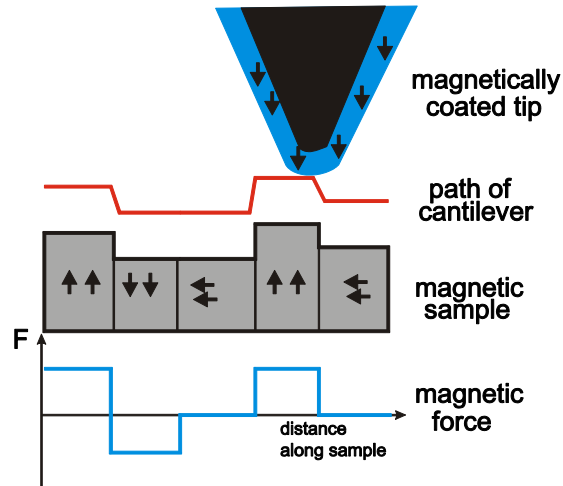


Figure 4.57. Principle of MFM imaging – "Lift Mode®". Two paths scanning. During the first scan the surface profile is recorded. At the second scan magnetic tip is elevated over the sample at constant height and tip-sample interaction is measured. This two paths method is generally used in all cases when long-range forces should be separated from short-range ones.

In this case each scanning line is scanned twice. During the first scan the probe is brought into the contact with a sample and profile of the surface along the scan line is recorded. During the second scan along the same line the separation between the probe and the surface is maintained at some constant value using data from the previously recorded surface profile. During this second scan the feedback is off and the probe-sample interaction force is recorded.

To evaluate magnitude of force appeared between probe and sample during magnetic force measurements let us use a scheme shown in Fig 4.58. Probe and surface are represented as two magnetic volumes with magnetization  $\vec{M}$  along connection line,  $z$  - probe-sample distance.

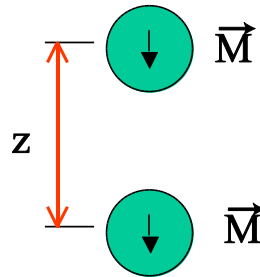


Figure 4.58. Model of magnetic interaction during magnetic force measurements. Two spheres with magnetic moment  $\vec{M}$  are at the distance  $Z$ .

Magnetic materials typically used for MFM tip preparation are some alloys on the base of cobalt or iron, so their magnetic moment  $\vec{M} = 10^6 \text{ A/m}$ ; volume  $V = 10^3 \text{ nm}^3$ ;  $z = 30 \text{ nm}$ .

Interaction force  $F$  is calculated from dipole approximation  $\vec{F} = \mu_0 (\vec{M} \cdot \vec{\nabla}) \vec{H}$ , which gives magnetic field distribution as  $\vec{H}(\vec{r}) = \frac{\mu_0}{4\pi r^5} (3(\vec{M} \cdot \vec{r})\vec{r} - \vec{M} \cdot r^2)$ . If  $\vec{M} = (0, 0, M_z)$ , then

$F_z = \frac{3\mu_0 (M_s V)^2}{4\pi z^4}$ . That gives  $Fz = 0.2pN$  and  $\frac{dF_z}{dz} = 4F_z/z = 0.3 \cdot 10^{-4} N/m$ . It means that observation of magnetic contrast is quite realistic both in static and in dynamic mode.

To understand magnetic contrast during MFM measurements let us consider two extreme domain configurations in the sample. The direction of magnetization of the probe is perpendicular to the sample surface. If magnetic domains are in-plane (Fig 4.59), then maximal gradient of the sample magnetic field is along the inter domains boundary and magnetic contrast is observed only at the domain boundaries.

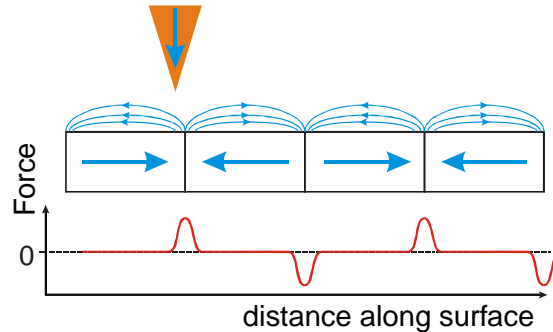


Figure 4.59. MFM imaging of in-plane magnetic domains. Magnetic contrast is observed only at the domain boundaries; its sign depends on direction of neighbor domain magnetization.

For out-of-plane magnetic configuration of the sample (Fig 4.60.) at the domain boundaries we should observe the change of the sign of magnetic contrast. But entire magnetic domain gives approximately the same value of magnetic contrast; opposite domains give different sign.

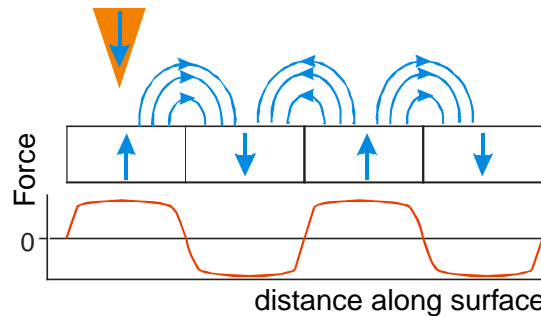


Figure 4.60. MFM imaging of out-of-plane magnetic domains. Magnetic contrast is observed for total sample surface.

The examples of MFM imaging are given in Fig 4.61 and Fig 4.62.



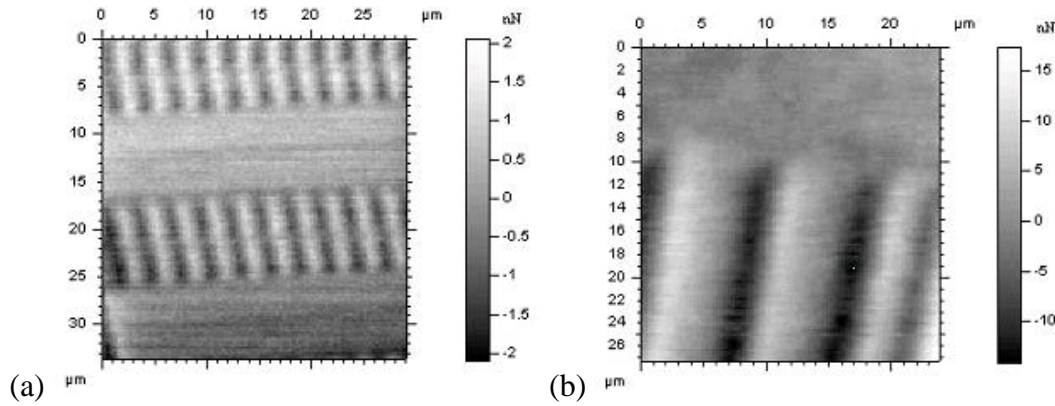


Figure 4.61. MFM images a) HDD WDC 1270 with capacity of 270.4MB; b) Diskette 1.44 Mb 3.5". Contact mode.

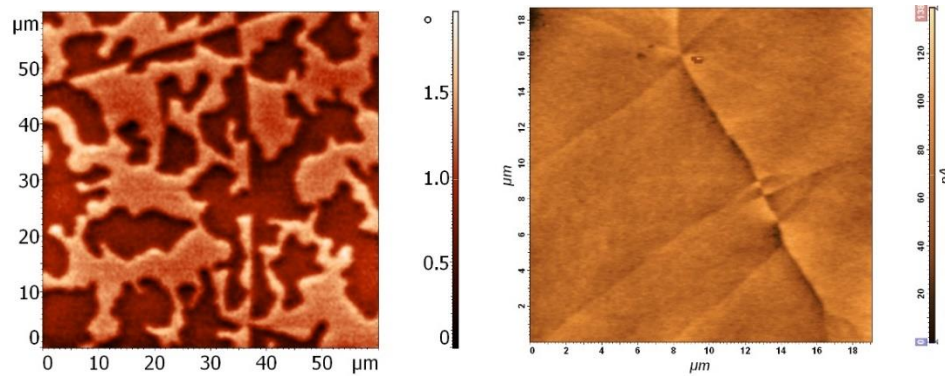


Figure 4.62. MFM images of (Au 5nm)/(Co  $d_{Co}$ )/(Au 5nm) trilayer;  $d_{Co}=1.6\text{nm}$  (a),  $d_{Co}=25\text{nm}$  (b). In AFM images sample surface is flat. Tapping mode.

#### 4.3.7 Electric force microscopy

Electric force microscopy (EFM, or Kelvin mode) utilizes similar to MFM setup for measurements of electric domains in specimens. It is also based on two-pass interleave technique – Lift Mode (Fig 4.63). During the first scan the sample surface profile is recorded, and during the second scan the distribution of electric field (or potential) along to the surface is obtained. In contrast to MFM measurements, for EFM measurements it is very easy to modulate tip-sample interaction during the second scan by applying small voltage  $V_{ac}$  between the conductive tip and sample. In such situation the applied voltage to the tip  $V_{tip}$  is  $V_{tip} = V_{dc} + V_{ac}\sin\omega t$ ; where  $V_{dc}$  - static voltage. Cantilever response (AFM feedback in this case is off) should modulate tip-sample distance in harmonic way as  $d = d_0 + A(w)\sin(\omega t + \varphi)$ . If we consider tip-sample as capacitor, then force of interaction is  $F = \frac{dE}{dz} = \frac{1}{2} \frac{dC}{dz} (V_{tip} - V_{surface})^2$ . And first harmonic of the interaction force  $F_{1\omega} = \frac{dC}{dz} V_{ac} (V_{tip} - V_{surface})$ . So if we introduces feedback loop which should stabilize  $F_{1\omega}$  as 0 with application of constant potential  $V_{dc}$  between tip and sample, then  $V_{dc} = V_{surface}$  under condition  $F_{1\omega} = 0$ . The realization scheme of EFM is shown in Fig 4.64. In addition to AFM stabilization loop

"AFM-part" explained above, the "EFM-part" consisting of comparator, lock-in detector and PID-regulator is added. And it operates during the second pass scan.

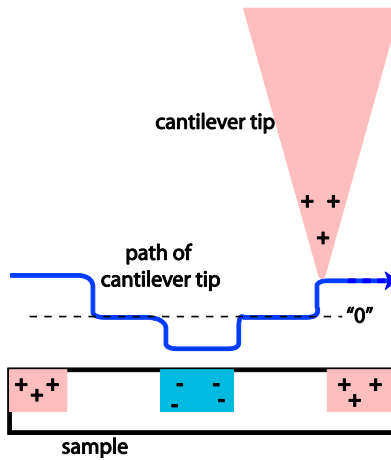


Figure 4.63. Principle of electrostatic force microscope (EFM). EFM measures electric field gradient distribution above the sample surface, through measuring local electrostatic interaction between a conductive tip and a sample using Lift Mode<sup>®</sup> method.

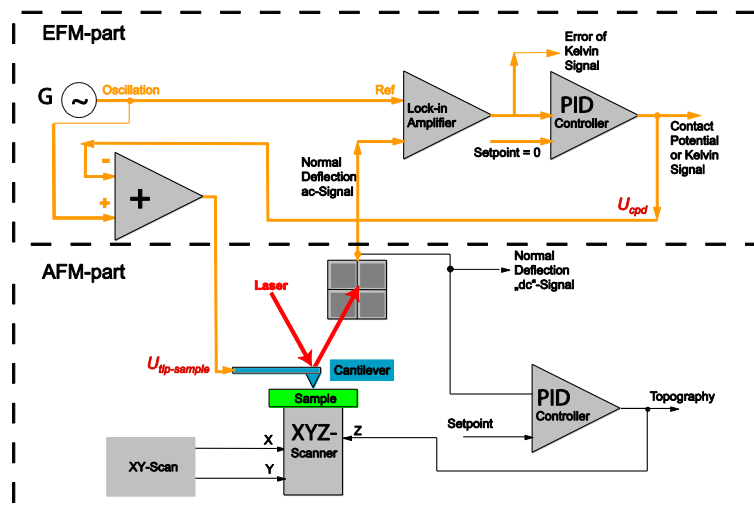


Figure 4.64. Kelvin probe microscopy measuring scheme. Typical part for AFM measurements "AFM-part" is extended with additional "EFM-part". External generator G applied additional voltage between tip-sample resulting in appearance of oscillating component of tip-sample force. After detection with lock-in amplifier PID-regulator provides dc-voltage equals to the local surface potential.

The example of EFM imaging is given in Fig 4.65. During first part of the experiment ferroelectric film was polarized by application of different voltages to the tip while it is scanned along horizontal lines with 1 mkm length. After polarization the same tip was used to detect induced electric charge (or potential) distribution.

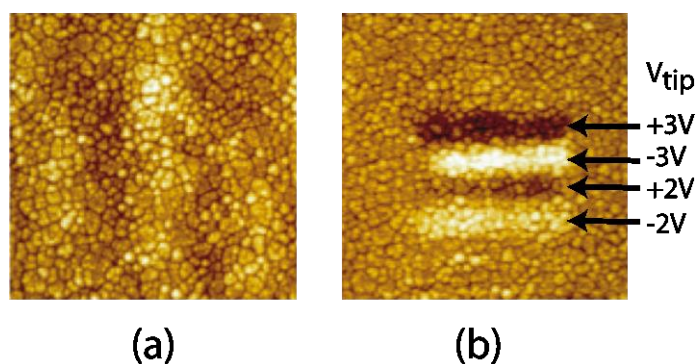


Figure 4.65. (a) Topography and FM-EFM (b) images of a ferroelectric film with electrical bits written onto it. The four bits were written with a metal-coated AFM tip at +3,-3,+2, and -2volts from top to bottom. The images were captured with the same tip, moments after polarizing the central region. Carrying +1V, the tip failed to polarize this ferroelectric film. 2 $\mu$ m scans (adapted from Application note "Electric Force Microscopy, Surface Potential Imaging, and Surface Electric Modification with the Atomic Force Microscope (AFM)", Veeco Instruments Inc, 2004, <http://www.veeco.com>).

#### 4.3.8 Scanning near-field optical microscopy

Scanning near-field optical microscopy (SNOM) represents a variety of scanning probe microscopy techniques, which combines high spatial resolution with sensitivity and spectroscopic capabilities inherent to optical methods of detection. In contrast to ordinary optical microscopy, which resolution is limited by diffraction of light, it relies on imaging in so-called near-field regime when distances between probe and object are much smaller than wavelength  $\lambda$  of the light.

Two main approaches (Fig. 4.66) to optical near-field microscopy exist: (i) aperture type SNOM and (ii) apertureless techniques. In the first case, a subwavelength size aperture on a scanning tip is used as an optical probe. This is usually an opening in a metal coating of either an optical fiber tip or of a cantilever. Excitation light is focused through the aperture resulting in evanescent electromagnetic field formation on the other side of the aperture at distances  $Z < 10$  nm close to its surface (near-field range). When a sample is scanned at a small distance below the aperture, the optical resolution of transmitted or reflected light is limited, in general, only by the diameter of the aperture and lateral resolution of 20 nm has been demonstrated (Dürig, Pohl et al. 1986).

Typical setup of aperture type SNOM is shown in Fig. 4.67. Specially etched optical fiber, which tip is playing the role of a subwavelength aperture, is used for sample excitation. Usually a tuning fork attached to the probe fiber is used to control spatial separation between the fiber tip and sample surface. Lateral oscillations of the probe tip at a mechanical resonance frequency are induced, and as the oscillating tip approaches the sample, forces between the tip and sample damp the amplitude of the tip oscillation. Feedback signal from the tuning fork is used to maintain constant distance between the probe and the sample surface by PZT. NSOM images are generated by scanning the

fiber tip over the sample in a two-dimensional raster pattern and collecting the emitted/scattered radiation in the optical far-field by a detector.

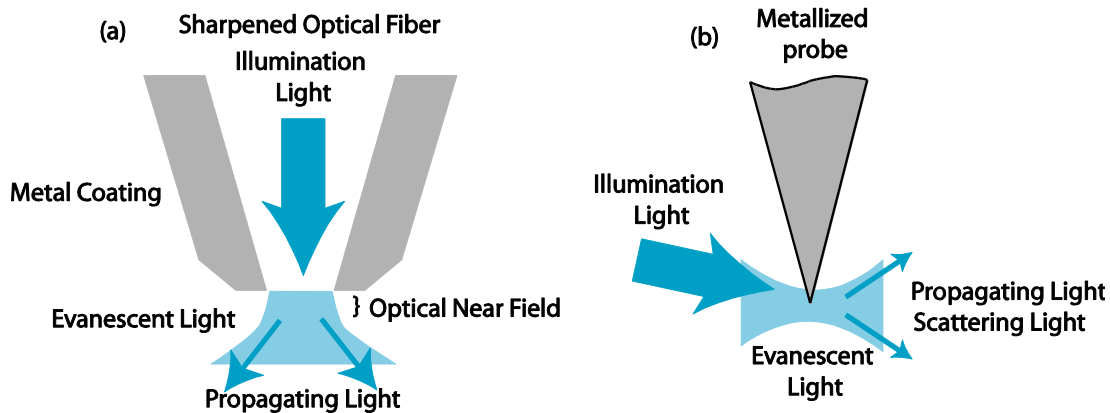


Figure 4.66. Principle of NSOM: Can be simply modeled by the electromagnetic interaction of two very closely positioned nano-objects, which represent a probe and sample. Aperture-type (a) and tip enhanced (b) SNOM principles.

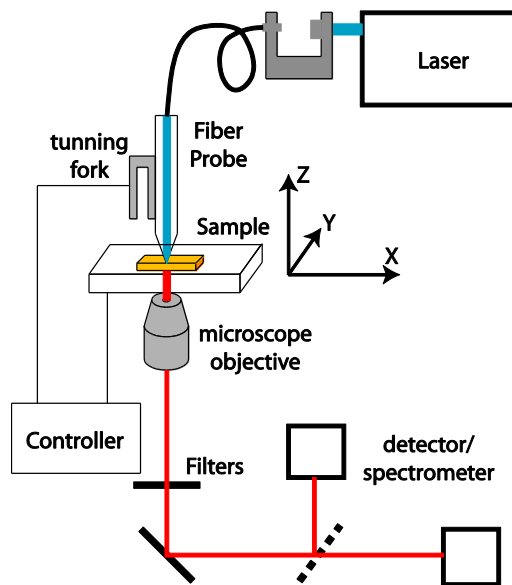


Figure 4.67. Aperture type SNOM. Specially etched optical fiber is used as a probe. It is glued to the side of the tuning fork which lateral oscillations can be excited. Probe-sample interaction is controlled through the resonant frequency shift of tuning fork owing to lateral oscillations damping. Scattered light is collected with microscope objective, filtered and analyzed.

#### 4.3.9 Apertureless techniques

Apertureless (or tip-assisted) techniques are based on the near-field optical phenomena as well, but do not require passing the light through an aperture (Fig. 4.66). A nano-antenna (a specially prepared metal-coated AFM probe) is excited by laser light tightly focused at its apex. Under specific conditions, the nano-antenna can localize and enhance the excitation electromagnetic field

and/or sample optical response in very close proximity to its apex. Thus, tip of the nano-antenna effectively works as a localized “nano-emitter” or “nano-scatterer” of light. When scanning the sample with respect to this nano-antenna, the obtained map of optical response (Raman or Rayleigh scattering, fluorescence, etc.) from the sample has lateral resolution which is determined by size of the electromagnetic field localization at the nano-antenna tip.

One of the most promising types of tip-assisted optical imaging techniques is Tip-Enhanced Raman Scattering (TERS or “nano-Raman”) where special nano-antenna with sharp metal tip is used to locally enhance the Raman scattering signal from a sample (Fig. 4.68). Fig. 4.69 shows enhancement of Raman spectra for carbon nanowire in proximity to the metal tip. Raman spectra provide structural information about chemical species and TERS can be used for identification of chemical substances and their distribution onto the sample surface.

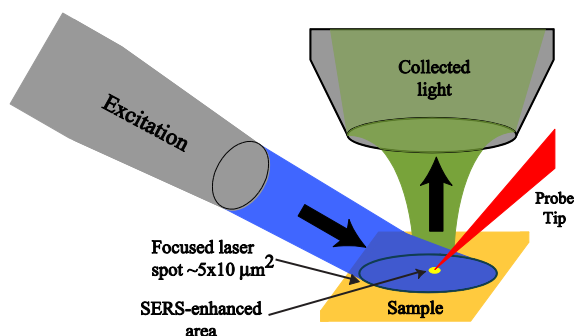


Figure 4.68. Apertureless type TERS. Specially inclined etched noble metal tip is approached to the sample. Spot around tip is illuminated with excitation laser and scattered light is collected with microscope objective.

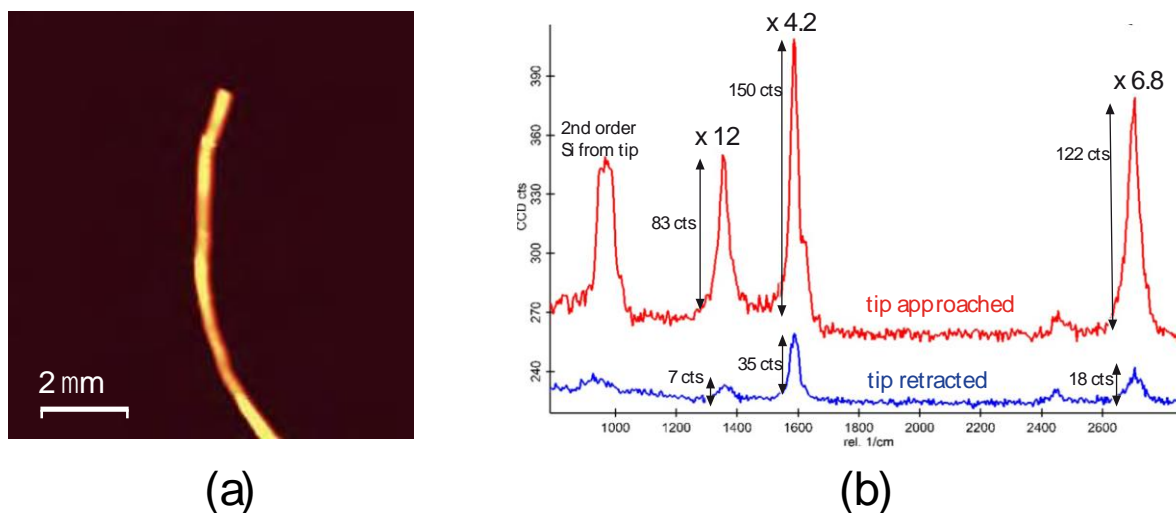


Figure 4.69. (a) AFM topography image of the carbon nanowire used for the TERS experiments in this study. (b) Tip-enhanced Raman spectroscopy on the nanowire. Blue spectrum: tip retracted, red spectrum: tip approached; enhancement of some lines are shown numerically. (from application

note "Tip-Enhanced Raman Spectroscopy (TERS) for Chemical Imaging at the Nanoscale" of Witec, <http://www.witec.de>).

#### 4.3.10 References

<https://www.ntmdt-si.com>

<http://www.witec.de>

<http://www.veeco.com>

<http://www.nanosensors.com>

[https://www.nobelprize.org/nobel\\_prizes/physics/laureates/1986/binnig-lecture.html](https://www.nobelprize.org/nobel_prizes/physics/laureates/1986/binnig-lecture.html), Nobel Prize in Physics 1986.

1. Albrecht T., Grütter P. et al. (1991), Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity, *Journal of Applied Physics*, 69 (2), pp ; 668-673.
2. Binnig G., Rohrer H. (1987), Scanning tunneling microscopy – from birth to adolescence, *Reviews of Modern Physics*, 59 (3), pp. 615-625.
3. Dürig U., Pohl W. et al. (1986), Near-field optical-scanning microscopy, *Journal of Applied Physics*, 59 (10), pp. 3318-3327.

## 4.4 Magnetometry

Magnetometry is the technique of measuring and mapping patterns of magnetism in different materials. Magnetometry measures perturbations in the ambient magnetic field caused by contrasts in magnetic susceptibility, i.e. the ability of substance to take on an induced magnetism caused by its immersion in the magnetic field.

The result of magnetometry analysis is the integral magnetic moment (or its projection on the direction of the external magnetic field) of the studied object in the external magnetic field. The pattern of the distribution of magnetic moments in the object can be obtained using a number of methods, such as magnetic force microscopy, neutron diffraction methods with polarized neutrons, etc. The unit of magnetic moment in the SI system is  $1 \text{ A} \cdot \text{m}^2$ . In the Gaussian system of units (as well as CGSM) which is the most common and convenient in magnetometric studies, the unit of magnetic moment is usually denoted as *emu* (*electromagnetic unit* of the magnetic moment). The ratio  $1 \text{ emu} = 10^{-3} \text{ A} \cdot \text{m}^2$  gives the relation between units of these two systems, that is the value of unit magnetic moment, the Bohr magneton, corresponding to the "quantum" of the magnetic moment of the electron being  $1 \mu_B = e \cdot \hbar / 2m_e = 9.27 \cdot 10^{-24} \text{ A} \cdot \text{m}^2 = 9.27 \cdot 10^{-21} \text{ emu}$  ( $\hbar$  is the Planck-Dirac constant,  $e$  is the elementary electric charge,  $m_e$  is the mass of the electron). Strength  $H$  of magnetic field applied to the studied object is measured in A/m (SI) or Oersteds (Oe in CGSM).

The results of magnetometry are the unique origin of information describing properties of modern materials, including nanostructured, where "direct" observation of the studied objects, nanostructures, is often problematic because of their small size. Analysis of magnetic properties of such materials allows "feeling" their structural peculiarities and analyzing indirectly atoms ordering inside material (crystalline ordering, nanoparticles size and distribution in composite materials, etc.).

### 4.4.1 Methods of magnetometry studies

The commonly used methods of magnetometry can be divided into two groups, "force" methods that measure force acting on magnetized sample in a variable magnetic field, and "induction" methods that measure the voltage induced by changing magnetic flux (for example, when removing a sample from a coil, or when the magnetized sample vibrates near the coil). The first group of techniques includes such methods as torque magnetometry and alternating gradient magnetometry (AGM) which is based on measuring the force induced by the gradient of a magnetic field. The sensitivity of AGM-magnetometry is rather high and close to  $10^{-8} \text{ emu}$ . The principle of the AGM magnetometer acting described in [1] is shown in Figure 4.70. During the measurement, the sample is cut into 3 mm x 3 mm, mounted on a non-magnetic probe (or at the end of a fiber) and suspended in the gap between a pair of the magnet poles and coupled AC field coils. The magnet poles provide a strong field to magnetize the sample in the measuring directions. The alternating gradient field coils exert a periodic force on the sample and deflect the sample with the magnitude proportional to the moment of the sample. The deflection of the probe is measured by a piezoelectric sensor

attached to the probe. When the AC field is tuned at the mechanical resonant frequency of the probe, the deflection of the probe is amplified (vibration amplitude increases). Therefore, the hysteresis loop of the sample can be obtained by detecting the corresponding piezoelectric voltage signal from the piezoelectric sensor [1].

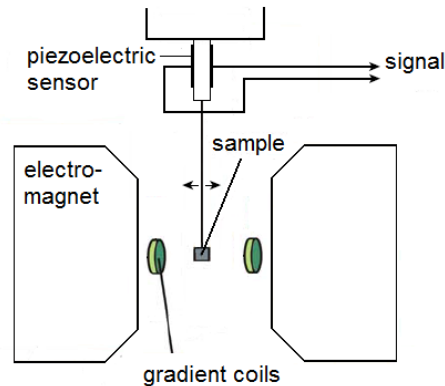


Figure 4.70. Schematic image of AGM magnetometer

The second group of magnetometry techniques includes vibrating sample magnetometry (VSM) and SQUID magnetometry. VSM (or Foner magnetometer) is the prevalent instrument to measure the basic magnetic properties of materials as a function of magnetic field, temperature and time. VSM is so named because the sample physically vibrates, typically at frequencies between 50 to 100 Hz. The idea of the method is based on the flux change in a coil when a magnetized sample is vibrated near it. The sample, commonly a small disk, is attached to the end of a nonmagnetic rod, the other end of which is fixed to a mechanical vibrator (Figure 4.71). The movement of the sample in the applied magnetic field alters the magnetic flux through the pick-up coils, which causes an induction voltage in the coils according to Faraday's law of magnetic induction. The induction voltage is proportional to the sample's magnetic moment. The small alternating magnetic flux is amplified, usually with a lock-in amplifier which is sensitive only to signals at the vibration frequency. The apparatus is calibrated with a specimen of known magnetic moment. The sensitivity of this method is usually about  $10^{-6}$  emu, but it can reach even  $10^{-8}$  emu.

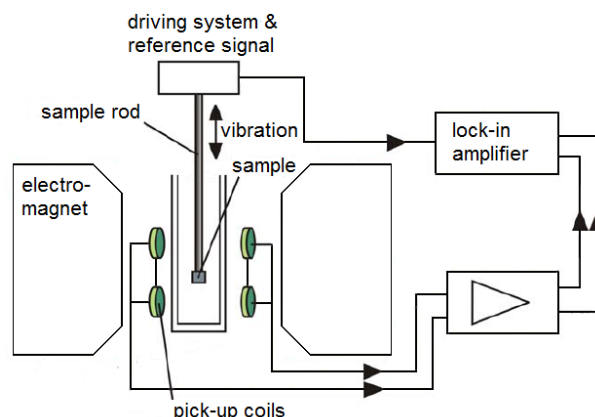




Figure 4.71. Schematic image of VSM magnetometer

SQUID (super-conducting quantum interference device) magnetometry is based on the Josephson junction effect, i.e. on the tunneling of superconducting electrons across a very narrow insulating gap, called a Josephson junction, between two superconductors (Figure 4.72). SQUID magnetometer utilizes the quantum properties of closed superconducting circuits and works basically as a converter of flux-to-voltage with high sensitivity. As the sample passes through the SQUID pick-up coils, the magnetic flux through the superconducting loop changes due to this movement. A changing magnetic flux through the ring generates a voltage and a current in the ring, according to Faraday's Law. The voltage of the junctions oscillates with a period equal to the quantum flux, i.e. each voltage step corresponds to the passage of a single flux quantum across the boundary of the ring, so that the change of the magnetic flux can be calculated by counting the number of oscillations and converted back to moment. Since a superconducting Josephson device requires low-temperature operation, it is usually used in conjunction with a superconducting solenoid. SQUID measurements over a range of fields and temperatures are time consuming. The sensitivity of SQUID is usually up to  $10^{-8}$  emu.

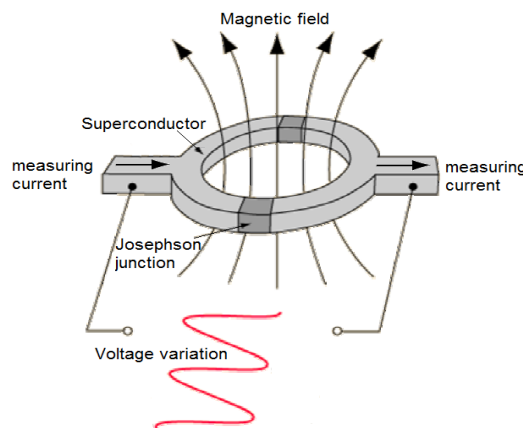


Figure 4.72. Schematic image of the operating principle of SQUID flux sensor

Magnetic force microscopy (MFM), a variation of conventional atomic force microscopy, is commonly used to detect the magnetic structure of materials. Schematic image of operating principle of magnetic force microscope [1, 2] is presented in Figure 4.73. The measurement process includes two stages. In the first pass, topographical data is taken in the tapping mode. The tip is then raised to a certain scan height in the second pass with a constant separation between the tip and sample surface. During the second pass, the long-range magnetic forces are detected which shift the resonance frequency of the cantilever and produce the magnetic image. The short-range Van der Waals forces vanish and thus the influence of topography can be minimized in the second pass. Therefore both the topography and the magnetic image can be obtained simultaneously for the same area. The tip is coated with a thin layer of magnetic material, which experiences a measurable force when it enters the field gradient where a domain wall meets the sample surface. The force acting on the tip can be measured optically, by the deflection of the cantilever holding the tip, or from the change in amplitude of the cantilever motion when set into resonant oscillation

[1]. MFM requires minimal surface preparation, and works on both conducting and insulating materials. The MFM offers the prospect of observing magnetic structures at extremely high resolution.

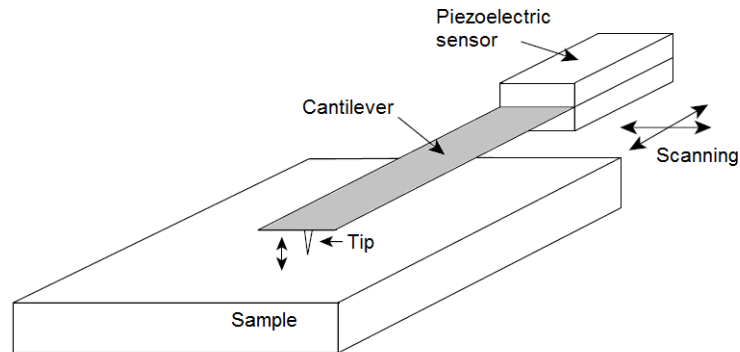


Figure 4.73. Schematic image of operating principle of magnetic force microscope

#### 4.4.2 Measuring and interpretation of magnetization curves

The most common source of information about magnetic properties of the studied object is so-called magnetization curves presenting the dependence of magnetic moment  $\mu$  on the external magnetic field  $H$ . Magnetization  $M$  is the magnetic moment of a unit volume of the material. The procedure for obtaining these dependences assumes the initial magnetization of the object at the maximal magnetic field  $H_{\max}$ . Further, the magnetic field decreases gradually to zero, after which a magnetic field in the opposite direction is applied to the sample, gradually reaching  $-H_{\max}$ . The registration of the magnetic moments corresponding to different values of applied field  $H$  allows obtaining the magnetization curve schematically shown in the Figure 4.74.

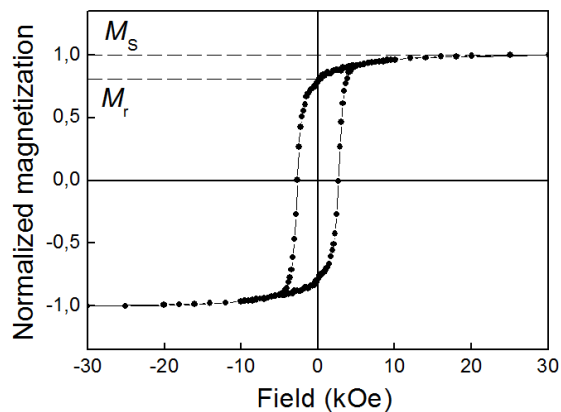


Figure 4.74. Typical magnetization curve  $M(H)$

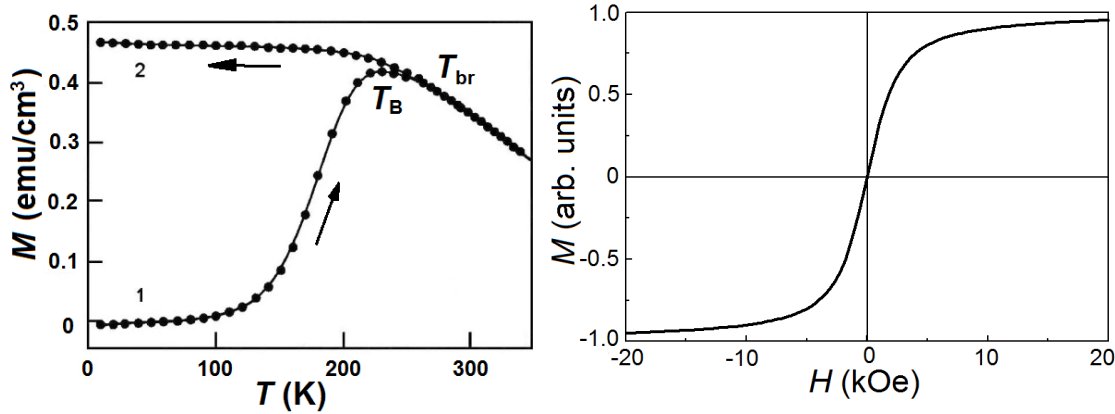
It should be noted that magnetization curves of magnetic materials exhibit often hysteresis due to the irreversibility of their magnetization reversal. Hysteresis is characterized by a mismatch of the descending (from  $H_{\max}$  to  $-H_{\max}$ ) and ascending (from  $-H_{\max}$  to  $H_{\max}$ ) branches of  $M(H)$

dependences, as it is shown in Figure 4.74. The registration of the complete magnetization curve (in two directions of  $H$  variation) makes it possible to determine the so-called coercive field  $H_C$  – negative magnetic field leading to zero magnetization. It is important to mention that initial saturation of magnetization is a necessary condition for determination of  $H_C$  and recording the complete magnetization curve. Saturation of the sample magnetization is achieved in the case of aligning the magnetic moments of all its structural elements (atoms, ions) in one direction of the external field. The magnetization thus reaches the maximal value  $M_S$ . Otherwise, if  $H_{\max}$  is insufficient for saturation of the studied sample, so-called minor hysteresis loops are detected, the field leading to zero magnetization obtained from these curves being not equal to  $H_C$  value. In addition to  $H_C$  value, saturation magnetization  $M_S$  and the remanent magnetization  $M_r$  can be determined from the magnetization curves (see Figure 4.74). The latter parameter is the value of magnetization characterizing the sample after decreasing the external field down to zero. For magnetization curves possessing hysteresis, the squareness  $S = M_r/M_S$  of the loops is one more important parameter.

Magnetization curves  $M(H)$  can be measured (i) at different temperatures for investigation of temperature effects in materials (phase transitions, thermal fluctuations of magnetic moments, superparamagnetic effects in nanoobjects, etc.), and (ii) for different spatial orientations of the external magnetic field with respect to the sample to study the effects of magnetic anisotropy.

#### 4.4.3 Analysis of temperature dependences of magnetization

Changes in the studied object temperature can provide significant modification of magnetization curves shape. Both  $H_C$  and  $M$  values of the sample can be modified. Measurement of temperature dependences of magnetization  $M(T)$  allows determination of temperature corresponding second-order phase transition in material, e.g. from ferromagnetic to the paramagnetic state (Curie temperature), which is characterized by a decrease in the magnetization down to zero with temperature increase. In addition, the objects studied by modern materials science often possess a small size (nanometers and even sub-nanometers) and, as a consequence, are characterized by a much more diverse magnetic effects and states, as compared with corresponding bulk materials. For example, the effect of superparamagnetism (SP) is typical for magnetic nanoobjects with linear size less than 10 nm. This effect originates from thermal fluctuations of magnetic moments in the objects with the energy exceeding energy of their magnetic anisotropy. Analysis of nanoobjects magnetic state can be performed by measuring and analyzing their temperature dependences of magnetization (or magnetic moment) in field cooled and zero-field cooled modes (FC and ZFC, respectively), i.e. measured after cooling in zero magnetic field (ZFC) and in small but non-zero magnetic field  $H_{FC}$  (FC). Typical FC and ZFC curves are shown in Figure 4.75.



*a*

*b*

Figure 4.75. Temperature dependences  $M(T)$  of superparamagnetic nanoobjects measured in ZFC (1) and FC (2) modes (*a*) as well as typical shape of their magnetization curve  $M(H)$  in SP state (at  $T > T_B$ ) (*b*)

They demonstrate the shape characteristic for particles in the SP state which possess dispersion in size. When cooling to low temperature without external field, disordering in the directions of magnetic moments originating from random orientation of the magnetic axes in a system of non-interacting nanoparticles keeps fixed (blocked). This leads to a zero value of total magnetization of the system. When the sample is heated in a small magnetic field (ZFC curve), its magnetization increases as a result of unblocking the magnetic moments of the particles and aligning them in the direction of the applied magnetic field. Peak of ZFC curve occurs due to the particles thermal energy  $k_B T$  increase with  $T$  and its reaching the value of magnetic anisotropy energy  $\Delta E_a$ . The temperature corresponding to maximal value of sample magnetization (peak of ZFC curve) defines blocking temperature  $T_B$ . Further increase in  $T$  providing domination of thermal energy over the anisotropy energy  $E_a$  leads to a transition of nanoparticles from nonequilibrium "blocked" state into SP state. It should be noted that magnetic granules with different sizes become SP at different temperatures. In this case, blocking temperature  $T_B$  determines the temperature of transition to the SP state for smaller particles, whereas particles of maximal size transit into SP state at the temperature corresponding coincidence of FC and ZFC curves (the so-called bifurcation temperature  $T_{br}$ ). One can see from the Figure 4.75 that the magnetic moments of majority particles are unblocked already at 250 K, but the weak splitting between FC and ZFC curves remains up to 340 K. Thus, the difference between these curves in the temperature region of  $T_B > T > T_{br}$  is associated with the dispersion of magnetic granules in size.

Nonzero magnetization of the material cooled in a magnetic field down to low temperatures (FC curve in Fig. 4.75) is explained by the fact that magnetic moments of the granules are fixed with  $T$  decrease, some of them being aligned in the direction of the applied field. Above  $T_B$ , magnetization curve  $M(H)$  of SP nanoobjects demonstrates characteristic "S-type" shape unsaturated over a wide range of fields accompanied with smooth and slow saturation, as well as zero  $H_C$  value (Fig. 4.75b).

$T_B$  value allows determination of SP granules size. In the simplest case, SP relaxation (or system thermal stability) in zero magnetic field can be described by Neel relation:

$$\tau = \tau_0 \exp\left(\frac{\Delta E_a}{k_B T}\right), \quad (4.56)$$

where  $\tau$  is relaxation time,  $\tau_0$  is characteristic time,  $\Delta E_a = KV$  is the energy barrier between two energy levels of system with minimal energy depending on magnetic anisotropy of studied system,  $K$  – constant of magnetic anisotropy,  $T$  – temperature.

Relaxation of magnetization of single particle through the energy barrier  $\Delta E_a$  in  $H$  field is described by relaxation time  $\tau$ . Taking into account the magnetization reversal by the coherent rotation mechanism and random direction of granules magnetic moments, one can obtain [3]

$$\Delta E_a(H) = KV \left(1 - \frac{H}{H_0}\right)^n, \quad (4.57)$$

where  $H_0$  is the external field eliminating energy barrier. As  $T_B$  being the temperature at which energy barrier is comparable with thermal energy, one can obtain the next relation from (1) for typical characteristic time of  $\tau_0 = 10^{-9}$  s

$$25k_B T_B(H) \approx \Delta E_a(H). \quad (4.58)$$

Extrapolation of  $T_B(H)$  in zero field leads to

$$25k_B T_B(0) \approx KV. \quad (4.59)$$

Basing on the value of  $K$  constant for nanoparticles in studied materials and estimated  $T_B$  value from  $M(T)$  dependences, it is possible to calculate the volume of magnetic particle. It should also be taken into account that  $K$  value of nano-sized particles is probably several orders of magnitude higher than for a bulk material due to the significant contribution of surface anisotropy [1, 3]. For example, for iron nanoparticles  $K \sim 10^6$  J/m<sup>3</sup>, whereas  $K = 4.8 \cdot 10^4$  J/m<sup>3</sup> for bulk  $\alpha$ -Fe [1, 3]. The volume of particles estimated by this method gives information on the size of SP granules, because exactly these granules participate in the peak of the ZFC curve formation.

One more way to determine the size of SP nanoobjects basing on the results of magnetometry is to approximate their magnetization curves in SP state, i.e. for any  $T > T_B$ , with Langevin functions  $L = \coth(a) - 1/a$ . This function describes magnetization of paramagnetic material [4] and is applicable to the description of particles in a SP state, since there is no magnetic interaction between them. For SP particles

$$M(H, T) = M_s \left[ \coth\left(\frac{\mu H}{k_B T}\right) - \frac{k_B T}{\mu H} \right], \quad (4.60)$$

where  $\mu = M_s V$  presents magnetic moment of nanoobject with  $V$  volume,  $M_s$  is saturation magnetization,  $H$  is the external magnetic field.

Assuming spherical shape of nanoparticles and log-normal distribution of their sizes

$$f(D) = \frac{1}{\sqrt{2\pi} \ln \sigma} \exp \left[ \frac{(\ln D - \ln \bar{D})^2}{2(\ln \sigma)^2} \right], \quad (4.61)$$

where  $\bar{D}$  is mean value,  $\sigma$  is dispersion in values, then the resulting magnetization can be presented as

$$M(H, T) = M_s \int_0^{V_{activ}} L \left( \frac{M_s V H}{k_B T} \right) f(V) dV, \quad (4.62)$$

where  $V_{activ} = 25k_B T / K$  is the size of maximal particle in SP state. This relation describe the share of  $M(H)$  presented in Figure 4.75b.

#### 4.4.4 Magnetic anisotropy and angular dependences of magnetic parameters

SP state characterizes nanoobjects only in a narrow size range (about several nanometers or less), which, moreover, do not exhibit magnetic interaction. Nanoobjects of a larger size, as well as continuous media characterized by magnetic ordering, exhibit thermally stable configurations of magnetic moments. Such materials generally demonstrate diverse magnetic parameters in different directions, i.e. they have a preferential direction for their magnetic properties realization (one or several), such as axis of easy magnetization. The dependence of materials magnetic properties on chosen direction indicates the presence of magnetic anisotropy. The most common type of magnetic anisotropy is uniaxial anisotropy, when only one easy-magnetization axis is characteristic for material (magnetic system). The origin of magnetic anisotropy of materials is in their crystalline anisotropy, shape anisotropy, magnetostriction, surface effects, exchange interaction, etc. Occurrence of magnetic anisotropy in material influences on the shape of its magnetization curves that varies significantly in different spatial orientations. The anisotropy axis dictates preferential direction for magnetic moments in the sample in the absence of external magnetic field. For small rotations of magnetization vector (magnetic moment) away from an easy direction, the anisotropy of the system acts like a magnetic field trying to hold it parallel to the axis. This field is called the anisotropy field and denoted as  $H_a$ . The anisotropy field is parallel to the easy direction and of a magnitude such that for small angular deviations from the easy axis (small angle  $\theta$ ) it exerts the same torque  $L = H_a \cdot M_s \cdot \sin \theta$  on  $M_s$  as the anisotropy itself [1]. The physical meaning of the anisotropy field providing the way for its calculation from magnetometry data implies the magnitude of the field necessary to magnetize the system in the direction of its hard axis, i.e. at  $90^\circ$  to its easy axis [1]. Anisotropy constant  $K$  is the main parameter characterizing magnitude of

magnetic anisotropy. It corresponds to energy barrier to be overcome by magnetic moment for orientation along the direction corresponding to the energy minimum.

Typical set of magnetization curves for a system with uniaxial magnetic anisotropy obtained at different angles  $\alpha$  to the axis of its easy magnetization is shown in Fig. 4.76. The shape of these curves is given by Eq. (8)

$$h \cdot \sin \theta + \frac{1}{2} \cdot \sin 2(\theta - \alpha) = 0 \quad (4.63)$$

where  $h = H/H_a$  is the external magnetic field normalized by anisotropy field value,  $\theta$  is the angle between magnetic moment orientation and the external field  $H$ .

This equation was obtained by Stoner and Wohlfarth [5] for a system with uniaxial magnetic anisotropy originating from shape anisotropy (for elongated ellipsoid, see Fig. 4.76). Its magnetization reversal in the external magnetic field is supposed to occur by coherent rotation of magnetic moments. Coherent rotation implies keeping parallel orientation of the magnetic moments of all components of the system (atoms, ions) permanently during the magnetization reversal. The energy minimum condition  $dE/d\theta = 0$  determines the equilibrium position of the magnetic moment in the external field, where  $E$  is the density of magnetic energy described by relation between two components, magnetic anisotropy energy and the energy of magnetic body in the external field,  $K \cdot \sin(\theta - \alpha) - M_S \cdot H \cdot \cos\theta$ .

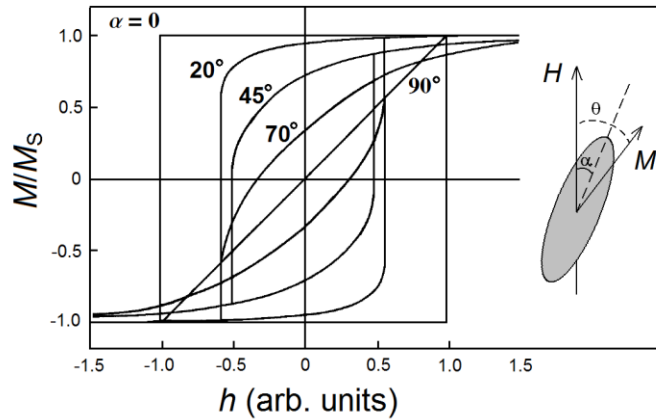


Figure 4.76. Normalized magnetization curves  $M/M_S(h)$  of the system with uniaxial magnetic anisotropy obtained at different angles with respect to its easy magnetization axis accompanied with schematic illustration of uniaxial system under external field

The legitimacy of the equation (8) is extended both for different types of uniaxial magnetic anisotropy (crystalline, stress anisotropy, etc.) [1], as well as for various magnetoanisotropic systems. It is shown that magnetic moments rotation is typical fragment of any magnetization reversal process occurring by various mechanisms. Additionally, coherent rotation is the simplest mechanism for model representation. It is appropriate for any orientations  $\alpha$  of external magnetic field with respect to easy magnetization axis of the system, in contrast to incoherent rotation mechanisms that cannot be realized for  $\alpha > 45^\circ$ . Being an ideal model representation that cannot be

completely realized for real objects, the coherent rotation mechanism however shows a good agreement with the experimental magnetometry data within the approximation of the magnetization curves of real anisotropic systems.

Application of the Stoner-Wohlfarth model for approximation of magnetization curves of magnetoanisotropic systems allows not only verification of magnetization reversal mechanism, but also determination the magnetic anisotropy field  $H_a$  and corresponding anisotropy constant  $K = (H_a \cdot M_s)/2$ . One more parameter that can be determined from the approximation is the angle  $\alpha$  characterizing orientation of easy magnetization axis of the system with respect to the direction of the applied magnetic field  $H$ . Mentioned parameters allow us to characterize magnitude and direction of the magnetic anisotropy of studied system.

The orientations corresponding to  $\alpha = 0$  and  $\alpha = 90^\circ$ , i.e. axes of easy and hard magnetization, are the most important for determining the parameters characterizing magnetic anisotropy. The field  $H_a$  of an ideal system with uniaxial anisotropy equals to the coercive field  $H_C$  of  $M(H)$  curve measured at  $\alpha = 0^\circ$  and to the saturation field of  $M(H)$  curve measured at  $\alpha = 90^\circ$ . The latter follows from the definition of the anisotropy field as a field necessary for magnetizing the system in the direction of its hard axis. It offers the simplest way for  $H_a$  estimation from the shape of the magnetization curve measured strictly along the hard axis of the system.

It should be noted that real systems often demonstrate several types of magnetic anisotropy simultaneously. For instance, the shape of object (as a film or elongated nanostructure) induces shape anisotropy that reduces or intensifies the "internal" anisotropy of the object due to its crystal structure, internal tension and deformation, magnetostriction effects, etc. Shape anisotropy arises in the object with a specific ratio of its sizes in different direction, i.e. at least one of its three sizes is finite (cylinder, film, sphere, etc.), due to the presence of free magnetic poles at the object surfaces, when magnetized in the external field. As a consequence, demagnetizing field  $H_d$  oriented oppositely to the magnetization vector  $M_s$  (magnetic moment) appears. Thus, the resulting field that influence on the magnetic moment of such an object (namely, spins inside the object) in an external field  $H$  is the difference between the absolute values of the external and demagnetizing fields  $H - H_d$ . In other words, object is magnetized at lower extent in an external magnetic field  $H$  than the value of this field assumes in the direction where demagnetization is present, i.e. in which the size of object is finite. The "degree of un-infinity" of an object is determined by its demagnetizing factors  $N$  in different directions. For example, a thin film has a zero demagnetizing factor (and, correspondingly, shows a zero demagnetizing field) in the film plane, whereas  $N = 4\pi$  (and  $H_d = 4\pi \cdot M_s$ ) along its normal.

Magnetization reversal of the system demonstrating several types of magnetic anisotropy is governed by "effective" magnetic anisotropy that is determined by the competition between them. Therefore, anisotropy field defined as the saturation field of magnetization curve in hard axis corresponds to the effective parameter  $H_{\text{eff}}$ , the field of "effective" magnetic anisotropy. The magnitude of effective magnetic anisotropy constant  $K_{\text{eff}}$  that determines magnetization reversal process depends both on the magnitude and mutual orientation of different components of magnetic



anisotropy detected in the studied sample [1]. The simplest way to determine the parameter  $K$  of a system with uniaxial magnetic anisotropy from the results of magnetometry is based on estimating the area between the magnetization curves measured in hard and easy axes, regardless of the fact, whether the system is reversed under the action of single magnetic anisotropy or the resultant of several components. According to this method, the area bounded by  $M(H)$  curves measured at  $\alpha = 0$  and  $\alpha = 90^\circ$  in the first quadrant is calculated. Strictly, this calculation can be carried out according to

$$K_{eff} = \int_{0_{in}}^{M_s} H dM - \int_{0_{out}}^{M_s} H dM, \quad (4.64)$$

where the first integral corresponds to the hard axis (in), and the second integral describes the easy axis (out). Schematically, interpretation of the method is shown in Fig. 4.77.

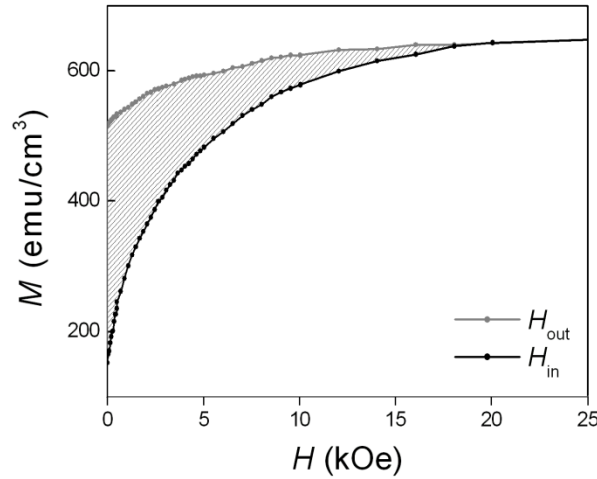


Figure 4.77. Magnetization curves  $M(H)$  of the system with uniaxial magnetic anisotropy obtained along its hard axis (in) and easy axis (out), as well as the square bounded by these curves (hatching) estimated for  $K$  calculation

One more important method for determination of magnetization reversal mechanisms in magnetic systems, as well as for estimation of parameters characterizing their magnetic anisotropy, is to measure the angular dependences of the coercive field. The  $H_C$  values can be obtained from corresponding magnetization curves measured at different  $\alpha$  angles to easy magnetization axis of the system (Fig. 4.76). Typical examples of  $H_C(\alpha)$  dependences corresponding to different magnetization reversal mechanisms are shown in Fig. 4.78.

According to the Stoner-Wohlfarth theory described above which is based on the assumption of non-interacting magnetically isolated grains, coercivity is defined by equations [5-7]

$$\begin{cases} H_C(\alpha)/H_C(0) = (\cos^{2/3} \alpha + \sin^{2/3} \alpha)^{-3/2} & \text{for } \alpha \leq 45^\circ, \\ H_C(\alpha)/H_C(0) = \cos \alpha \cdot \sin \alpha & \text{for } 45^\circ < \alpha \leq 90^\circ. \end{cases} \quad (4.65)$$

These relations set the shape of  $H_C(\alpha)$  curve corresponding to coherent rotation mechanism and presented in Figure 4.78.

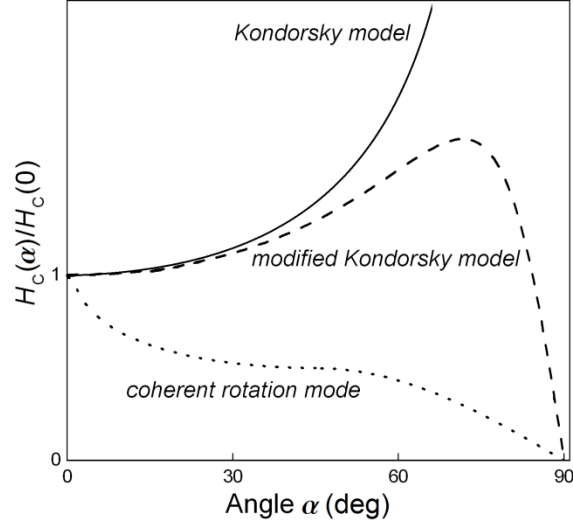


Figure 4.78. Theoretical angular dependences  $H_C(\alpha)$  corresponding to magnetization reversal according to Stoner-Wohlfahrt (coherent rotation) and Kondorski (domain walls motion) models

The dependence of  $H_C$  on the orientation of external magnetic field  $H$  for ideal multidomain system is described in the frame of Kondorski model [8, 9] by the relation

$$H_C(\alpha) = H_C(0) / \cos \alpha, \quad (4.66)$$

where  $H_C(0)$  is the coercive field measured along its easy magnetization axis. It should be noted that in the real systems the Kondorski mechanism is not observed in the whole range of  $\alpha$  angles. As a rule, it is replaced by the mechanism of coherent rotation of magnetic moments described by Stoner and Wohlfarth [5] at the angles close to hard magnetization axis ( $\alpha \sim 90^\circ$ ). The coercive field is then reduced to zero along the axis of hard magnetization. Such a "modification" of the Kondorski mechanism by an additional contribution of coherent rotation can be described by the relation [10]

$$H_C(\alpha) / H_C(0) = \frac{(N_x + N_N) \cos \alpha}{N_z \sin^2 \alpha + (N_x + N_N) \cos^2 \alpha}, \quad (4.67)$$

where parameter  $y = (N_x + N_N) / N_z$  determines the relation between intrinsic magnetic anisotropy of system and demagnetizing effects originating from magnetostatic energy,  $N_x$  and  $N_z$  are demagnetizing factors of the system (here, for the film) in two orthogonal directions, and  $N_N = H_a / M_S$  is formal parameter similar to demagnetizing factor that characterizes other type of anisotropy than magnetostatic (shape). Estimation of  $N_N$  value from approximation of  $H_C(\alpha)$  dependence of the system that demonstrates magnetization reversal mainly by domain walls motion allows determining its anisotropy field from  $N_N = H_a / M_S$  relation.

#### 4.4.5 References

1. B.D. Cullity, C.D. Graham. Introduction to magnetic materials. 2nd ed. Wiley–IEEE Press, Hoboken, N.J. (2009). 568 p.
2. D. Sarid. Scanning Force Microscopy with Application to Electric Magnetic and Atomic Forces. Oxford University Press (1994).
3. N.M. Dempsey, L. Ranno, D. Givord, J. Gonzalo, R. Serna, G.T. Fei, A.K. Petford-Long, R.C. Doole, D.E. Hole. J. Appl. Phys. 90 (12), 6268 (2001).
4. I.S. Jacobs, C.P. Bean. In Magnetism. Vol.3. (Eds G.T. Rado, H. Shuhl). Academic Press, New York (1963). P. 271.
5. E.C. Stoner, E.P. Wohlfarth. Phil. Trans. Roy. Soc. A 240, 599 (1948).
6. E.C.S.a.E.P. Wohlfarth, IEEE Trans. Magn. 27 (4), 3475 (1991).
7. S. Chikazumi. Physics of Ferromagnetism. Oxford University Press, New York (1997).
8. E. Kondorski. J. Phys. (USSR) 2, 161 (1940).
9. N.P. Suponev, R.M. Grechishkin, M.B. Lyakhova, Yu.E. Pushkar. J. Magn. Magn. Mater. 157/158, 376 (1996).
10. M. Mathews, E.P. Houwman, H. Boschker, G. Rijnders, D.H.A. Blank. J. Appl. Phys. 107, 013904 (2010).

## 4.5 The essence of ellipsometric method, the fundamental ellipsometry equation

### 4.5.1 Introduction

The development and widespread practical application of ellipsometric control method, which has been known since the end of the 18<sup>th</sup> century, is caused by the requirements of semiconductor microelectronics that demonstrated the rapid growth in the middle of the 20<sup>th</sup> century. A simple, affordable and reliable method for high-precision control of the thickness of the semiconductor structures' dielectric layers was required. The ellipsometry was the best one to satisfy all these requirements. Moreover, the possibilities inherent in the physical principles of the method turned out to be much wider and allowed not only to measure the thickness of the layers, but also to characterize their physical and chemical properties as well as to investigate the processes at the surface of the coating. The development of this method continued to follow several connected tracks.

First, the method of hardware was developed: ellipsometers of wide spectral range, high-speed ellipsometers as well as devices with specific capacity were created. At the same time, the methodical bases were also in progress, the models for the ellipsometric measurements interpretation were designed, the algorithms and methods for the numerical solution of the ellipsometry problems were formulated. Ellipsometry went beyond microelectronics and extended to other sciences: solid-state physics, surface physics, materials science, optical coating technology, polymer chemistry and electrochemistry, biology, medicine, etc.

Ellipsometry is a highly sensitive and accurate method for studying surfaces and interfaces of various media.

The method is based on the change in polarization state of light after reflection from the boundary between two media. Incident on the surface of the plane-polarized light by reflection and refraction becomes elliptically polarized, which is determined by the optical properties of the reflective surface, and the thickness and refractive index of the film located thereon.

The light source used in ellipsometry is monochromatic or laser radiation. There is also the direction of the spectral ellipsometry in a wide range of wavelengths (from UV to near IR), which is used in studies of the atomic composition of inhomogeneous and anisotropic surfaces and films. Ellipsometers are used to study the physicochemical properties of the surface, its morphology, to measure the thickness of multilayer structures and the characterization of the optical properties of thin films.

Using ellipsometry it is possible to characterize the composition of the composite compounds, the density of foreign nanoinclusions, structural quality of the material, the quality of the interfaces; detect changes caused by temperature changes or exposure to electrical, magnetic and mechanical fields. In addition, the advantage of this method is non-destructive and noninvasive measurement effects. The method relates to the rapid methods, allowing using the ellipsometry for monitoring directly in the creation of structures or the study of various types of physical effects. The method

is widely used in the production and research of semiconductor materials and devices, crystal physics, electronics, optics and medicine.

Application:

- the measurement of the optical constants and thickness of thin films of materials, diagnostics of the layers with gradient properties, the study of the energy band material structure.
- the characterization of mechanical, structural, physical and chemical properties of materials, and a micro-porosity and density measurement of microinclusions, surface topography.
- residual control of the surface layers and the degree of surface treatment and mechanical stresses in the fibers.
- research piezoelectric, electromagnetic, magneto-optical properties of materials [1-4].

There are special installations, which are called ellipsometers, to obtain data using the above-mentioned method. Figure 4.79. shows one of such devices.



Figure 4.79. The ellipsometer

#### 4.5.2 Ellipsometric measurements

All ellipsometric measurements are based on the transformation of light polarization by optical linear elements, i.e. by polarizing prisms and phase shifters. The essence of the measurements can be explained by the circuit in Figure 4.80.

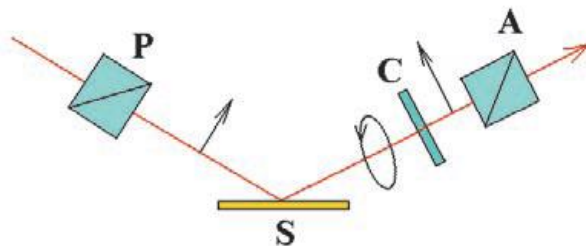


Figure 4.80 The scheme of ellipsometry measurements

P is a polarizer, S is a reflecting surface, C is a compensator (phase-shifting element), A is an analyzer

The protometric ellipsometer operating principle can be described as follows. The light beam, emitted from the source, passes through a polarizing prism (polarizer P) and becomes linearly polarized. After the reflection from the surface S polarized light becomes elliptical. These ellipse parameters are analyzed using the phase-shifting plate (compensator C) and the second polarizing prism (analyzer A). The light beam is completely extinguished at some azimuth positions of the optical elements, and a detector registers zero signal. While determining experimentally these positions, ellipsometric parameters are measured. It is possible to fix two of these elements, and to rotate at a constant angular velocity (for example, the analyzer) the third one. The amplitudes of the Fourier components of the signal taken from the photo detector also allow us to calculate the ellipsometric parameters  $\Psi$  and  $\Delta$ , which characterize relative change of the amplitudes for p- and s- polarizations and phase displacement.

The parameters of the polarization ellipse, i.e. the orientation of its axes and eccentricity, are determined by the optical properties of the reflective structure and the angle of light incidence. The ratio of complex reflection coefficients for the two types of light-wave polarization is measured in the experiment: in the incident plane (p) and perpendicular to (s). This ratio is usually expressed by means of ellipsometric parameters  $\Psi$  and  $\Delta$ , which characterize the relative change of the amplitudes for p- and s-polarization between them:

$$\tan \Psi \cdot e^{i\Delta} = \frac{R_p}{R_s} \quad (4.68)$$

The ellipsometric measurements are more informative than photometric, as two values are measured at the same time: the amplitude and the phase parameters  $\Psi$  -  $\Delta$ . Therefore, the equation (1) enables us to determine any two parameters of the model describing reflection coefficients  $R_p$  and  $R_s$ . The latter depends on the optical properties of the structure under investigation, as well as on the angle of incidence and wavelength[1-2].

Fig. 4.81 shows the ellipse explaining the principle of method.

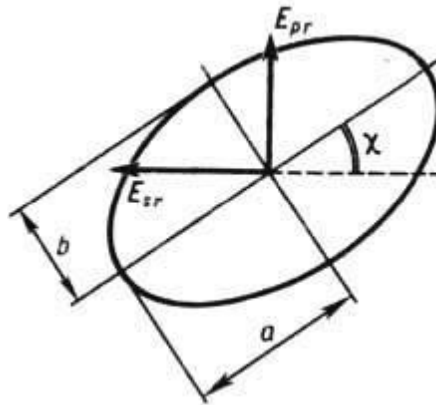


Figure 4.81. The ellipse

If the linearly polarized plane wave with electric vector  $E$  falls on a flat surface of the medium to be examined ( $E_p$  component lies in the plane of incidence and perpendicular to it -  $E_s$ ), the mirror image arises  $E_r$  plane wave with the components (expression 4.69)

$$\begin{aligned} E_{pr} &= R_p E_p = r_p e^{i\delta_p}, & E_p E_{pr} &= R_p E_p = r_p e^{i\delta_p} E_p \\ E_{sr} &= R_s E_s = r_s e^{i\delta_s}, & E_p E_{sr} &= R_s E_s = r_s e^{i\delta_s} E_p \\ \frac{E_p}{E_s} &= tg\Psi \cdot e^{i\Delta} \\ \Delta &= \delta_p - \delta_s, \end{aligned} \quad (4.69)$$

The resulting vector  $E_r$  will describe the ellipse.

$$\gamma = \pm \arctg\left(\frac{b}{a}\right), \quad (4.70)$$

where  $a$ ,  $b$  are semi axes of the ellipse. These parameters can be linked by means of the following formulae (expression 4.71)):

$$\begin{aligned} E_{sr}^2 + E_{pr}^2 &= a^2 + b^2 ; \\ E_{sr}^2 - E_{pr}^2 &= (a^2 - b^2) \cos 2\Psi ; \\ \pm E_{sr} E_{pr} \sin \Delta &= ab ; \\ 2E_{sr} E_{pr} \cos \Delta &= (a^2 - b^2) \sin 2\Psi \end{aligned} \quad (4.71)$$

$R_p$  and  $R_s$  values are determined by the incidence angle  $f$  and the optical properties (refractive index  $n$  and the absorption index  $\Psi$ ). Thus, while measuring  $a$  and  $b$  and knowing  $f$ , it is possible to define the optical properties. ( "+" sign in (2) and (3) corresponds to the left rotation, when viewed against the beam, the sign "-" corresponds to the right one)

It is possible to divide the light beam reflected from the sample into separate polarized components and measure the intensity of each of them. Then the azimuthal elements rotation is not required at all, and fast response can be achieved. This is a static photometric measurement circuit. Another class of devices is fast laser ellipsometers. These devices are used not only in research but also in high-tech manufacturing as a control tool.

To carry out local surface measurements and microscopic studies special equipment of the probe spot size of several micrometers is manufactured.

#### 4.5.3 The methods of measurements using the laser ellipsometer

Let's consider the example of measuring the optical coatings performance using the ellipsometer LEF-3M-1 which has the ability to collect ellipsometric data at different angles of incidence and length of continuous emission wavelength  $\lambda = 6328$  (He-Ne laser).

The ellipsometer LEF-3M-1 [1] is a polarizing goniometer, on movable arms of which polarizing elements are arranged: a polarizer, an analyzer and a compensator, and a light source and a photodetector as well.

The limb of the goniometer provides the installation of the given incidence angle. The polarizing elements are placed in their own limbs, and can rotate freely. The light source should provide the reception of the working beam with a high degree of monochromaticity and narrow radiation pattern in order to consider a wavefront flat.

The prisms of Iceland spar type Glan-Foucault or Glan-Thompson are used as a polarizer and an analyzer (Fig. 4.82.). Zero reference azimuth of these elements corresponds to the plane with the greatest degree of polarization.

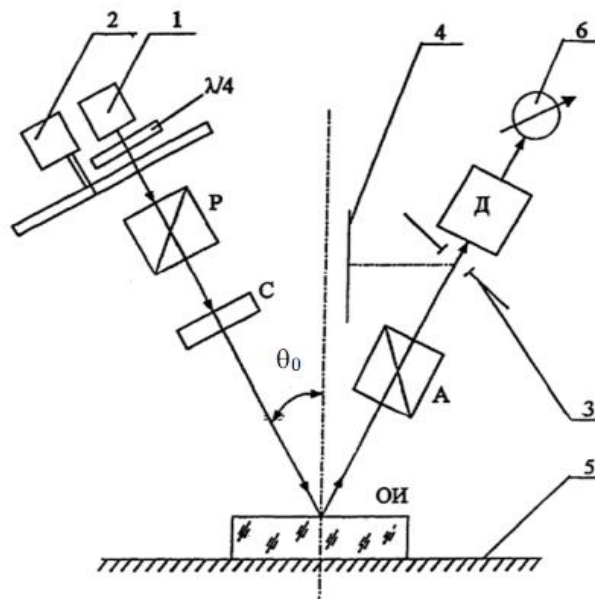


Figure 4.82. The diagram of the ellipsometer LEF-3M-1. (1 is a radiation source; 2 is a modulator; 3 is an aperture mirror; 4 is a screen; 5 is a stage; 6 is a logger)

The compensator is a plate of birefringent crystal, cut and matched in thickness so that the phase difference between the extraordinary and ordinary rays is  $90^\circ$ . Zero reference azimuth compensator corresponds to the plane along which the light propagates with the highest rate, i.e. the so-called fast axis compensator. Since crystalline compensator transmittance values along the fast and slow axes do not coincide, the real compensator is characterized by at least two parameters: the phase transmittance  $\delta_c$  and coefficients ratio along the fast and slow axes  $p_c$ .

While doing the research the sample was placed so that the axis of rotation of the arms passed through the reflective surface and the plane of incidence was perpendicular to it. The incidence angle is measured from the perpendicular to the reflecting surface.



Then the conducting of ellipsometric measurements is the following: the elliptical polarization is achieved by the rotating polarizer at the output of the compensator (i.e. in the light beam incident on the sample), so the light becomes linearly polarized after reflection. In this case, it can be put out by appropriate rotation of the analyzer, and using the photodetector for observing the zero (or minimum) intensity of the radiation. The linearly polarized light is delivered on the reflecting surface with the orientation of the polarization plane at which the light after reflection becomes elliptically polarized enough to have a linear polarization after passing the compensator. Hence, in this case, reflected radiation is put out by means of the rotating analyzer. For this reason, the measurement scheme described above is called a zero ellipsometry scheme. Its main advantage lies in the fact that with the help of the photodetector it is possible to determine the analyzer and polarizer positions, resulting in minimum light intensity at the output of the device. This intensity value is of no interest in itself. According to the comparative analysis, the zero method has the highest metrological characteristics. The highest accuracy of the method is provided by the compensator when the fast axis orientation is  $C_0 = \pm 45^\circ$ .

Let us write the azimuths of the polarizer and analyzer in an informative position (position damping) as P and A. The re-position damping is achieved with the help of the polarizer or analyzer rotation. All possible positions of polarization elements, leading to the light extinction at the ellipsometer exit, form ellipsometric zones. If such positions are broken into angular intervals and the azimuths of polarization elements fall there, 8 zones are formed, each of which consists of four subbands.

The most important findings of this examination are the following: if P and A are measured in any subband, the polarization state of the reflected light can be unambiguously determined only if the compensator parameters  $\delta_c$  and  $p_c$  are known. The nonmonotonic character of the  $\delta_c$  and  $p_c$  measure indicates the existence of multiple causes of instability: in addition to its own temperature changes as a result of the warming up of the device, there might be a slight compensator twist due to the differential heating of the ellipsometer arms.

Considering the stated amendments for the optical characteristics of the films a multi-zone method for measuring the azimuth of polarization elements P and A is used at each of the angles  $\varphi_1 = 70^\circ$  and  $\varphi_2 = 73^\circ$  which are close to the value of Brewster's angle for silicon.

#### 4.5.4 Methods of measurement using the spectral ellipsometer

Let's study the technique of ellipsometry measurements by the example of a multichannel spectral ellipsometer ESM-512.

The spectral ellipsometer ESM-512 with the binary modulation of the polarization state (BMSP) is used to determine the optical material parameters with the thickness of the optically transparent and translucent layers in multilayer structures, and the quality of processing and the surface roughness. The original polarizing devices and rapid measurement algorithms are used in the

instrument. The absence of moving polarizing elements in the spectral ellipsometer ESM-512 allows improving the accuracy and precision of the measured parameters.

The spectral ellipsometer ESM-512 consists of a power supply, optomechanical part of the ellipsometer with a built-in 8 LEDs source of light, small-sized diffraction monochromator, the line of photodiodes, and the system of controlling, interfacing and registering ellipsometric parameters with the software kit. The operating principle of the spectral ellipsometer ESM-512 is based on the switching of the polarization state, wherein the test sample is struck alternately by the beam of white light which is converted into a sequence of orthogonally polarized beams with azimuth of linear polarization P and P + 90 degrees. At each polarization state the light reflected from the sample is divided alternately into a series of orthogonally polarized beams with azimuth of linear polarization of A and A + 90 degrees, then falls on the monochromator forming the spectrum on the line 512 photodiodes, which is analyzed by the electronic registration system of ellipsometric parameters.

Fig. 4.83. shows the external view of the device.

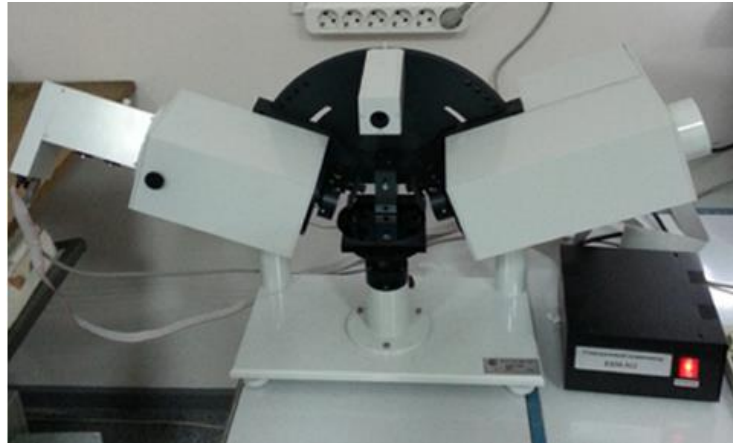


Figure 4.83. The external view of the spectral ellipsometer ESM-512 together with its power source

The ellipsometry provides the measurement of the relative phase shift of the two orthogonally polarized components and their relative change in the beam and the sample interaction. For the ellipsometry with such configuration as polarizer-sample-analyzer the light intensity on the photodetector is described by the formula:

$$I = I_0 \cdot (\sin^2 A \cdot \sin^2 P + \cos^2 A \cdot \cos^2 P \cdot \tan^2 \Psi + 0.5 \cdot \sin 2A \cdot \sin 2P \cdot \cos \Delta \cdot \tan \Psi) \quad (4.72)$$

where P and A are the polarizer and the analyzer azimuths;  $I_0$  is the coefficient that is independent of P and A;  $\Psi$  and  $\Delta$  are the ellipsometric angles determining the ratio of the complex amplitude reflection coefficients  $R_p$  and  $R_s$  for the p and s polarizations (formula 1).

In this model of the spectral ellipsometer we use the original method of polarization and optical measurements with switching of the polarization state, wherein the test sample is struck alternately

by the radiation with two orthogonal states of polarization with azimuths  $P$  and  $P + 90^0$  and the photodetectors signals are analyzed for the azimuths of the analyzer  $A$  and  $A + 90^0$ :

$$I_1 = h_a \cdot I_0 (\sin^2 A \sin^2 P + \cos^2 A \cos^2 P \tan^2 \Psi + 0.5 \sin 2A \cdot \sin 2P \cos \Delta \cdot \tan \Psi) \quad (A, P)$$

$$I_2 = I_0 (\cos^2 A \sin^2 P + \sin^2 A \cos^2 P \tan^2 \Psi - 0.5 \sin 2A \cdot \sin 2P \cos \Delta \cdot \tan \Psi) \quad (A+90, P)$$

$$I_3 = h_a \cdot I_0 (\sin^2 A \cos^2 P + \cos^2 A \sin^2 P \tan^2 \Psi - 0.5 \sin 2A \cdot \sin 2P \cos \Delta \cdot \tan \Psi) \quad (A, P+90)$$

$$I_4 = I_0 (\cos^2 A \cos^2 P + \sin^2 A \sin^2 P \tan^2 \Psi + 0.5 \sin 2A \cdot \sin 2P \cos \Delta \cdot \tan \Psi) \quad (A+90, P+90), \quad (4.73)$$

$h_a$  is a coefficient characterizing the sensitivity ratio of two channels in the analyzer.

For each azimuth  $P$  and  $P + 90^0$  the ratio of the photo detectors signals at analyzer azimuths  $A$  and  $A + 90^0$  are measured. Using these ratios  $b_1 = I_1 / I_2 / h_a$  and  $b_2 = I_3 / I_4 / h_a$  the ellipsometric parameters of the ratios  $\Psi$  and  $\Delta$  are defined:

$$\tan^2 \Psi = (x_1 - b_1 b_2 x_2 + c) / (b_1 b_2 x_1 - x_2 + c) \quad (4.74)$$

$$\cos \Delta = [b_1 x_3 - \sin^2 A \sin^2 P + (b_1 x_4 - \cos^2 A \cos^2 P) \tan^2 \Psi] / 0.5 (b_1 + 1) \sin 2A \sin 2P \tan \Psi, \quad (4.75)$$

here  $c = b_2 (\sin^2 A \sin^2 P - \cos^2 A \cos^2 P) + b_1 (x_4 - x_3)$

$x_1 = \sin^2 A$ ,  $x_2 = \cos^2 A$ ,  $x_3 = \cos^2 A \sin^2 P$ ,  $x_4 = \sin^2 A \cos^2 P$

For the chosen sample model considering the measured values  $\Delta$  and  $\Psi$  the layers parameters can be calculated based on the known ellipsometric formulae [1,2].

Before performing the measurements the ellipsometer requires testing and calibration. The checks are carried out if there is a suspicion of measurement inaccuracies or measurements are not possible.

For this, it is suggested:

- Set the angle of incidence =  $90^0$ , disconnect the compensator, remove micro add-on device.
- Put the test mirror in the analyzer under the beam and make sure that the beam hit the center of the circle on the ground glass in the test window.
- Remove the test mirror. In the group "Testing" click "Read" to display the spectrum of the LED intensities for the configured device. Set in the selection field "LD" the record "All" (all LEDs), in the field "t,ms" the value of 1000, tick the "Auto" check box. Press the "Measuring" button to display the spectrum of the LED intensities. The second chart shows what the ADC measures, the first one shows current in the photodiodes [Pa], calculated by dividing the integration time on each LED. If the measured spectrum is not significantly different from the read one in the intensity and position of the maxima, we can assume that the position of the LEDs and the beam is tuned.
- Put the compensator under the beam and check the spectrum matching with and without the compensator (click the "Measuring" button in the group "Testing"). The compensator is installed automatically with a tick in the button "Compensator".

- Put the micro add-on device under the beam and check the matching of the rays with and without the add-on device on the test mirror and the maxima positions of the spectrum as well (click "Measuring" in the group "Testing").
- Set the arms of the polarizer and the analyzer at an angle of 70 degrees.
- Set the plane-parallel (not curved) sample on the table.
- Check that the surface of the sample should be 2 mm below the base plane of the semicircle.
- Set the crosses matching in the autocollimator with the help of the screws of the table tilt. Check the position of the beam on the ground glass. If there is a match, check the position of the maxima in the spectrum (click "Measuring" button in the "Testing" group). If there is a discrepancy of more than 3 nm or mismatch on the ground glass, you need to set the beam on the ground glass of the analyzer with the help of the screws of the tilt and height of the stage in the same position as viewing against the light. The height of the stage is recommended to be installed with the running micro add-on device, and the tilts – without micro add-on device. Then achieve a good maxima match in the spectrum by means of a more fine adjustment. If there is no matching of the lines in the autocollimator, it is necessary to adjust the autocollimator to achieve the lines matching with the help of four screws at the bottom of the autocollimator. Next, after the change of the sample the tilt adjustment is performed only with the help of the autocollimator.

If you can not make the adjustment, you should contact the manufacturer for advice or call for the representatives to make the adjustment.

The parameters that are to be calibrated:

- the ratio range of the channels sensitivity corresponding to the two polarizations after the analyzer.
- the angles of the polarizer and analyzer with respect to the plane of incidence.
- the angle of incidence.
- the  $\tan \psi$ ,  $\Delta$  spectra of the compensator.

Full calibration is carried out by the manufacturer. The user can optionally perform the following calibration:

The entire range of spectral wavelengths is closed off by 8 LEDs, each of which has its own range of wavelengths. The LEDs are installed on the disk, which is rotated by a stepper motor. When you run the program, the disk is set to the starting position determined by the limit switch and it turns to set the white LED.

The number of LEDs for measurement is selected automatically from the posted wavelength band. After the measurement the white LED is set.

The signal at the photodiodes illumination forms because of the charge integration and is proportional to the time integration. When measuring the integration time is automatically selected so that the maximum signal for each LED was at approximately 0.7 the range of the ADC, but not more than  $\frac{1}{4}$  of the time set as Max.time (4 beams are). Therefore, the more Max.time is up to a

certain point, the more precise are the measurements, but the possible measurement time is more as well. In most cases, the sufficient accuracy is achieved when Max.time = 1 second, which corresponds to approximately 10 seconds while measuring the full range.

Table 4.4. The parameters of LED sources

LED number	Wavelength of the maximum	Wavelength band
0	355	348-362
1	370	362-377
2	385	377-393
3	400	393-408
4	420	408-435
5	white	435-690
6	740	690-758
7	800	758-810

After measuring the spectrum intensity, the background illumination (backlighting) is measured and subtracted from the measured signals. Then the ratio of rays and the required parameters is calculated. While measuring the parameters are displayed for each LED, but they can be obtained considering the background of the previous measurement, or using the instruction <Background meas.> or a zero value for the first measurement. Therefore, the intermediate results can be inaccurate (they are displayed for the information that measurements are normal). When measured at a wavelength depending on the time it is necessary to measure the background signal using the instruction <Background meas.> before the start of the measurements.

It is recommended that the ambient light contact with the inlet of the analyzer should be avoided while measuring and calibrating. Although some small permanent exposure does not affect the measurement, but the exposure that changes while being measured leads to additional errors. The exposure change occurs when people move around the ellipsometer or when there are some changes of the sunlight passing through the window.

Measurement preparation:

1. Set the angle of incidence, if you want to change it. In most cases, the angle should be 70 degrees.
2. Install the sample so that the beam is in the right place of the sample and align the tilts of the stage according to the autocollimator.

3. If the thickness of the sample changes, adjust the height of the stage according to the ingress of the beam in the desired location of the ground glass in the analyzer (you can use micro add-on device for higher accuracy) and then align the tilts of the stage according to the autocollimator.

4. Remove the alignment mirror in the analyzer.

Measuring procedure:

1. Set the required parameters in the group "Acquisition".
2. Select the angle of incidence in the "Setup" window so that it corresponds to the fixed one.
3. Press <Setup F> in the sample model for a call to the corresponding calibrated angle of incidence or set the desired value in the editable field without selecting "Setup" in the window. The angles values A, P are set automatically from the "Setup" window. Make sure you never adjust the angles A, P in a manual way in the "Setup" window, they should be defined only as a result of the calibration.
4. If the measurements are made with a compensator, an approximate sample model must be entered, otherwise the calculation  $\cos \delta$  of the sample may be incorrect (see para. "The use of the compensator.").
5. Select the "Smooth region (+-pixels)" parameter in the "Setup" window : it is possible to choose approximately 10 for thin films (if there are no sharp peaks in the spectrum), otherwise – approximately 3-5.
6. Press the button <Start(L)> to start the measurement.

Set the values "WaveL." and "Period" in the group "Acquisition". When measuring there is averaging over the number of pixels defined as the maximum of the parameters "Smooth region (+-pixels)" and "Average (pixel)". Before the measurement, it is recommended that the background illumination should be measured using the instruction <Background meas.>. Press the button <Start (t)> to start the measurement.

The compensator is installed in the beam path automatically when you select the "Compensator" check box. Phase shift of the compensator is between 75-85 degrees over the whole spectrum. The compensator is recommended to be used to measure only thin samples, where  $\delta$  is within the limits of 160 - 180 degrees or 0 - 20 degrees. However, it is not recommended to use the compensator for the samples where the  $\delta$  and the compensator in sum are more than 180 or 360 degrees in the spectrum. Since the measurement includes the calculation of both  $\cos \delta$  sample and the compensator, its adjustment for producing  $\cos \delta$  (or  $\delta$ ) of the sample can be ambiguous. To remove the ambiguity in the calculation the assumed (given) model of the sample is used. However, this is not possible near the critical points (180, 360 deg.), because of the finite accuracy of the given model (the exact model of the sample is not known beforehand). When measuring using the compensator a rough sample model is necessary to be set.

The apparatus is equipped with a micro add-on device, which is a combination of two removable units that are located at the polarizer exit and the analyzer input and have lenses designed to focus the beam. To carry out local measurements these units should be placed on the way of the beam.

Note that the absolute accuracy of measurement using the add-on device is worse than without it because the lenses have a small anisotropy, and the angle of incidence and the azimuth (P-A) may differ from the calibrated ones for the inevitable inaccuracies lenses position. A shift of 1 mm of focusing relative to the beam centre at a focal length of 45 mm results in an error of 1 degree in the angle of incidence. Micro add-on device is recommended to be used in the most difficult cases, when the sample surface is very uneven or if you want to measure on a small sample area.

See published works [4-9] for further details about the spectral ellipsometry method.

#### 4.5.5 References

1. Р. Аззам, Н. Башара. Эллипсометрия и поляризованный свет. М.: "Мир". 1981.
2. М.М.Горшков. Эллипсометрия. Москва, "Советское радио". 1974.
3. D.A.G. Bruggeman, Ann.Phys. (Leipzig) 24 (1935) 636.
4. В.И. Ковалев, А.И. Руковишников, П.И. Перов, Н.М. Россуканый, Л.А. Авдеева. Разработка оптических методов и аппаратуры для контроля технологии и параметров полупроводниковых структур нано- и микроэлектроники. *Радиотехника и электроника* 44 (1999) 1059
5. В.И. Ковалев, А.И. Руковишников. Компактный многоканальный спектроэллипсометр для ex-situ и in-situ измерений. *Приборы и техника эксперимента* (2003) № 2, 164-165.
6. Leontyev A.V., Kovalev V.I., Khomich A.V., Komarov F.F., VV Grigoryev, AS Kamishan. PMMA and polystyrene films modification under ion implantation studied by spectroscopic ellipsometry. *Proceedings SPIE* 5401 (2004) 129-136.
7. Khomich A.V., Kovalev V.I., Vedenev A.S., Kazanskii A.G., Forsh P.A., He D., Wang X.Q., Mell H. Optical and photoelectrical characterization of as-deposited and annealed PECVD polysilicon thin films. *Proceedings SPIE* 5401 (2004) 200-207.
8. F.F. Komarov, A.V. Leontyev, A.V. Khomich, V.I. Kovalev. The formation of dielectric microwaveguides at the system polymer/SiO<sub>2</sub>/Si by using of ion irradiation. *Vacuum* 78 (2005) 617-622.

## 4.6 Mössbauer spectroscopy

Mössbauer spectroscopy is the multipurpose experimental technique that is applied for investigation of local atomic ordering and magnetic properties in numerous scientific fields including basic research in physics and chemistry, materials science, biology, geophysics, etc. This technique is based on the effect of the resonant and recoil-free emission and absorption of gamma radiation by atomic nuclei bound in a solid. This effect is also called *Mössbauer Effect* after the name of the scientist – Rudolph Mossbauer, who observed and explained this effect in 1957, and was awarded the Nobel Prize in Physics in 1961 for this discovery [1, 2].

Mössbauer effect is a process in which a nucleus emits or absorbs gamma rays without loss of energy to nuclear recoil because the momentum of recoil is delivered to a surrounding crystal lattice rather than to the emitting or absorbing nucleus alone. When this occurs, no gamma energy is lost to the kinetic energy of recoiling nuclei so that emission and absorption occur at the same energy. Due to very narrow linewidths of resonant absorption energy it is very sensitive to small changes in the energies of nuclear transitions reflecting interactions between a nucleus, its electrons and electrons of neighboring atoms. This is the basis for Mössbauer spectroscopy, which combines the Mössbauer effect with the Doppler effect, namely, change in frequency (or wavelength) of a wave for studied sample that is moving relative to the wave source to detect and analyze these interactions.

Considered interactions of nucleus with neighboring atoms results in the splittings and shifts, so far drawing to the formation of hyperfine structure of Mössbauer spectra characterized with energies as small as  $10^{-5}$  -  $10^{-8}$  eV. This means that only isotopes possessing lifetime  $\tau$  of excited levels larger than  $10^{-8}$  s could be applied for Mössbauer spectroscopy. Among acceptable Mössbauer isotopes iron (Fe) is the element that is most often observed through the Mössbauer Spectroscopy technique, as its isotope  $\text{Fe}^{57}$  is abundantly present, emits a low energy  $\gamma$ -ray, and has an excited nuclear state of long duration, all of which are necessary for an observable Mössbauer spectrum. Some other elements that could be applied for Mössbauer technique are Ruthenium (Ru), Iridium (Ir), Gold (Au), Tin (Zn), Antimony (Sb), Europium (Eu), Gadolinium (Gd), Dysprosium (Dy), Erbium (Er), Ytterbium (Yb), among others.

Registration of Mössbauer spectrum includes measurement of count rate of  $\gamma$ -quanta (or their relative absorption) as a function of the velocity of the source.

### 4.6.1 Mössbauer spectroscopy instrumentation

Typical Mössbauer spectrometer consists of a radioactive source (1), sample (2), detector (3), and a drive (4) that moves the source or the absorber (see the scheme in Figure 4.84). The drive usually moves the source away from the sample, with a linear variation in velocity (between 1-100 mm/s) with respect to time in order to change the range of investigated energies as a result of the Doppler effect. Absorber contains the same type of isotope as the source. Mössbauer spectroscopy could be



realized in two main modes - transmission and back-scatter, which differs by the detector's position, with respect to the source and the sample.

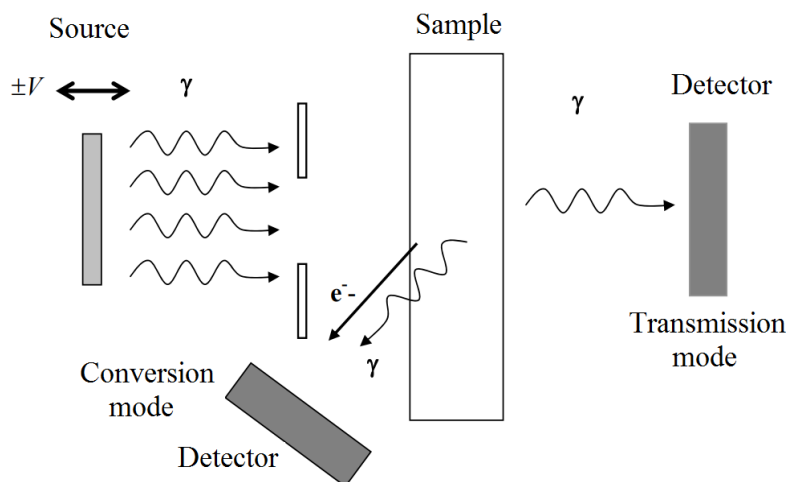


Figure 4.84. Scheme of Mössbauer spectrometer

The energy scale used for Mössbauer spectra with respect to the x-axis is most commonly in terms of the source's velocity, which is measured in mm/s. As it is already mentioned, Mössbauer spectroscopy is extremely sensitive in detection of relative variation of energy, which in case of isotope  $\text{Fe}^{57}$  is  $10^{-13}$ . That is why this technique allows to observe changes of energies of  $\gamma$ -quanta originating from the interaction of probe nuclei with magnetic and electric fields created in crystals by electrons and surrounding atoms. Such interactions results in the splitting of energy levels of nuclei which is reflected in recorded Mössbauer spectra. By analyzing of such spectra one could obtain the information on magnetic properties of studied materials (described by effective magnetic hyperfine field  $H_{\text{hf}}$ ), on lateral symmetry of wave functions of electrons (electrical quadrupole splitting  $\Delta$ ), and on charge density on nuclei and chemical bounding (isomer shift  $\delta$ ).

**Isomer shift**  $\delta$  that originates from the Coulomb electrostatic interaction of nuclei with external electron shell (generally, this is  $s$ -shell) results in the shift of energy levels and, consequently, shift of the resonant line in the spectrum (see Figure 4.85a). Systematic analysis of isomer shifts in various iron compounds and alloys reveals that values of  $\delta$  depend on the valence state of the iron ion. For example,  $\text{Fe}^{3+}$  (electronic configuration  $3d^5$ ) possesses lower values of  $\delta$  varying between 0.05 and 0.5 mm/s as compared to  $\text{Fe}^{2+}$  (electronic configuration  $3d^6$ ) varying between 0.65 and 1.45 mm/s. So far, value of  $\delta$  extracted from the concrete spectrum allows definition of valence state of iron ion and Fe site occupancy in a particular site in studied material. More generally, value of  $\delta$  characterizes type of chemical bonding, phase composition of solids and kinetics of phase transitions and chemical reactions.

**Electrical quadrupole interaction** is governed by electric interaction of quadrupole moment of nuclei with electric field gradient formed by electrons in its nearest surrounding. Such interactions

draw to the splitting of resonant lines and observation of quadrupole doublet on the spectra (see Figure 4.85b). It is established that electric field gradient is proportional to  $r^{-3}$ , where  $r$  is the radius of nuclei. Consequently, the effect on the  $\Delta$  values of atoms (or ions) closest to the nuclei as well as partially filled electron shells are most pronounced.

Experimental data prove that quadrupole splitting for  $\text{Fe}^{2+}$  ions ( $\Delta = 1.7 - 4 \text{ mm/s}$ ) are much larger than those for  $\text{Fe}^{3+}$ , where 3d – electrons with parallel orientation of spins form spherically symmetric charge distribution and relatively small electric field gradient is associated only with neighboring ions. Larger gradient for  $\text{Fe}^{2+}$  originates from uncompensated electron on the d – shell of ion that promotes large contribution into electric field gradient.

Values of quadrupole splitting directly relate to the valence state of atoms, crystalline symmetry and type of chemical bonding in solids, thus (being combined with isomer shifts) allowing precise identification of phase composition of materials.

**Magnetic Zeeman hyperfine splitting** originates from the interaction between dipole magnetic moment of nuclei  $\mu$  with hyperfine magnetic field of nuclei  $H_{\text{eff}}$ , formed by its own electrons. Magnetic dipole interactions are observed mainly in magnetically ordered materials (ferro-, antiferro- or ferrimagnets). Such interactions results in the splitting of resonant lines into several lines in accordance with possible transitions between sublevels of basic and excited energy levels of the nuclei. For example, for  $^{57}\text{Fe}$  nuclei there are 6 allowed transitions that are reflected in 6 spectral lines called magnetic sextet (see Figure 4.85c). Analysis of position of spectral lines and their intensity allows determining of the magnetic field strength  $H_{\text{eff}}$  within nuclei. Values of these fields are very sensitive to the peculiarities of electronic structure of materials and contain information about their magnetic properties.

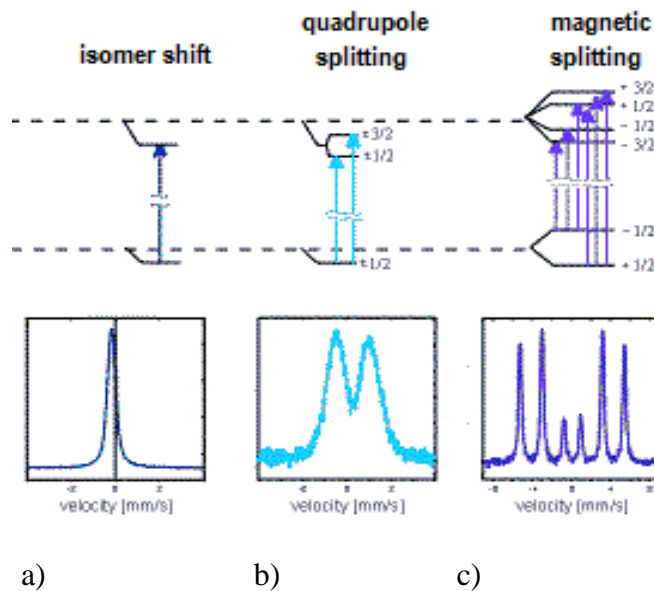


Figure 4.85. Diagrams of hyperfine splittings of the energy levels of nuclei and correspondent types of Mössbauer spectra [3]: a – isomer shift and shift of resonance line; b – quadrupole splitting and quadrupole doublet; c – Zeeman magnetic splitting and magnetic sextet.

It is worth noticing that interpretation of Mössbauer spectra is quite nontrivial task. Before fitting procedure one should propose the model for approximation of the spectra that includes set of possible subspectra that corresponds to phases or just local states of Mössbauer isotope in the particular sample. Also contribution of this or that phase as well as their magnetic characteristics should be taken into account. Such model could be proposed basing on preliminary study by other structural techniques like X-ray or electron diffraction combined with available data on morphology and magnetic properties of studied materials.

#### 4.6.2 Mössbauer spectra of nanocrystalline systems and their interpretation

There is quite limited set of experimental techniques that allows investigation of nanocrystalline systems (powders, thin films, nanostructured composites), containing nanoparticles (nanoinclusions) with diameter less than 10 nm. Mössbauer spectroscopy offers exclusive opportunity to identify phase composition of nanocrystallites by studying local atomic arrangement there. Moreover, by fitting of the spectra one could estimate the relative contribution of every particular phase or just Fe ions in particular valence state.

Below we will consider several typical examples illustrating specific features of Mössbauer spectra recorded on nanocrystalline materials and their fitting.

One of the typical features of nanocrystalline structures is effect of superparamagnetic relaxation – that is quasi-paramagnetic behavior of nanoparticles of very small size due to thermally induced random orientation of magnetic moments there. Superparamagnetic effect is vividly reflected in Mössbauer spectra at temperatures above *blocking temperature*  $T_B$  (see part *Magnetometry*). As considered above, superparamagnetic state of magnetically ordered material originates from spontaneous change of magnetic moments orientation because of thermally activated coherent rotation of all spins. This may happen under the condition when characteristic time  $\tau$  of applied experimental technique is larger than time of superparamagnetic relaxation  $\tau_{sp}$  of studied material. In accordance with Bean-Livingstone criterion, value of  $T_B$  is defined as the critical temperature above which magnetization process in agglomeration of non-interacting nanoparticles of similar value becomes typical for superparamagnetic state. In this case correlation between  $T_B$ , nanoparticles volume  $V$  and  $\tau_{sp}$  is described by the relationship:

$$T_B = \frac{K_{eff} \cdot V}{\ln\left(\frac{\tau_{sp}}{\tau_0}\right)} \cdot k_B, \quad (4.76)$$

Where  $K_{eff}$  – is effective magnetic anisotropy constant,  $\tau_0$  – constant in the range  $10^{-11}$  -  $10^{-9}$  s.

Medium sizes of nanoparticles and correspondent blocking temperatures could change noticeably for different materials.

To illustrate superparamagnetic behavior one should refer to the Fig. 4.86 that summarizes Mössbauer spectra for nanocomposite samples containing FeCo-based nanoparticles of 3-6 nm in diameter assembled inside insulating matrix of aluminum oxide. Fig. 4.86a demonstrates the change of spectra with the increase of nanoparticles concentration  $x$ . At low concentration of magnetic nanoparticles ( $x = 34\%$  and  $40\%$ ) Mössbauer spectra reveal asymmetric doublet. This doublet is associated with FeCo-based nanoparticles with medium diameter  $\langle D \rangle = 4,5$  nm that is definitely below the threshold size for transition from ferromagnetic to superparamagnetic state. That is why observation of such doublet is interpreted as the evidence of superparamagnetic relaxation in FeCo nanoparticles above  $T_B$ . Fig. 4.86a reveal that shape of Mössbauer spectra measured at room temperature changes with the increase of nanoparticles concentration  $x$  from superparamagnetic doublet towards well ferromagnetically split sextet through partially magnetically collapsed sextet at  $x = 47\%$ . Such variation of spectrum means that increasing  $x$  promotes formation of ferromagnetically interacting network of nanoparticles which even at room temperature diminishes the superparamagnetic behavior of every separate particle. Unusual shape of the spectrum observed for sample with  $x = 47\%$  illustrates the intermediate state of material combining superparamagnetic state and ferromagnetically ordered state of some particles depending on their size. This means that in sample containing nanoparticles with wide size dispersion some of nanoparticles already agglomerated into the network, while some part still remains in isolated state thus demonstrating superparamagnetic relaxation at room temperature.

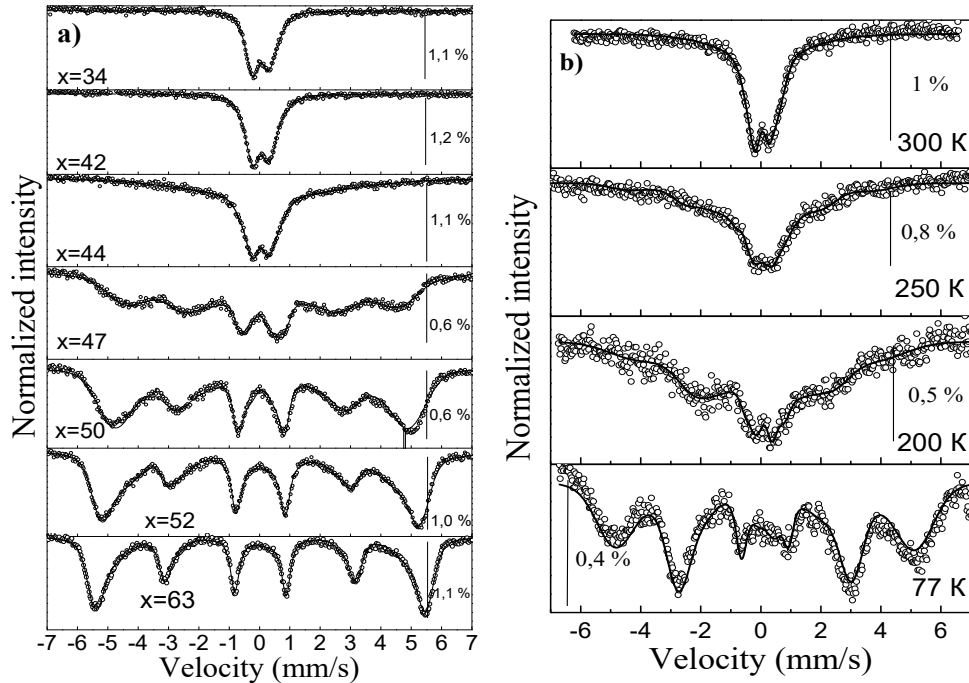


Figure 4.86. Experimental (points) and approximated (curves) Mössbauer spectra of  $(\text{FeCo})_x(\text{Al}_2\text{O}_3)_{100-x}$  ( $34\% \leq x \leq 63\%$ ) as a function of FeCo contribution  $x$  (a) and of measurement temperature  $T$  (b).

In order to propose proper model for fitting of Mössbauer spectra containing nonmagnetic contribution (single line or doublet) one must be sure it could be associated with superparamagnetic state of the sample but not with just paramagnetic phase. To verify that it is necessary to perform set of measurements with decreasing temperature. In case of superparamagnetic relaxation, the spectrum will transform into the sextet at measurement temperature below  $T_B$ . Such transformation is well illustrated in Fig. 4.86b that summarizes the spectra measured at 300 K, 250 K, 200 K и 77 K. It is seen that at  $T = 77$  K spectrum contains well splitted sextet supporting ferromagnetic state of FeCo-based nanoparticles as expected.

Temperature evolution of Mössbauer spectra allows estimation of  $T_B$  value for particular sample that is defined as the temperature when nonmagnetic and magnetically split contributions into the spectrum become equal. Basing on temperature dependencies of both contributions plotted in Fig. 4.87., value of  $T_B$  is estimated approximately as 126 K.

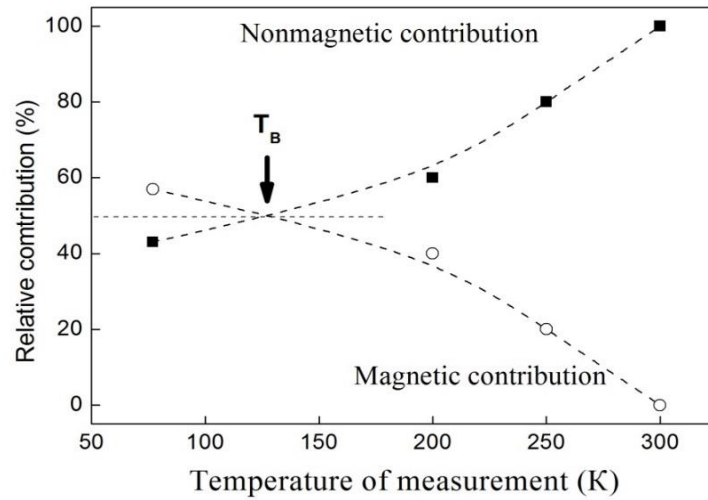


Figure 4.87. Estimation of the blocking temperature  $T_B$  from the temperature dependencies of magnetically split (magnetic) and nonmagnetic contributions Mössbauer spectra of  $(\text{FeCo})_{40}(\text{Al}_2\text{O}_3)_{60}$  sample.

Medium size of nanoparticles could be estimated as well from the temperature dependencies of Mössbauer spectra. Hyperfine magnetic splitting in the spectra becomes apparent when time of superparamagnetic relaxation  $\tau_{sp}$  is equal to characteristic time  $\tau$  for Mössbauer spectroscopy. Characteristic time  $\tau$  is proportional to the frequency of Larmor precession of the spin of nucleus  $\nu_L$  defined as:

$$\nu_L = \frac{\mu_I}{2\pi\hbar I} H_{eff}, \quad (4.77)$$

where  $\mu_L = 0.153 \cdot 10^{-27}$  J/T is magnetic moment of  $^{57}\text{Fe}$  nucleus in excited state,  $I = 3/2$  – spin of this nucleus,  $H_{eff} = 33.3$  T. Value of  $\nu_L$  calculated basing on this formula is  $2.6 \cdot 10^7 \text{ s}^{-1}$ , that corresponds to characteristic time  $\tau = 3.8 \cdot 10^{-8} \text{ s}$ .

Next, value of  $\tau_0$  could be estimated from the following expression:

$$f_0 = 1/\tau_0 = (K_{\text{eff}} \cdot A)/(\rho \cdot N_A \cdot h), \quad (4.78)$$

where  $A = 61$  g/mol – calculated molar mass of FeCo-based nanoparticles,  $\rho = 8.3 \cdot 10^3$  kg/m<sup>3</sup> – density of nanoparticles,  $N_A$  – Avogadro number. Magnetic anisotropy constant  $K_{\text{eff}} = 10^6$  J/m<sup>3</sup> as defined by nanosized Fe granules. Calculated value of  $\tau_0 = 5 \cdot 10^{-11}$  s.

Finally, from equation (4.77) the following expression is derived:

$$K_{\text{eff}} \cdot V_{kr}/k_B \cdot T_B = \ln(\tau_{sp}/\tau_0) = \ln(3.8 \cdot 10^{-8}/5 \cdot 10^{-11}) = 6.6, \quad (4.79)$$

where  $V_{kr}$  - critical volume of nanoparticles for the onset of magnetic hyperfine splitting (when  $\tau_{sp} = \tau$ ),  $T_B$  – blocking temperature estimated from temperature dependency of nonmagnetic and magnetically split contributions shown in Fig. 4.87.

Substitution of  $T_B \approx 126$  K into the formula (4.68) allows calculation of  $V_{kr}$  value for nanoparticles which are in nonmagnetic state at room temperature (either superparamagnetic or weakly interacting). Medium diameter of nanoparticles estimated from  $V_{kr}$  equals to 3 nm. This value is very close to the values directly observed by transmission electron microscopy thus confirming applicability of considered approach for estimation of nanoparticles size from Mössbauer spectra.

In conclusion, Mössbauer spectroscopy is very informative technique providing complex qualitative and quantitative information about local atomic structure and magnetic state of materials, containing correspondent isotope. Phase composition and magnetic state are identified basing on the analysis of hyperfine parameters (isomer shift, quadrupole splitting, magnetic hyperfine splitting) extracted from experimental Mössbauer spectra by their fitting.

#### 4.6.3 References

1. R.L. Mössbauer, Z. Physik, 1958, 151, 124.
2. R.L. Mössbauer, Naturwissenschaften, 1958, 45, 538.

<http://iacgu32.chemie.uni-mainz.de/moessbauer.php>.

## 4.7 Scanning electron microscopy

### 4.7.1 Outlook

In this chapter, we will consider analytical methods based on probing of objects under study by a high-energy (about  $10^4$  eV) electron beam. In this case, along with reflection of the incident electrons from the surface, other interesting and useful effects take place: emission of secondary electrons, creation of vacancies in the internal electron shells of atoms, which results in emission of Auger electrons and high-energy photons. All these effects can be used for analysis of the surface morphology and elemental composition of the investigated objects.

Electron beam probing is of an exceptional interest for the study of semiconductor materials and devices. In this case, incident electron beam leads to generation of multiple electron-hole pairs. Investigation of spectral composition of the radiation arising due to their radiative recombination is the basis of cathodoluminescence analysis. If a barrier structure (*p-n* or heterojunction, Schottky contact) is formed on the basis of the studied semiconductor, the separated electrons and holes will provide a current in the external circuit, which magnitude is determined by the recombination rate of the generated charges. Since the recombination rate is strongly sensitive to presence of structural defects in material, analysis of the current distribution induced by the electron beam is an extremely valuable method for investigation of defects and their recombination activity.

### 4.7.2 Physical basis of scanning electron microscopy

If electron beam is used as a probe, information about an object of interest can be obtained due to the following effects:

1. Emission of secondary electrons (surface morphology mode).
2. Elastic electron scattering (phase contrast mode).
3. Diffraction of reflected (backscattered) electrons.
4. Emission of characteristic X-ray radiation.
5. Cathodoluminescence.
6. Excitation of Auger electrons.

For the experimental realization of the above-mentioned methods, one use devices called as scanning electron microscopes (SEM). Schematic diagram of the scanning electron microscope (Figure 1) includes an electron source, a system for electron focusing and scanning (an electronic column), a sample chamber, detectors for signal recording, and a pumping system necessary to create a vacuum in a microscope chamber.

The interaction of the scanning electron probe with a local inhomogeneity and its environment leads to a difference in the corresponding measured signals by a value, which can be recorded. As

a result, during scanning the surface with an electron beam, a contrast is formed. The resolution of the scanning electron microscope is determined:

- by the size of the electron probe. In modern microscopes, the "diameter" of the electronic probe is a few nanometers;
- by sensitivity of the equipment, which records the change in response.
- by the size of signal generation area.

All modern electron microscopes allow the formation of contrast both using secondary and elastically backscattered electrons. The possibility of separate detection of secondary and elastically backscattered electrons is related to difference in their energies: secondary electrons have an energy about 20 eV, whereas backscattered ones – about 20 keV.

The number of electrons entering the detector is determined by the angle between the incident electron beam and the normal to the surface of the sample. This is true both for backscattered and secondary electrons. Changes in the surface slope about  $1^\circ$  are sufficient for registration.

There is a direct analogy between the mechanism of contrast formation in SEM and regularities of perception of the optical image by the human eye. In both cases, the brightness of the image region is determined mainly by its slope with respect to the incident light or electron beam. This explains the ease, with which the eye almost automatically perceives the SEM image as a three-dimensional image of the object due to changes of brightness.

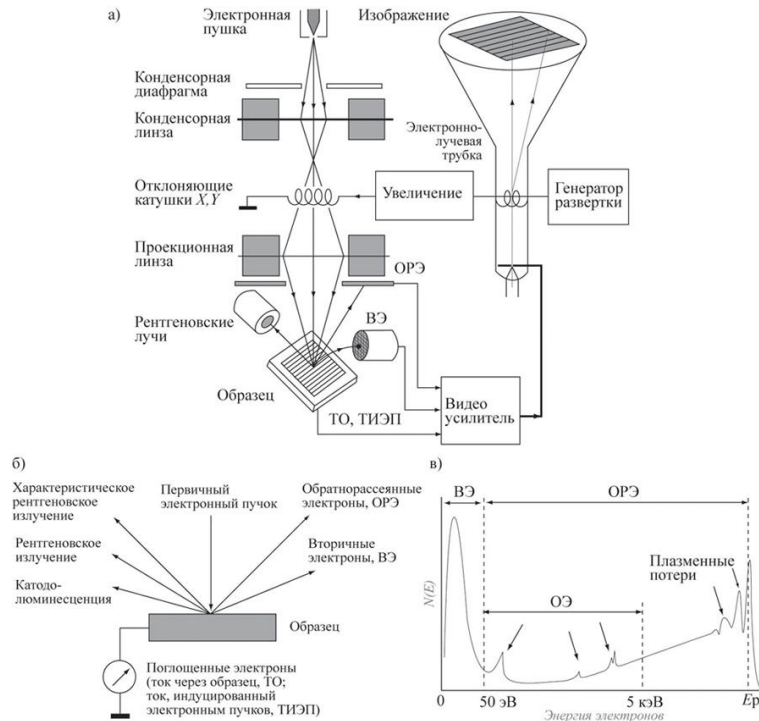


Figure 4.88. Scheme of scanning electron microscope (a); schematic representation of the processes occurring under impact of electron beam (b); energy spectrum of detected electrons (c).



In the case of backscattered electrons, there is another mechanism of contrast formation. The number of backscattered electrons increases with increasing atomic number of the element bombarded by the primary electrons. Therefore, for a heterogeneous distribution of atoms of different sorts over the surface, the regions enriched with heavy atoms appear lighter, whereas the regions enriched with light atoms are darker.

Dependence of the formed contrast on the operation mode in SEM is illustrated in Fig. 4.88. In this example, the object under study was a ceramics of strontium cobaltite  $\text{SrCoO}_x$ . As is seen from Fig. 4.88a, where the contrast is obtained in the mode of detection of secondary electrons, the sample has a polycrystalline structure with a grain size of several tens of micrometers. One can see that both in grains and along the grain boundaries, there are round-shape inclusions of a much smaller size. Image presented in Fig. 4.88a does not allow determining, whether these inclusions belong to the same phase as the matrix, or else represents another phase. At the same time, as can be seen from Fig. 4.88b, in the backscattered electron detection mode, the inclusions appear much more dark indicating that they have differ composition as compared to the matrix.

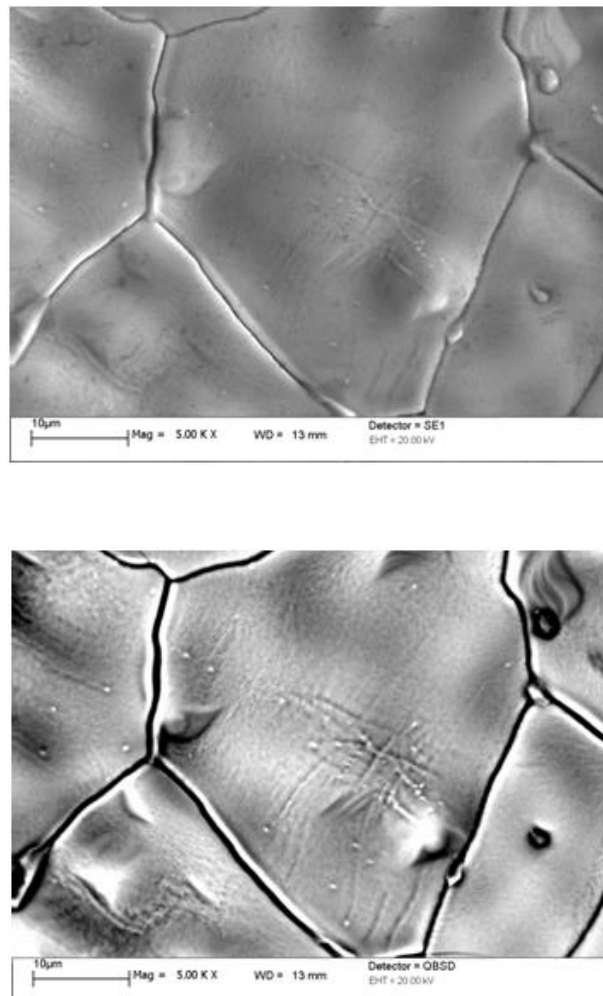


Figure 4.89. SEM images of strontium cobaltite ( $\text{SrCoO}_x$ ) ceramics [1]

Top image: detection of secondary electrons; bottom image: detection of backscattered electrons.

#### 4.7.3 Analysis of the elemental composition using electron probe

Irradiation with high-energy electrons results in formation of vacancies in the inner electron shells of atoms. In this case, the excited atoms can release their energy with the emission of either X-ray photons or electrons. The energy of these emitted particles is uniquely determined by the position of the energy levels of electrons in atoms, which is individual for each chemical element making it possible to determine the chemical composition of the objects under study.

The method of the elemental composition analysis based on determination of the energy of the characteristic X-ray radiation quanta was called the energy-dispersive X-ray analysis (EDX).

Let us consider the generation of X-rays by an electron beam. After the appearance of a vacancy in the inner shell of an atom, excitation can be relaxed due to transition of an electron from a higher-energy shell with emission of an X-ray quantum. The set of possible radiative transitions and, consequently, the energy spectrum, are determined by the following selection rules:

$$\begin{aligned}\Delta l &= \pm 1, \\ \Delta j &= 0, \pm 1\end{aligned}\tag{4.80}$$

where  $l$  is the orbital quantum number ( $l=0, 1 \dots n-1$ ),  $j = l \pm s$  ( $s = \frac{1}{2}$  is the spin quantum number).

Detectability of the EDX method is determined mainly by the presence of X-ray bremsstrahlung, which occurs when the incident electrons slow down in the analyzed object and is characterized by a continuous spectrum. As an example, Figure 3 shows the EDX spectrum of a bismuth oxosulfide film. Along with the peaks corresponding to bismuth, sulfur and oxygen, there is a continuous background of bremsstrahlung. In practice, the EDX method enables one to register concentrations of elements of the order of one tenth of atomic percent.

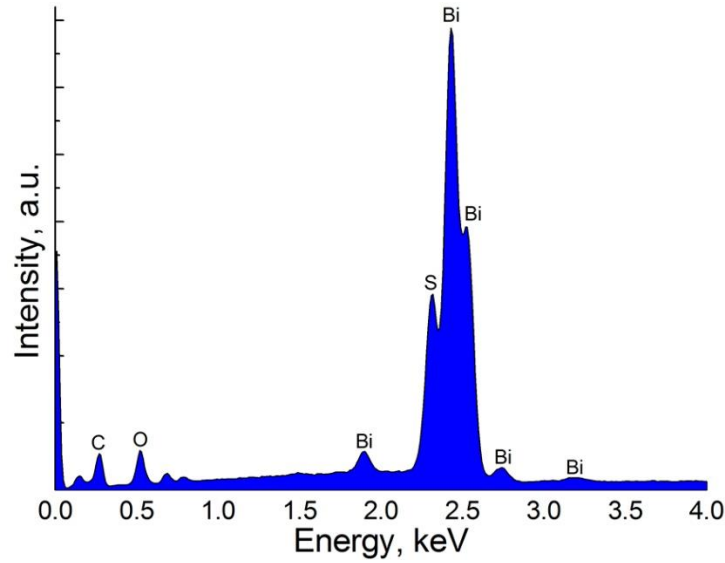


Figure 4.89. EDX spectrum of bismuth oxysulfide powder [2]

To estimate the spatial resolution of methods of the elemental composition analysis based on the interaction of electron beams with solids, let's estimate the stopping range of electrons in them.

The stopping range of electron can be represented as:

$$R = \int_{E_0}^0 \frac{dE}{dE/dx} . \quad (4.81)$$

Here energy losses  $dE/dx$  are determined by the following equation:

$$\frac{dE}{dx} \approx \frac{NZ_2}{E} \ln \left( \frac{E}{I} \right), \quad (4.82)$$

where  $N$  is the concentration of atoms in the target,

$Z_2$  is the atomic number of the target,

$I$  (eV)  $\approx 10Z_2$  is the ionization energy of the target atoms.

One use usually the following empirical relation:

$$R = \frac{K}{\rho} E_0^\gamma, \quad (4.83)$$

where  $K$  is the substance independent constant,

$\rho$  is the substance density,

$1.2 < \gamma < 1.8$ .

It follows from (4) that mass range of electrons  $\rho R$  does not depend on the substance, but is determined only by their energy.

The mass range of electrons relative to X-rays generation  $\rho R_x$  is less than full range  $\rho R$ , since a nonzero electron energy is necessary for generation of X-rays.

There is an empirical equation:

$$\rho R_x = 0.064(E_0^{1.68} - E_B^{1.68}), \quad (4.84)$$

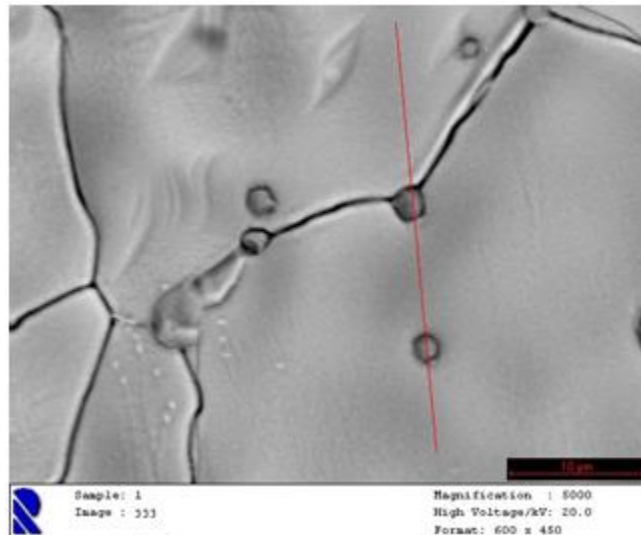
where  $R_x$  is in micrometers,

$\rho$  is in  $\text{g/cm}^3$ ,

$E_0, E_B$  are in keV ( $E_B$  is the energy required for generation of X-ray radiation, or electron binding energy).

Thus, the region of X-ray generation is of the order of a micrometer (a few hundreds of nanometers for targets composed by heavy elements). Since the value reciprocal to the X-ray absorption coefficient is much larger, the size of the region, which forms the characteristic X-ray radiation, is determined by the size of its generation area.

As an illustration of application of the EDX method, let's return to the example that has been shown in Figure 4.89. Analysis of contrast in the mode of detection of elastically reflected electrons showed the presence of regions with a presumably different atomic composition in the strontium cobaltite ceramics. To determine the quantitative characteristics of atomic composition, a linear scanning of the studied sample by an electron beam along the line shown in Figure 4.90a with simultaneous registration of the characteristic X-ray emission spectrum has been done. As can be seen from the Figure 4.90b, small inclusions of the round form are essentially enriched with cobalt, which has a smaller atomic number ( $Z_{\text{Co}}=27$ ) as compared to strontium ( $Z_{\text{Sr}}=38$ ).



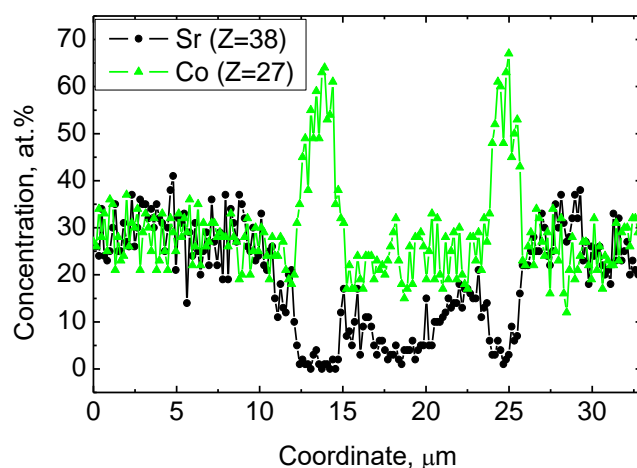


Figure 4.90. Distribution of concentrations of Sr and Co in strontium cobaltite ceramics determined by EDX with linear scanning [1]

#### 4.7.4 Auger-electron spectroscopy

As it was mentioned above, irradiation with fast electrons can lead to the formation of vacancies on the inner electron shells of atoms. In this case, the excited atoms can release their energy with the emission of either photons or electrons. The latter process represents the physical basis of the Auger electron spectroscopy method, in which the atomic composition of material is determined using measurement of the energy spectrum of the electrons emitted during the irradiation by fast electrons. This process is schematically shown in Fig. 4.91.

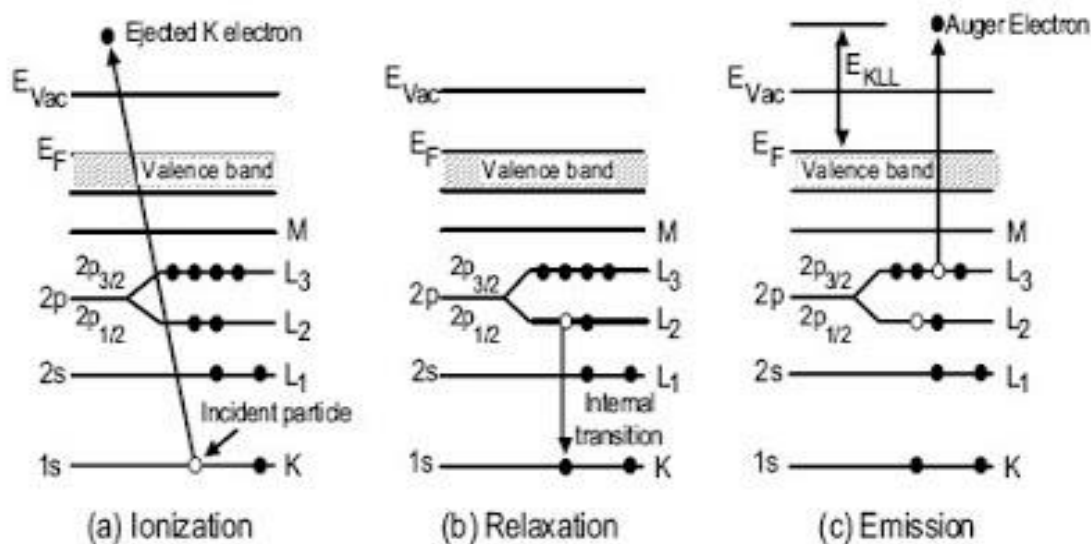


Figure 4.91. Sequence of electronic transitions in the Auger process:

- (a) – creation of vacancy in the K-shell; (b) electron transition from  $L_2$ - to K-shell;
- (c) – emission of electron (Auger-electron) from  $L_3$ -shell

The identification of atoms using the spectroscopy of core levels is based on the values of the binding energy of electrons. In the Auger electron spectroscopy, the energy of the emitted electron is determined by the difference in binding energies, which accompanies the relaxation of the atom excitation with redistribution of its electron shells and the emission of Auger electrons with characteristic energies.

Figure 4.92 shows non-radiative Auger-processes of excitation relaxation, in which the atom remains in the final state with two vacancies (or holes). If one of the vacancies of the final state lies in the same shell as the primary vacancy (although not at the same level), the nonradiative transition is called the Koster-Kronig transition. This transition is of a great importance, since the probabilities of the Koster-Kronig transitions exceed significantly the probabilities of the normal Auger transitions and, hence, affect the relative intensities of the detected Auger lines. The example in Fig. 4.92 shows that if  $L_1$ -level contains a vacancy, A fast transition from  $L_2$  to  $L_1$  (Koster-Kronig transition) will occur. Consequently, the intensity of the transition from  $M$ -shell to  $L_1$  vacancy will decrease.

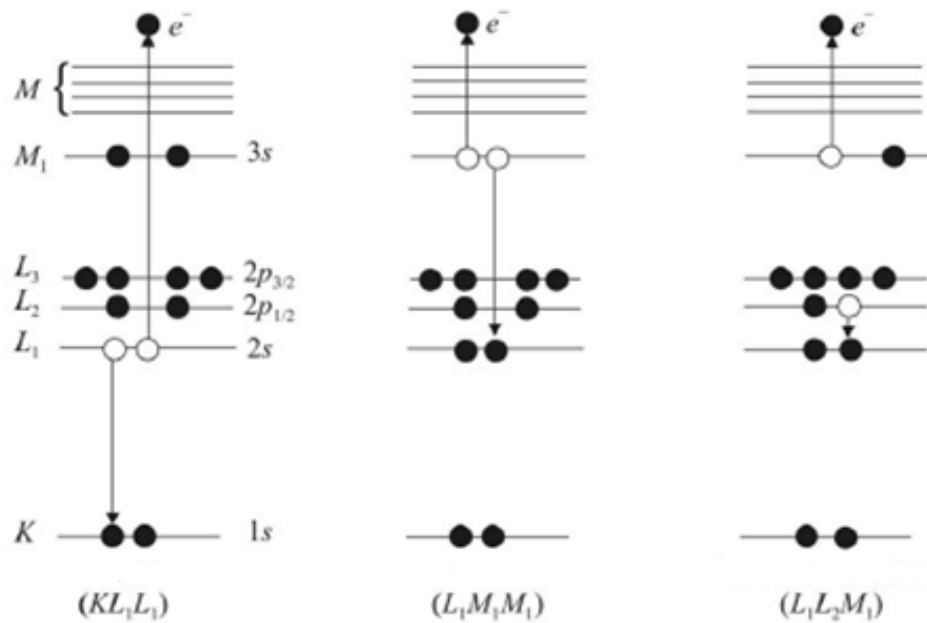


Figure 4.92. Schematic representation of various two-electron processes of excitation relaxation.  $KL_1L_1$  Auger-transition corresponds to initial hole in K-shell, which is filled with an electron from  $L_1$ -level, and at the same time another  $L_1$ -electron is emitted.  $LM_1M_1$  Auger transition corresponds to process with initial  $2s$ -vacancy.  $L_1L_2M_1$  Koster-Kronig transition corresponds to initial vacancy in the  $L_1$ -level, which is filled with electron from the same  $L$ -shell, but another level ( $L_2$ )

The identification of peaks in the spectrum of secondary electrons arising from Auger emission is currently not difficult. There is a large amount of databases with experimental Auger spectra, in which the main Auger peaks are marked with indication of their energies making it possible to reliable identification of elements in most cases.

There is a fundamental difference between the depth of analysis in the Auger electron spectroscopy and the energy-dispersive X-ray spectroscopy. In both cases, the analyzed particles (Auger electrons, X-ray quanta) are generated in the region with a micrometer-order depth. However, the mean free path of Auger electrons is about 1–3 nm, which determines the depth of analyzed region. To obtain information about deeper layers of the studied objects, the Auger electron spectroscopy is combined with ion etching.

The yield of Auger electrons for a free atom is determined by the product of the ionization cross section by electron impact and the probability of the emission (yield) of the Auger electron ( $1 - \omega_X$ ), where  $\omega_X$  is the probability of X-ray emission (fluorescence yield).

In the general case, the lifetime of the excited state  $\tau$  is determined by the sum of all possible relaxation processes. Radiative transitions occur with probability  $W_X$ . Auger transitions have a probability  $W_A$ , and  $W_K$  is the probability of the Koster-Kronig transitions (for which a hole is filled by an electron from the same shell). Since there are no other excitation mechanisms, one can write

$$\frac{1}{\tau} = W_X + W_A + W_K, \quad (4.85)$$

For transitions to the vacancies in the  $K$ -shell (as well as for holes at  $L_3$ - and  $M_5$ -levels), the Koster-Kronig transitions do not occur, and the probability of X-ray emission  $\omega_X$  is given by the following semiempirical relation

$$\omega_X = \frac{W_X/W_A}{1 + W_X/W_A}, \quad (4.86)$$

where

$$\frac{W_X}{W_A} = (a + bZ + cZ^3)^4. \quad (4.87)$$

As can be seen from the given relations, Auger transitions predominate for elements with small  $Z$  values. For them, emission of the Auger electrons is an important mechanism of relaxation of  $K$ -vacancies. Such dependence on  $Z$  does not mean a decrease in the Auger transition velocity at large  $Z$  values, but only emphasizes that X-ray transitions become the predominant way of relaxation of excitation at large  $Z$ .

The detectivity of the Auger electron spectroscopy is related to a strong background of secondary electrons in the energy spectrum of the electrons emitted by the object under investigation. Auger-electron transitions are usually manifested only as small features superimposed on this background. Therefore, it is common practice to use differentiating techniques and obtain  $dN(E)/dE$  function.

This makes it possible to minimize the contribution from a slowly varying background. An example of application of the spectrum differentiation technique for a case of incident electrons with an energy of 2 keV and a cobalt sample is shown in Figure 4.93. For a “pure” spectrum  $N(E)$ , mainly, the presence of a sharp peak related to elastically scattered electrons and an almost flat background are characteristic. The arrows in Fig. 4.93a show the transition energies for oxygen and cobalt. In the derivative of the spectrum (Fig. 4.93b), *LMM* signal of cobalt and *KLL* signal of oxygen are clearly detectable. The typical value of the registration limit is  $10^{-3}$ , i.e.  $\approx 0.1$  at.%.

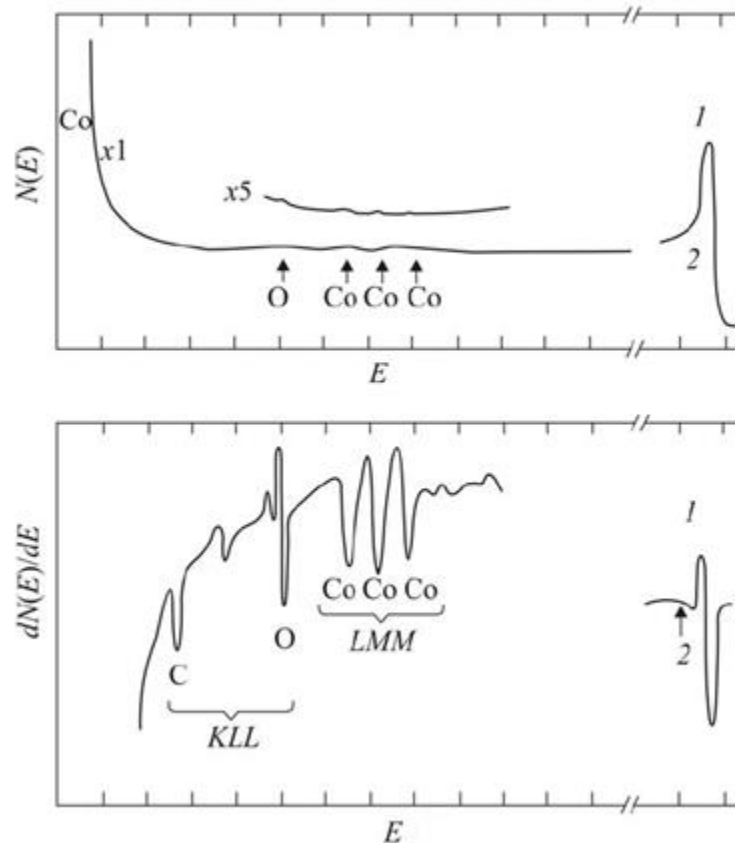


Figure 4.93. Comparison of  $N(E)$  spectrum (a) with its derivative  $dN(E)/dE$  (b) for electrons with an energy of 2 keV incident on the cobalt sample:

- 1 – peak of elastically backscattered electrons;
- 2 – peak corresponding to excitation of plasma oscillations

To perform an absolute quantitative concentration analysis using AES, it is necessary to find a relation between the current of Auger electrons of a given element and the concentration of this element in the near-surface region. Determination of the absolute concentration of an element  $X$  in a matrix using the yield of Auger electrons  $Y_A$  is complicated by the influence of the matrix on the backscattered electrons (which are able to cause the Auger effect along with incident electrons)



and the exit depth of the Auger electrons. The yield of *KLL* Auger electrons  $Y_A(t)$  generated in a thin sample layer with a thickness  $\Delta t$  at a depth  $t$  is determined by the relation

$$Y_A(t) = N_X \Delta t \cdot \sigma_e(t) [1 - \omega_X] \exp \left[ -\frac{t}{\lambda \cos(\Theta)} \right] \cdot I(t) \cdot T \cdot \frac{d\Omega}{4\pi}, \quad (4.88)$$

where  $N_X$  is the number of atoms of type  $X$  per unit volume;

$\sigma_e(t)$  is the ionization cross section at depth  $t$ ;

$\omega_X$  is the fluorescence yield;

$\theta$  is the registration angle;

$\lambda$  is the depth of yield;

$T$  is the transmittance of the analyzer;

$d\Omega$  is the solid registration angle;

$I(t)$  is the excitation electron flux at depth  $t$ .

The density of the exciting electron flow can conveniently be divided into two components:

$$I(t) = I_p(t) + I_b(t) = I_p(t)(1 + R_b(t)), \quad (4.89)$$

where  $I_p$  is the flux of primary electrons at a depth  $t$ ,  $I_b$  is the flux of backscattered primary electrons,  $R_B$  is the backscattering coefficient.

In quantitative Auger electron spectroscopy, the method of external standards is used most often. If external standards with known atomic concentration  $N_X^S$  of element  $X$  are used, the atomic concentration  $N_X^T$  in the sample under study can be found using relation for the yields of Auger electrons:

$$\frac{N_X^S}{N_X^T} = \frac{Y_X^S}{Y_X^T} \frac{\lambda_X^T}{\lambda_X^S} \frac{1 + R_B^T}{1 + R_B^S}. \quad (4.90)$$

Within this approach, knowledge of the ionization cross section of the fluorescent yield is not required, since yields of Auger electrons from the same chemical element are measured and compared. In addition, if the composition of the standard sample is close to the studied sample, the elemental composition can be determined directly from the ratio of the Auger electron yields (provided that the measurements are made identically). However, if the composition of the standard is significantly different from the sample under study, it is necessary to take into account the

influence of the matrix on probability of backscattering of primary electrons and the depth of electron yield. Parameters that include an inelastic mean free path length of electrons and strongly depend on the matrix must be corrected.

As it was noted above, in order to obtain in-depth concentration distributions of elements, the Auger electron spectroscopy method is combined with ion etching. For objects of a complex composition, their etching rate by ion beam is unknown as a rule. Therefore, to re-calculate the etching time scale to the depth scale, additional methods are necessary. As an example, Figure 8 shows the Auger profiles of the element concentrations for  $\text{In}_2\text{O}_3$  film grown on the ITO glass substrate and sensitized with CdS nanoparticles. To obtain the depth scale, cross-sectional SEM image of the film has been used.

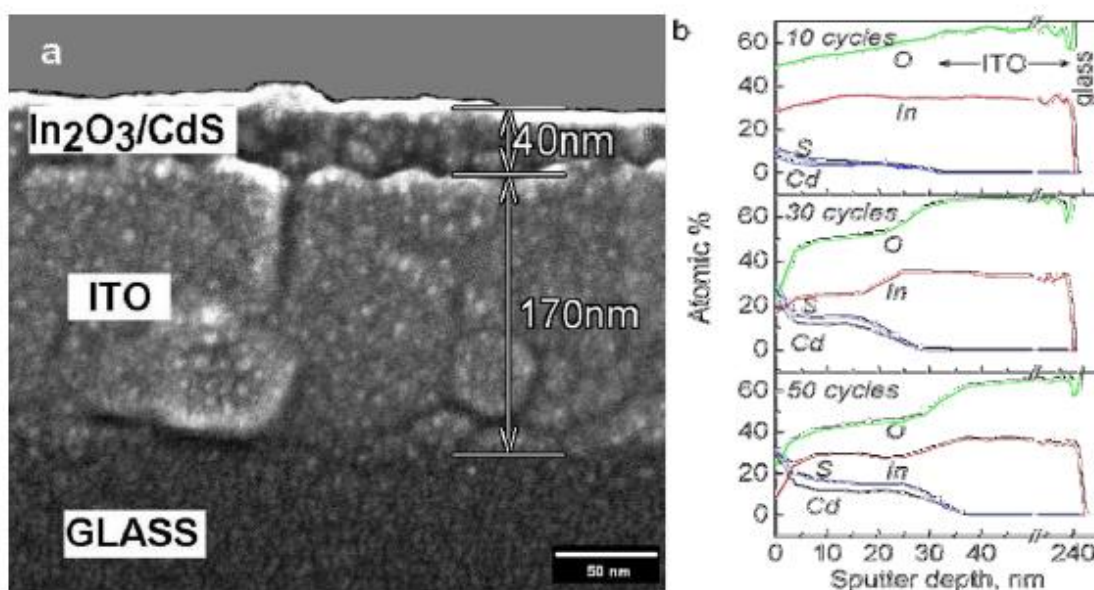


Figure 4.94. Cross-sectional SEM image of  $\text{In}_2\text{O}_3/\text{CdS}$  film grown on the ITO glass (a), Auger-profiles of element concentration (b) [3]

#### 4.7.5 Cathodoluminescence

As is known, cathodoluminescence (CL) is the appearance of light in the ultraviolet, visible and infrared spectral regions under irradiation by high-energy electrons. For a long time, interest in this physical phenomenon has been associated with the research and use of lumiphors for coatings of screens in cathode ray tubes. To date, the main purpose of the cathodoluminescence, along with other analytical methods, such as photoluminescence, optical absorption and reflection of light, is associated with the study of the electronic spectrum of solids (mainly, those energy levels that participate in radiative recombination of charge carriers).

The creation of highly efficient semiconductor lasers, light-emitting diodes, photodetectors, which often represent structures with a sub-micron size of layers, requires the application of analytical methods that are highly localized and allow the control of the distribution of radiative recombination centers created both by deliberate introduction of active impurities during the growing process and uncontrolled defects, which present in semiconductor structures. In this sense, the scanning electron microscopy provides great opportunities for local cathodoluminescence analysis.

The most obvious advantages of cathodoluminescence in comparison with photoluminescence are the following. First, the high energy of the incident electrons  $E_i$  ( $10^3 - 10^4$  eV) makes it easy to excite electron-hole pairs in materials with arbitrarily large band gap energy  $E_g$ . This makes cathodoluminescence an almost indispensable method for analysis of wide-gap semiconductors and dielectrics, where the application of optical excitation is extremely difficult. Secondly, since the  $E_i / E_g$  ratio equals to  $10^3 - 10^4$ , each incident electron is able to generate a corresponding number of electron-hole pairs, which multiplies the intensity of the luminescence signal.

As an example, Fig. 4.95 shows the cathodoluminescence spectra of synthetic diamond. As can be seen, a narrow line ( $\approx 235$  nm) corresponding to the recombination of free excitons, as well as a number of bands corresponding to defects creating energy levels in the forbidden gap of diamond are observed in the spectra. The ratio of the intensity of the free exciton line to the intensity of the defective lines is maximal for the region 2 indicating the lowest concentration in it of defects, which play the role of radiative recombination centers. This simple example shows the high information value of the cathodoluminescence method.

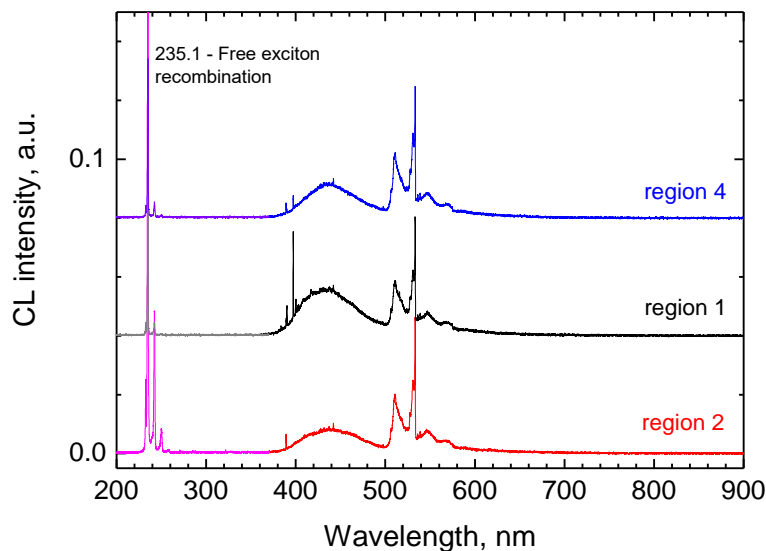


Figure 4.95. Cathodoluminescence spectra of synthetic diamond crystal measured at a temperature of 80 K under excitation of different sample regions [4]

#### 4.7.6 Electron beam probing for study of semiconductor materials and devices

As is known, the photovoltaic effect can be realized in any semiconductor structure, in which there exists an internal electric field separating non-equilibrium charge carriers generated by incident optical radiation. At the same time, the basic material for the creation of the solar cell is silicon, on the basis of which at the present time more than 85% of all solar cells are manufactured. Therefore, we consider below the application of the scanning electron microscopy for investigation of those properties of silicon, which are especially important for its use in production of solar cells.

One of the characteristic features of crystalline silicon is small (in comparison with many other semiconductor materials) value of the light absorption coefficient. This leads to the fact that a significant part of quanta is absorbed in silicon at a large depth. The calculation shows that the thickness of the silicon solar cell should be about 100  $\mu\text{m}$  to absorb 95% of light quanta with an energy exceeding the band gap energy. In order to contribute to electrical power of solar cell by light generated electron-hole pairs, the minority carriers must reach the region, in which the electric field of  $p$ - $n$  junction is localized. In a typical solar cell, this field is located in about 1  $\mu\text{m}$  thick layer near the face surface of the wafer. It is clear that carriers generated rather far from the  $p$ - $n$  junction can give a contribution to the electric power of solar cell only in the case, if they can reach  $p$ - $n$  junction until their recombination (death). They need to travel a distance of about several tens of micrometers in the diffusion process. In other words, the material from which the solar cell is made should be characterized by high values of lifetime  $\tau$  and the diffusion length  $L$  of the minority charge carriers, which is achieved only at a high degree of structural perfection of material. These two parameters –  $\tau$  and  $L$  – are fundamental for assessing the suitability of crystalline silicon for the creation of solar cells.

It should be noted that the diffusion length of carriers is more significant parameter than the lifetime for characterizing the quality of photovoltaic materials and the solar cells created on their basis. Indeed, the lifetime characterizes only the processes of carrier recombination, while from the standpoint of the solar cell operation, the most important factor is the efficiency of carriers collection, which is determined not only by the rate of their recombination, but also by their diffusion coefficient, i.e., in fact, by the diffusion length.

Due to the extremely importance of  $L$  parameter, let's acquaint with the physical basis of the methods for determining it. If single-crystal silicon is used for manufacturing solar cells, the lifetime and diffusion length are, as a rule, rather uniformly distributed in the wafer plane, so the methods for their determination do not require high locality. On the contrary, if polycrystalline silicon is used for fabrication of solar cells, it is preferably to apply the methods with high spatial resolution that would enable one to track the influence of structural defects (such as grain boundaries, dislocations, etc.) present on such material on its properties.

Methods for determination of the diffusion length are based, as a rule, on the analysis of a signal that is formed by non-equilibrium carriers generated by a light or electron beam and separated by

some internal electric field. The latter can be created by the space charge of the  $p$ - $n$  junction, the Schottky barrier or the semiconductor-electrolyte contact. The variation of the distance between the region of generation of non-equilibrium charge carriers and the region of localization of the separating electric field (which can be realized by mechanical displacement of a light or electronic probe, and also by varying the spectral composition of light or the energy of an electron beam), in turn, results in a change in the photoresponse, which will depend, obviously, on the value of the diffusion length.

Method for determination of the diffusion length based on the analysis of the spectral dependence of the photocurrent upon illumination of a crystal with a  $p$ - $n$  junction formed at the surface (or a made solar cell) with variable wavelength light is one of the most common. In order to determine the diffusion length in the material, a Schottky barrier in the "semiconductor-electrolyte" system or the bending of the energy bands always present on the surface can be used instead of the  $p$ - $n$  junction.

As an alternative to the methods based on optical excitation of non-equilibrium charge carriers, the methods based on irradiation of material or solar cell under study with an electron beam with an energy from several keV to several tens of keV can be used.

To measure the lifetime and diffusion length of minority carriers, the method of electron beam induced current (EBIC) is widely used. This method can be easily realized in a scanning electron microscope. One of its advantages is related to an internal amplification of the signal, which arises due to the fact that each primary electron generates a large number (usually  $10^3 - 10^4$ ) of electron-hole pairs leading to a high quantum yield.

Let us give the relations, which form a mathematical basis for analysis of results of such experiments. An electron-induced current collected by a  $p$ - $n$  junction or a Schottky barrier can be calculated by knowing the spatial distribution of the non-equilibrium charge carriers generated by the electron beam, which obeys the continuity equation:

$$\nabla^2 p(x, y, z) - \frac{p(x, y, z)}{L^2(x, y, z)} + \frac{G(x, y, z)}{D} = 0 \quad , \quad (4.91)$$

where  $p$  is the excess concentration of minority carriers,

$L$  is the local value of the diffusion length,

$D$  is the diffusion coefficient of minority carriers,

$G$  is the generation rate of minority carriers,

$x$  и  $y$  are the cartesian coordinates in the plane of the  $p$ - $n$  junction or the Schottky barrier,

$z$  is the coordinate in the perpendicular direction.

The boundary conditions for (12) have the following form:  $p(x, y, w) = 0$ ,  $p \rightarrow 0$  при  $z \rightarrow \infty$ ,

where  $w$  is the depth of the lower boundary of the space-charge region of the  $p$ - $n$  junction or the Schottky barrier. The collected current of the minority carriers can be represented in this case in the form

$$I = eD \iint \frac{\partial p}{\partial z}(x, y, w) dx dy . \quad (4.92)$$

An electron beam-induced current can also be represented as

$$I = e \iiint G(x, y, w) \psi(x, y, w) dx dy dz , \quad (4.93)$$

where the function  $\psi(x, y, z)$  represents the probability of collection of generated carriers and can be interpreted as a current collected from a single charge created at a point with coordinates  $(x, y, z)$ . Function  $\psi(x, y, z)$  can be found from the solution of equation

$$\nabla^2 \psi(x, y, z) - \frac{\psi(x, y, z)}{L^2(x, y, z)} = 0 , \quad (4.94)$$

the boundary conditions for which take the form

$$\psi(x, y, z < w) = 1, \quad \psi(x, y, \infty) = 0 . \quad (4.95)$$

The first condition means, in fact, the absence of recombination in the space charge region. The solution of Eq. (4.94) under the conditions (4.95) for a homogeneous sample enables one to obtain an analytic expression for the function  $\psi(x, y, z)$ :

$$\psi(x, y, z) = \begin{cases} \exp\left(-\frac{z-w}{L}\right), & z > w \\ 1, & z < w \end{cases} . \quad (4.96)$$

However, it is necessary to note that taking into account surface recombination substantially complicates the relations. To reduce the effect of surface recombination, it is effective to use a method based on measurement of the dependence of the induced current on the penetration depth of incident electrons, i.e. on the energy of the electron beam. For a homogeneous sample with a Schottky barrier, the current of the collected minority carriers can be represented in the form

$$I = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{w'}^{\infty} G(x, y, z) \exp\left(-\frac{z-w}{L}\right) dx dy dz + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{t_m}^w G(x, y, z) dx dy dz , \quad (4.97)$$

which enables us to determine the value of the diffusion length. At the same time, the disadvantage of this method is related to necessity to use sufficiently high electron energies when studying samples with large diffusion lengths. Moreover, analysis of Eq. (19) shows that the diffusion length can be reliably determined only under the condition  $L < R$ , where  $R$  is the projection range of the incident electrons.

An exceptionally wide field of application of the electron (light) beam induced current method is the investigation of electrically active defects in inhomogeneous semiconductors (for example, polycrystals). The so-called recombination contrast is used as a quantitative characteristic showing the degree of electrical activity of the defect:

$$C = \frac{I_d - I_0}{I_0}, \quad (4.98)$$

where  $I_d$  and  $I_0$  are the values of induced current at the defect and away from it, respectively. It is necessary to emphasize, however, that the influence of a defect on the magnitude of the induced current is determined not only by a different recombination rate on it, but also by a built-in electric fields, which can lead to an additional separation of non-equilibrium carriers and an increase in the detected signal. Attempts of quantitative interpretation of the induced current signal were carried out in the eighties for dislocations and grain boundaries. However, the assumptions used in the proposed models for calculation of the parameters of electrically active defects from the induced current distributions do not take into account their interaction with point defects and can hardly be considered adequate, since it is well known that there are always special defect-impurity atmospheres in the vicinity of dislocations and grain boundaries. Therefore, to date, the study of recombination contrast in inhomogeneous samples is used mainly a the qualitative evaluation of electrically active defects. It should be noted that the recombination contrast can also be studied by scanning the sample being investigated not by a light beam.

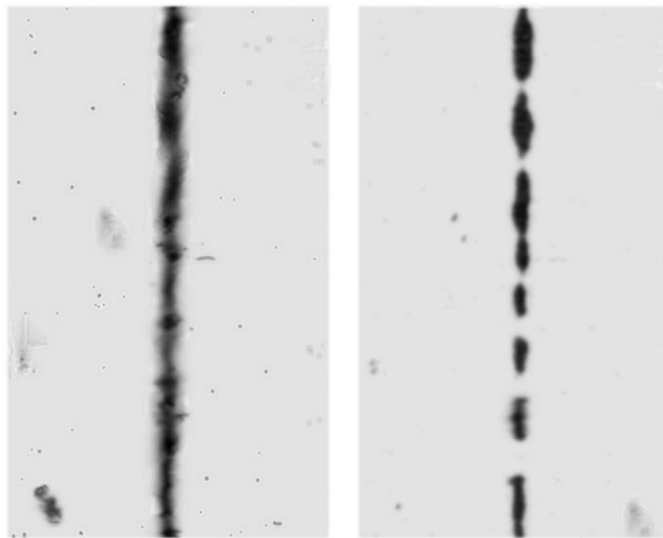


Figure 4.96. Recombination contrast for grain boundaries of various types in silicon.

Dark areas correspond to small current values, light areas correspond to small values [5]

The need to create contacts to the sample and form barriers for separation of non-equilibrium carriers is a significant drawback of the electron (light) beam induced current method, which limits its application. As an alternative to the electron (light) induced current method, the method of

surface electron-induced potential in scanning electron microscopy can be considered. The undeniable merit of this method is related to an expressness and non-destructive nature.

The scheme of this method is shown in Fig. 4.97. The investigated semiconductor crystal 1 is grounded on the stage of the scanning electron microscope 2, and its upper surface remains free. If the sample is irradiated with an electron probe 3, the non-equilibrium carriers generated in the volume 4 are separated by a field of the space-charge region (the presence of which is due to  $p$ - $n$  junction or of surface electronic states), thereby leading to a change in the surface potential. Obviously, this change will be sensitive to the diffusion length of minority carriers, which is the basis of the method. At a small (about 0.5 mm) distance from the surface of the test sample, an annular detector 5 is placed, the signal in which is induced due to the capacitive coupling between it and the grounded sample and enters the measuring device 6.

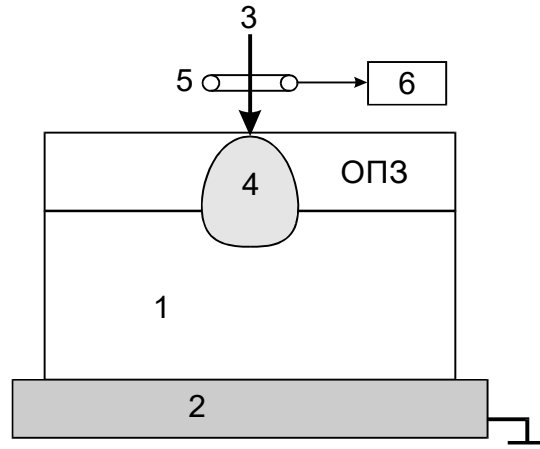


Figure 4.97. Scheme of registration of surface electron-induced potential in a scanning electron microscope.

The calculation of the diffusion length is based on the following relations. Under excitation of non-equilibrium carriers by an electron probe, the change of the surface potential can be represented as

$$\pm \Delta V = C \beta \ln \left( 1 + \frac{G_{eff} I_0}{I_s} \cdot \exp \left( -\frac{X}{L} \right) \right), \quad (4.99)$$

or

$$\pm \Delta V = C \beta \ln \left( 1 + \frac{\Delta n}{N_0} \cdot \exp \left( -\frac{X}{L} \right) \right), \quad (4.100)$$

where  $C$  is the capacitive coupling factor of the detector,

$\beta = kT/e$  is the thermal potential,



$G_{\text{eff}} = G_0 S / (1 + S)$  is the effective carrier generation rate,

$G_0 = (E_0 / E_{\text{eh}}) (1 - k)$ ,

$S$  is the reduced surface recombination rate,

$E_0$  is the energy of the incident electrons,

$E_{\text{eh}}$  is the energy of formation of electron-hole pair,

$k$  is the fraction of the reflected energy of the probing electron beam,

$I_0$  is the electron probe current,  $I_s$  is the reverse saturation current of the p-n junction,

$\Delta n$  and  $N$  are concentrations of non-equilibrium and equilibrium carriers, respectively.

The "+" sign is selected for the depleted near-surface layer in the case of  $n$ -type base, and "-" for the  $p$ -type base. Unfortunately, the above-given relations do not allow to calculate directly the diffusion length, since they contain parameters that are almost indeterminable (for example,  $G$ ,  $I_s$ ,  $C$ ,  $S$ ). To avoid this difficulty, one usually proceed as follows. Under series scanning with an electron probe with different current values  $I_{01}$  and  $I_{02}$  along  $x$  coordinate normal to  $p$ - $n$ -junction,  $\Delta V(x)$  dependencies are registered. One can always choose such points  $x_1$  and  $x_2$  on these dependencies that the following condition will be true

$$\Delta V(x_1, I_{01}) = \Delta V(x_2, I_{02}), \quad (4.101)$$

i.e.

$$\Delta V_1 = C \beta \ln \left( 1 + \frac{G_{\text{eff}} I_{01}}{I_s} \cdot \exp \left( -\frac{X_1}{L} \right) \right) = \Delta V_2 = C \beta \ln \left( 1 + \frac{G_{\text{eff}} I_{02}}{I_s} \cdot \exp \left( -\frac{X_2}{L} \right) \right), \quad (4.102)$$

from which it follows that

$$\ln \left( \frac{I_{01}}{I_{02}} \right) = \frac{X_2 - X_1}{L}. \quad (4.103)$$

Note that for determination of the diffusion length, it is necessary to know only the difference in coordinates corresponding to equal changes of the surface potential.

The method of surface electron-induced potential allows independent determination of the lifetime of minority charge carriers. It should be noted that the relaxation of the surface potential induced by an electron probe is not exponential, therefore, the analysis of the kinetics of the decay in the signal is not applicable for the determination of the lifetime. Typically, a phase shift between the irradiation pulse and the maximum of the detected signal is registered. In the case of scanning with an electronic probe along a coordinate normal to the plane of p-n junction, the gradient of the phase shift is given by:

$$L \frac{d\varphi}{dX} = 0.62 (\omega \tau)^{1/2} \cdot \left( 1 + 2.8 (\omega \tau)^{-3/2} \right)^{-1/3}, \quad (4.104)$$

where  $\omega$  is the modulation frequency of the electronic probe.

#### 4.7.7 References

1. A.V. Mazanik, unpublished.
2. E.A. Bondarenko, E.A. Streltsov, M.V. Malashchonak, A.V. Mazanik, A.I. Kulak, E.V. Skorb, Giant Incident Photon-to-Current Conversion with Photoconductivity Gain on Nanostructured Bismuth Oxysulfide Photoelectrodes under Visible-Light Illumination, *Advanced Materials*, 2017, DOI: 10.1002/adma.201702387.
3. N.V. Malashchonak, S.K. Poznyak, E.A. Streltsov, A.I. Kulak, O.V. Korolik, A.V. Mazanik, Photoelectrochemical and Raman characterization of  $\text{In}_2\text{O}_3$  mesoporous films sensitized by CdS nanoparticles, *Beilstein Journal of Nanotechnology*, 2013, 4, 255–261.
4. N.M. Kazuchits, A.V. Mazanik [et al.], to be published.
5. A. Fedotov, N. Drozdov, E. Katz, Yu. Ilyashuk, A. Mazanik, A. Ulyashin, Transformation of electrical activity of extended defects in silicon polycrystals under annealing and hydrogen plasma treatment, *Physica Status Solidi (a)*, 1999, 171, 353–363.

## 4.8 Diffraction methods of analysis

### 4.8.1 Introduction

Study of materials by X-ray diffraction (XRD) is one of the main analytical methods used in the materials science. The application of this method is intended most often to qualitative and quantitative determination of the phase composition of investigated objects. However, due to ability to determine the parameters of crystal lattice with a high accuracy, the method makes it possible to study the transformation of the crystal structure under various impacts: irradiation with high-energy particles, deformation, variation of temperature.

A low lateral resolution is essential disadvantage of the XRD analysis. This drawback is eliminated by replacing X-ray radiation with an electron beam. Since both elastically scattered and transmitted through the object electrons diffract, the electron diffraction (ED) can be realized in both scanning and transmission electron microscopes. Another important advantage of using electrons in diffraction studies is significantly larger (approximately by three orders of magnitude) cross sections for their scattering by atoms. This becomes especially important for the study of thin (about  $10^{-8}$  m or less) films and layers, when the application of the XRD method becomes problematic. In this case, low-energy electrons (with an energy of less than  $10^3$  eV) are used. The depth of their penetration into material does not exceed several atomic layers, which enables to obtain a diffraction pattern formed predominantly by a thin near-surface layer of the object.

### 4.8.2 Theoretical basis of diffraction analysis

Determination of the phase composition of solids is a widespread problem of materials science. Complex systems consisting of atoms of several sorts are, as a rule, multiphase. For example, in the "bismuth-vanadium-oxygen" system, the presence of bismuth and vanadium oxides, as well as bismuth vanadate is possible; the latter can be crystallized in several crystalline modifications. Each crystalline phase has its own unique crystal lattice, which is characterized by a certain symmetry and unique numerical parameters (the lengths of the cell parameters and the angles between them).

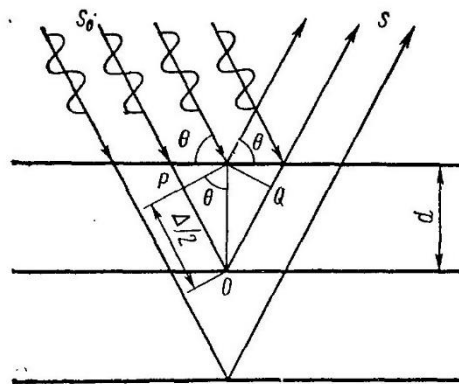


Figure 4.98. Diffraction on atomic planes

Let us consider a monochromatic radiation incident on a system of atomic planes parallel to each other. If the rays are mirrored from neighboring planes, then the condition for their interference amplification has the form

$$2d \sin(\Theta) = n\lambda \quad (4.105)$$

where  $d$  is the distance between neighboring crystallographic planes, with atoms of which X-rays interact (Fig. 4.98),

$\theta$  is the diffraction angle;

$n$  is the order of the diffraction maximum (the order of "reflection");

$\lambda$  is the wavelength of monochromatic X-rays incident on a crystal.

The formula (4.105) is called the Bragg's law. It is the basis for determining the phase composition and lattice parameters from the analysis of both X-ray and electron diffraction patterns. Knowing the wavelength of the radiation and the diffraction angle, it is easy to calculate the interplanar distance using the relation (4.105). Further, starting from the relationship between the interplanar distances and the lattice parameters, it is possible to determine their values. For example, for the cubic lattice

$$d = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \quad , \quad (4.106)$$

where  $a$  is the lattice parameter,

$h$ ,  $k$  and  $l$  are the Miller indices of the corresponding family of planes. Similar relationships are known for other types of lattices.

The concept of the reciprocal lattice is also often used for analysis of diffractograms. The Bragg's condition (1) can be represented in another way: the interference amplification of the rays scattered by the lattice occurs, if the change of wavevector of radiation is equal to one of the reciprocal lattice vectors of the crystal.

The so-called Debye-Scherrer method (powder method) is used most often for determination of the phase composition of the investigated objects. In this method, the object of investigation is a polycrystal consisting of a set of individual single crystals. In this case, for planes with any Miller indices, there is always a certain number of crystallites with orientation corresponding to a "reflecting" position. If the object under investigation represents a single crystal or a polycrystal with large grain sizes, it is necessary to prepare a powder to realize the Debye-Scherrer method.

The scheme of implementation of the Debye-Scherrer method in the so-called Bragg-Brentano geometry is shown in Figure 4.98. The source of monochromatic X-ray radiation and detector of the scattered rays move towards each other at the same angular velocity.

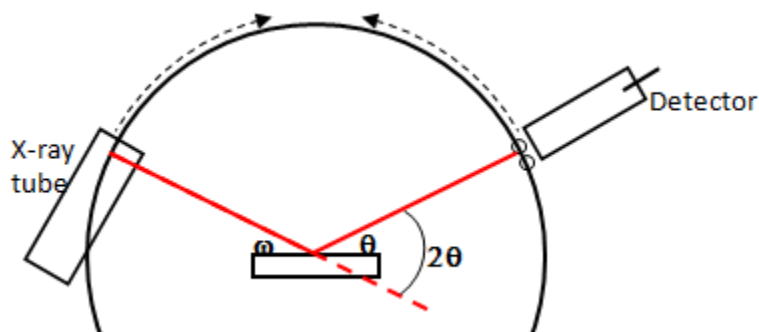


Figure 4.99. Bragg-Brentano geometry

Figure 4.100 represents an example of the solution of the most common problem of X-ray diffraction analysis related to the qualitative determination of the phase composition. The investigated objects represent  $\text{Cu}_2\text{O}$  films electrochemically deposited on FTO (fluorine doped tin oxide) substrates. As is known, copper can be oxidized to oxidation states Cu(I) and Cu(II), therefore, in the general case, it is possible to observe the presence of different phases (Cu,  $\text{Cu}_2\text{O}$ , CuO) in the given system. As can be seen from Fig. 4.100, there is a significant number of peaks in the diffraction patterns. Their comparison with existing databases makes it possible to qualitatively establish the phase composition. It can be seen that variation of the conditions of electrochemical deposition does not results in the appearance or disappearance of any peaks, which indicates that the phase composition of the films is unchanged.

As it was mentioned in the Introduction, diffraction analysis is often used to analyze changes in the system under the influence of various external factors. The example given in Figure 3 demonstrates the change in the phase composition of the above-considered  $\text{Cu}_2\text{O}$  film after its illumination in an electrolyte solution. As can be seen from Fig. 4, the X-ray patterns show peaks corresponding to elementary copper, which points to the reduction of  $\text{Cu}_2\text{O}$  during the cathode photoelectrochemical process.

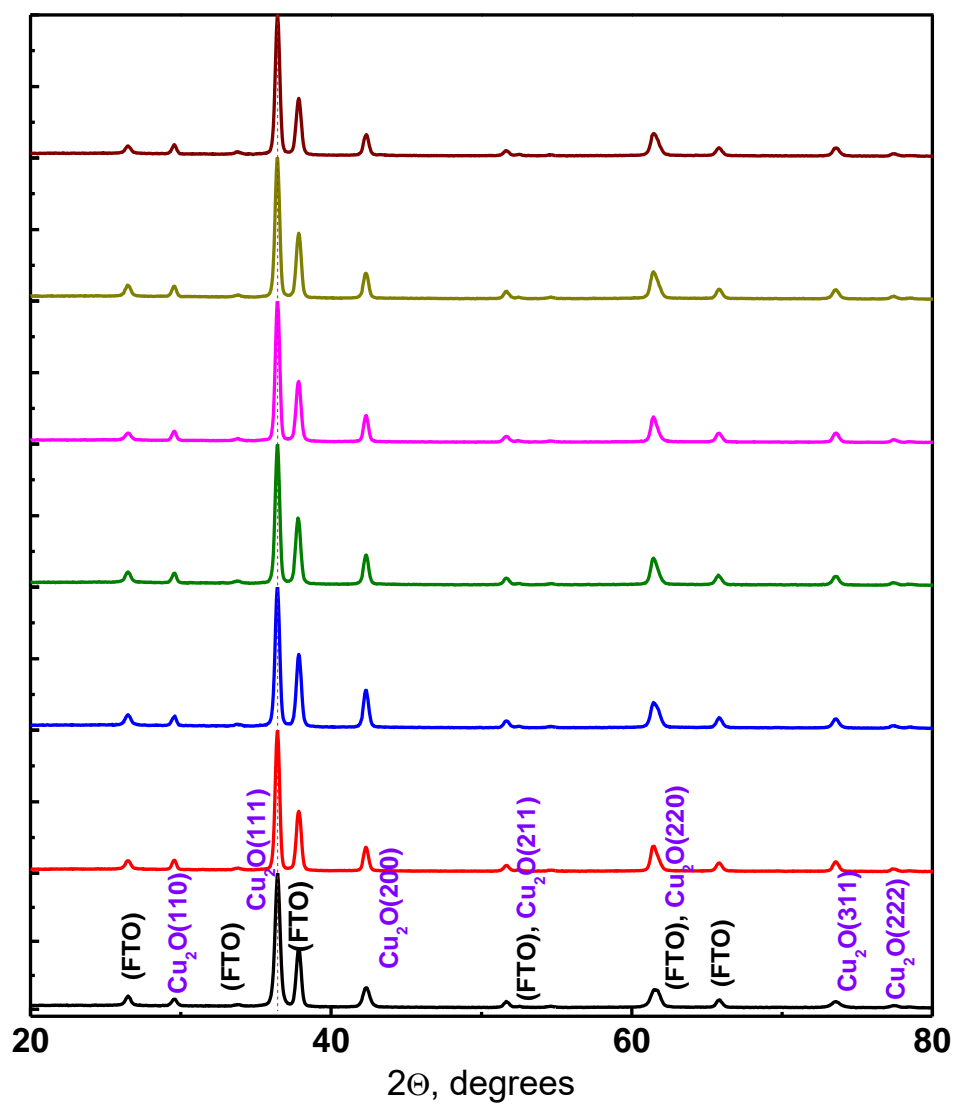


Figure 4.100. X-ray diffraction patterns of  $\text{Cu}_2\text{O}$  films electrochemically deposited on FTO substrate [1]

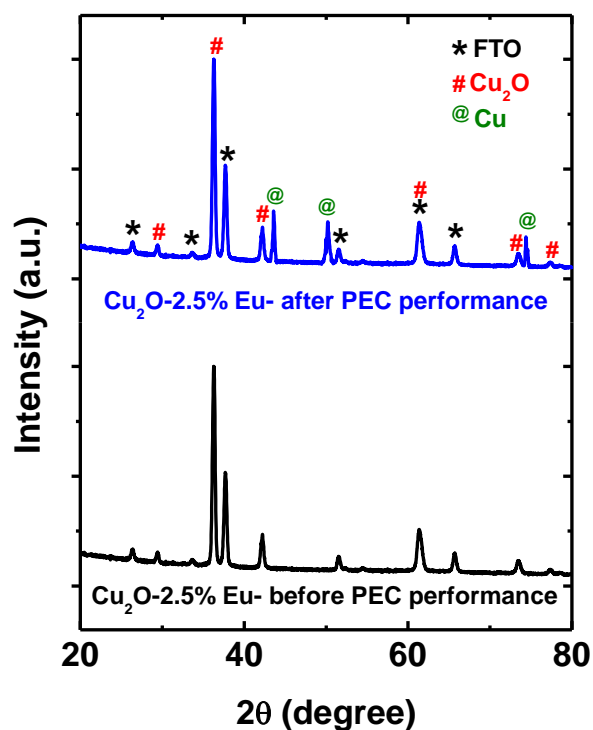


Figure 4.101. X-ray diffraction patterns of  $\text{Cu}_2\text{O}$  films electrochemically deposited on FTO substrate before and after electrochemical tests [1]

#### 4.8.3 Determination of the texture of polycrystalline objects

As is known, the texture refers to the presence of a preferential orientation of crystallites (grains) in polycrystalline objects. Obviously, X-ray diffraction allows an answer to the question of whether or not a texture exists. As an example, Fig. 4.102. shows the diffractograms of bismuth oxybromide films synthesized at the department of electrochemistry of BSU using two different techniques. Bismuth oxybromide is known as a wide band gap semiconductor (band gap energy of 2.9 eV), which has a high photocatalytic activity under illumination by light with spectral composition corresponding to the fundamental absorption. As can be seen from Figure 4.102., all objects are characterized by the same set of reflexes, which indicates the identity of their phase composition. However, the relative intensity of the reflexes is different for different samples, which indicates a difference in the predominant orientation of the crystallites.

#### 4.8.4 Broadening of X-ray reflexes

Thus, the position of the X-ray reflexes makes it possible to determine the interplanar distances for the known wavelength of the incident radiation, which opens the possibility of establishing the phase composition. The width of reflexes also contains important information about the object under study. There are the following mechanisms of reflex broadening: (i) instrumental, (ii)

determined by a finiteness of the coherent scattering region, (3) associated with the presence of elastic deformations in the crystal. Instrumental broadening is due to an imperfect monochromaticity of the radiation used, as well as to the finite (non-zero) dimensions of the source and X-ray detector. To determine the instrumental contribution to the broadening of X-ray reflexes, one use the crystals of a high structural perfection, for which the contribution of other mechanisms to the overall broadening of the lines is negligibly small.

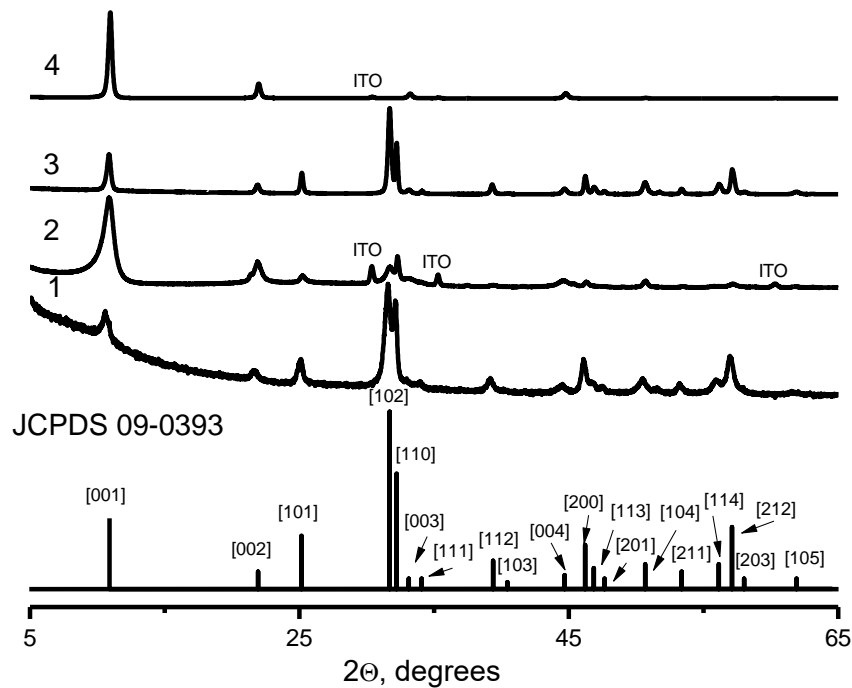


Figure 4.102. X-ray diffraction patterns of BiOBr samples [2]: 1 – powder of type 1; 2 – film of type 1; 3 – powder of type 2; 4 – film of type 2

Knowing the instrumental contribution to the broadening of the reflex, one can calculate the important parameters of the object, on which the diffraction takes place: the size of the coherent scattering region and the magnitude of the elastic deformations. To do this, one use the Williamson-Hall formula:

$$\beta = \frac{k\lambda}{D \cos(\Theta)} + 4\epsilon \tan(\Theta) \quad , \quad (4.107)$$

where  $\lambda$  is the wavelength of X-ray radiation,

$D$  is the size of the coherent scattering region,

$\Theta$  is the diffraction angle,

$\epsilon$  is the lattice strain,



$\beta$  is the full width at the half-maximum of reflex in the  $2\Theta$  scale (in radians),

$k$  is a parameter close to unity.

As can be seen from equation (3), in order to determine the size of the coherent scattering region and the value of the relative deformation, it is necessary to have at least two (preferably several) peaks in the diffractogram. As an example, Figure 6 demonstrates the Williamson-Hall plot for nanocrystalline cerium oxide.

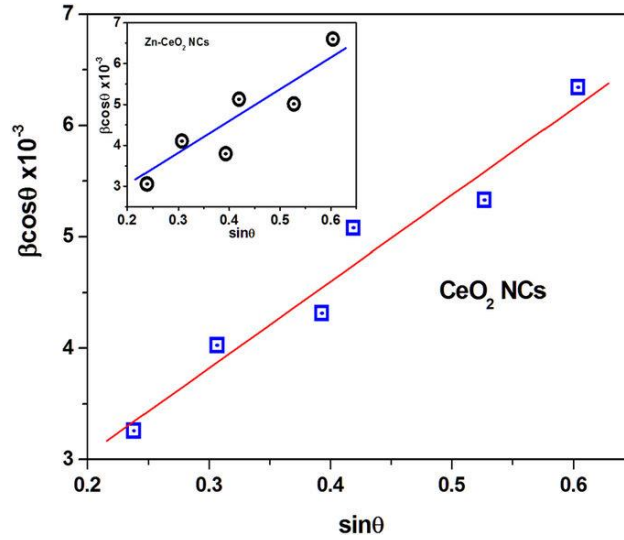


Figure 4.103. Williamson-Hall plot for nanocrystalline cerium oxide [3]

In the particular case of the absence of deformations, formula (4.107) can be rewritten as

$$D = \frac{k\lambda}{\beta \cos(\Theta)} \quad (4.108)$$

The relation (4.108) is known as the Scherrer formula. It formally implies that even a single diffraction reflex is sufficient to determine the size of the coherent scattering region. In modern materials science, the Scherrer formula is often used to estimate particle sizes. However, it should be remembered that the method based on the relation (4.108) is indirect, therefore, it is preferable to compare the values of the parameter  $D$  obtained from the formula (4.108) with the results of electron microscopy.

#### 4.8.5 Diffraction of electrons

As it was noted above, the disadvantage of X-ray diffraction analysis is related to the low lateral resolution of the method. However, this drawback is eliminated by replacing X-ray radiation with an electron beam.

Diffraction can be realized both for a beam of electrons, which are elastically reflected from the object under study, and for a beam of electrons passing through the sample. Diffraction of elastically reflected electrons can be realized in a scanning electron microscope (see the previous chapter), diffraction of electrons passing through the object is realized in a transmission electron microscope.

Both scanning and transmission electron microscopes make it possible to reveal details of the microstructure of an object that are beyond the resolution of the optical microscope (i.e., less than 0.2  $\mu\text{m}$ ). This is due to the fact that the de Broglie wavelength for electrons with energies typical for scanning and transmission electron microscopy is much smaller than the wavelength of optical radiation. As is known, the de Broglie wavelength of an accelerated charged particle is determined by the relation

$$\lambda = \frac{h}{\sqrt{2mE}} = \frac{h}{\sqrt{2meU}} \quad , \quad (4.109)$$

where  $h$  is the Planck's constant,  $m$  and  $e$  are the mass and charge of an electron, respectively,  $E$  is the electron energy,  $U$  is the accelerating voltage.

For a typical electron energy of 10 keV in a scanning electron microscope, we obtain  $\lambda_{\text{SEM}} = 1.2 \cdot 10^{-11}$  m, and for a typical electron energy of 100 keV in a transmission electron microscopy  $\lambda_{\text{TEM}} = 3.9 \cdot 10^{-12}$  m.

Modern transmission electron microscopes possess a subnanometer resolution, which can be as good as a tenth of a nanometer. It is also possible to obtain in them a diffraction pattern from the sample region with a linear size of the order of 1  $\mu\text{m}$  – the so-called microdiffraction pattern. Obviously, this possibility is extremely useful for analyzing inhomogeneous objects.

The most effective use of the transmission electron microscopy is achieved if the objects under study have a thickness of about a hundred nanometers. The preparation of such samples is often the most difficult task that must be solved for the study of objects by the transmission electron microscopy. For this purpose, ultrasonic dispersion, ion etching and other methods of sample preparation are widely used.

#### 4.8.6 Scheme of transmission electron microscope

Since electrons possess an electric charge, one can control their trajectory by means of electric and magnetic fields. This makes it possible to focus the electron beam and thus obtain electronic images.

The source of high-energy electrons (as a rule, their energy is from 100 to 300 keV) is the so-called electron gun. In fact, it is a three-electrode electrostatic lens (Figure 4.103.). The electron source is a cathode (a heated filament of tungsten or lanthanum hexaboride or a field emission cathode), which has a high negative potential with respect to the grounded anode. A slightly more negative potential compared with the cathode is applied to a focusing electrode. The field between the

cathode and the focusing electrode acts as a collecting lens. The anode is grounded, i.e., has a large positive potential with respect to the cathode and the focusing electrode.

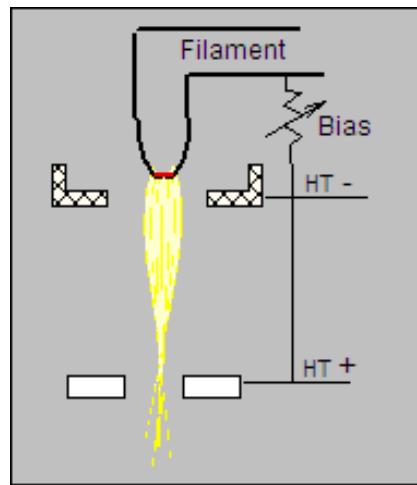


Figure 4.104. Scheme of a three-electrode electron gun.

Thus, the electron gun is the source of a narrow intense beam of fast electrons. To control the movement of electrons, one uses electromagnetic lenses, which represent solenoids (coils with a winding with a constant electric current).

As is known from general physics, when an electron moves in a magnetic field, its trajectory changes, if there is a non-zero velocity component perpendicular to the magnetic field. Since the electron also has a velocity component parallel to the magnetic field, it moves in the solenoid along a helicoidal trajectory. Electrons that exit simultaneously from any point A along the magnetic field lines will collect simultaneously at a certain point B. By analogy with the optics, we can say that the point B is the image of the point A (while the vector AB is parallel to the magnetic field lines). Such image will not be enlarged, since the magnetic field is uniform in a long solenoid and, therefore, the focal length  $F = \infty$ . An enlarged image can be obtained by means of an inhomogeneous magnetic field possessing axial symmetry.

Magnetic lens is a solenoid enclosed in a shell of a soft ferromagnet having an air gap inside the solenoid, where a strong non-uniform magnetic field is created. In the short-focus lens, in addition, there are pole pieces that allow localization of a particularly strong magnetic field in a small region.

The change of current in the coils of the solenoid makes it possible to vary the intensity of the magnetic field, and hence the focal length (force) of the lens. Long-focus lenses of electron microscopes give a magnification of about 10, and short-focus lenses of the order of 100 times.

The principal optical scheme of the transmission electron microscope is shown in Figure 8.

The electron gun and the electromagnetic condenser lens together with the condenser diaphragm form on the sample a slightly divergent beam of fast electrons. The sample is placed in the front focal plane of the strong objective lens. Electrons scattered by the sample are focused in the back focal plane of the objective, where the aperture diaphragm is located; the electrons scattered at different points of the sample at the same angle are collected in the same point. Thus, in this plane the primary diffraction image of the object (the electron diffraction pattern) is formed. The interference of these diffracted beams forms in a plane optically conjugated with the front focal plane of the lens a primary microscopic image of the object that is inverted and magnified by approximately 100 times. In this plane, the selector diaphragm is placed. Focusing is carried out by changing the current in the winding of the objective lens. The image from the plane of the selector diaphragm is transferred by a weak intermediate lens to the plane of the field diaphragm with a small magnification (up to 10 times), and the main projection lens (strong) transfers this image from its front focal plane to its back focal plane, where the screen is located, with magnification of about 100 times. Thus, the total magnification of a microscope with three image forming lenses can reach  $10^5$  times. Due to the use of two lenses in the projection system, it is possible to easily change the magnification in very wide range (starting from optical ones) with a corresponding change in the imaged field of the object and with the unchanged size of the final image. The change in magnification is achieved by smoothly changing the current in the intermediate lens and stepwise change of the magnification range in the projection lens.

The presence of an intermediate lens also makes it easy to obtain on the screen an enlarged image of the diffraction pattern. Indeed, it is possible to establish the current of the intermediate lens so that there is a primary diffraction image in its front focal plane, which will be transferred in the plane of the screen by the intermediate and projection lenses with an increase of several hundred times. Thus, by changing the current of the intermediate lens, it is possible to translate the microscope from the mode of microscopic image observation to the mode of observation of the diffraction pattern (microdiffraction electron pattern) and vice versa.

In the formation of the final diffraction pattern on the screen, only the rays that correspond to the region of the primary microscopic image limited by the selector diaphragm participate, in other words, only the rays scattered from that part of the object limited by the image of the aperture of the selector diaphragm in the plane of the object contribute to this image. Varying the position of the selector diaphragm or moving the sample, it is possible to obtain diffraction patterns from different regions. The dimensions of these regions are determined by the physical size of the selector diaphragm and the focus length of the objective lens, i.e. by the size of image of the selector aperture in the plane of the object. This size is equal to the size of the selector diaphragm (usually about 0.1 mm) divided by the lens magnification (about 100) and corresponds to approximately one micrometer.

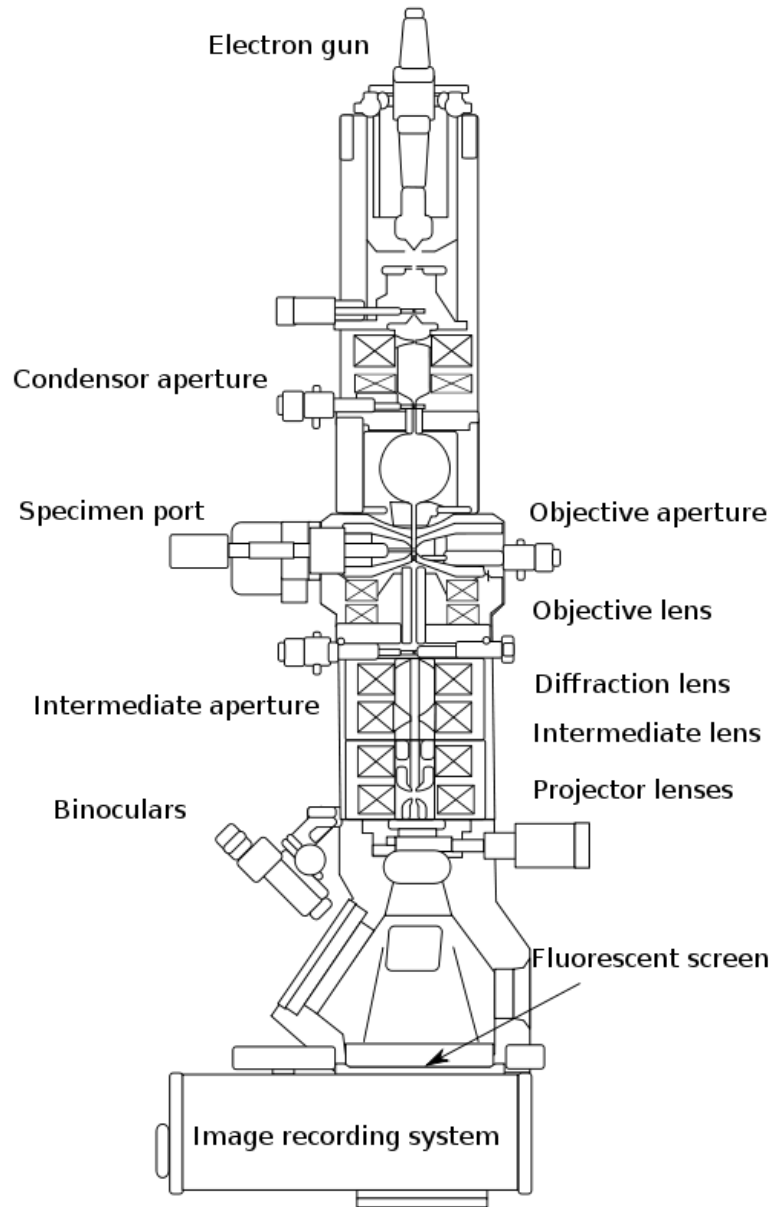


Figure 4.105. Principal optical scheme of transmission electron microscope

When considering the diffraction of electrons in a transmission electron microscope, it should be taken into account that the deflection angle of the diffracted electron beam  $2\Theta$  is inversely proportional to the corresponding interplanar distance of the crystal. Indeed, the de Broglie wavelength for an electron with an energy of 100 keV is equal to  $3.9 \cdot 10^{-12}$  m, which is much smaller than the typical values of interplanar distances. Therefore, it follows from the Wolf-Bragg formula (1) that

$$2\Theta = \frac{k\lambda}{d} . \quad (4.110)$$

It is easy to understand that for polycrystalline objects the diffraction pattern will be represent a set of concentric circles (rings), each of which corresponds to diffraction on a certain family of planes. In this case, the diameters of the rings are inversely proportional to the corresponding interplanar distances. To determine the absolute values of interplanar distances, one use a reference sample – a crystal of a known phase, for example, a thin aluminum foil.

As an example, Figure 4.106. shows the electron diffraction patterns of polycrystalline indium oxide film before and after the formation of cadmium sulfide nanoparticles on the surface. The small size of crystallites of indium oxide (about 50 nm) and cadmium sulfide nanoparticles (several nanometers) leads to the fact that the electron diffraction pattern is recorded from a region containing a huge amount of crystallites. Therefore, the diffraction pattern is set of homogeneous axially symmetric rings. It is possible to clearly distinguish the rings corresponding to the  $\text{In}_2\text{O}_3$  phase and rings corresponding to hexagonal CdS.

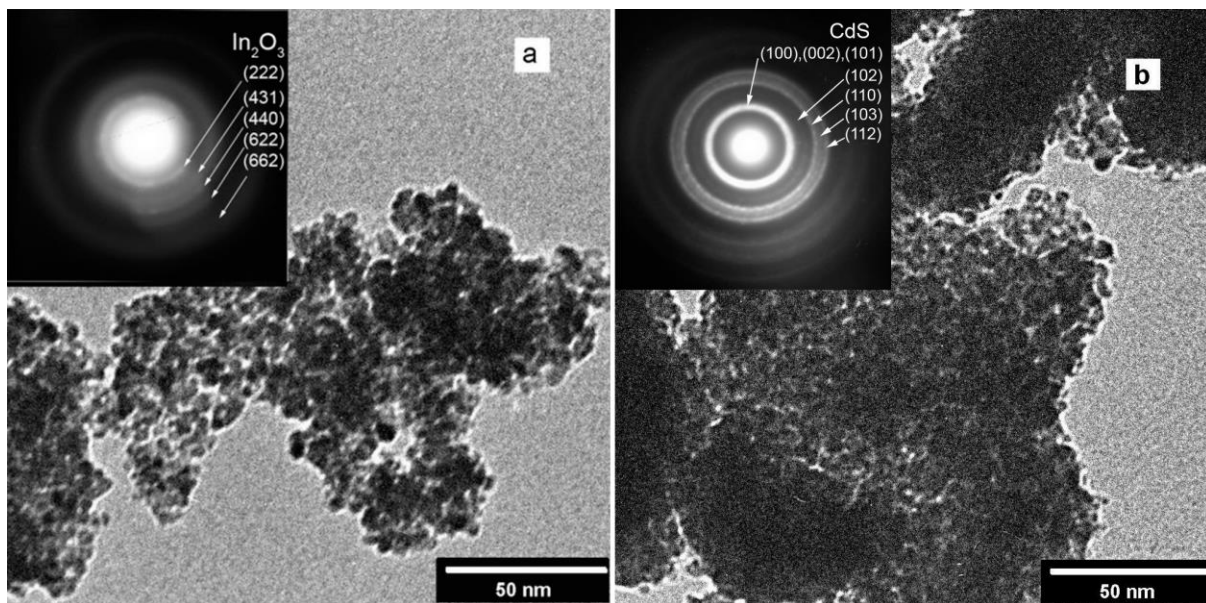


Figure 4.106. Electron diffraction pattern of polycrystalline indium oxide [4]

If the electron beam is diffracted on a single crystalline object, the diffraction pattern represents a set of individual points (Fig. 4.106.). Each such point obviously corresponds to the nod of the reciprocal lattice of the crystal. If the Miller indices are known for a plane perpendicular to the incident ray, it is easy to determine the corresponding interplanar distances from the position of the reflections.

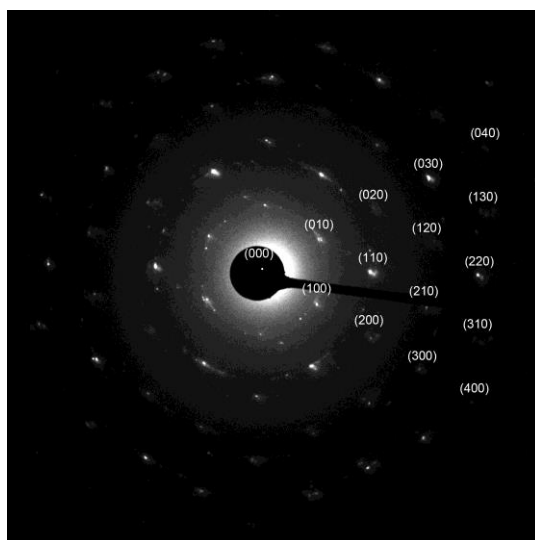


Figure 4.107. Electron diffraction pattern for a polycrystalline BiOS film recorded from a single crystalline region [5]

#### 4.8.7 Low energy electron diffraction

It is necessary very often in research practice to determine the structure of thin (in the limit – single layer) near-surface regions. Obviously, the use of the above considered X-ray diffraction methods for this purpose is impossible, since the contribution of one monolayer to the formed diffraction pattern will be negligibly small. Therefore, to solve the problems of determining the atomic structure of the surface, the method of low energy electron diffraction (LEED) is widely used.

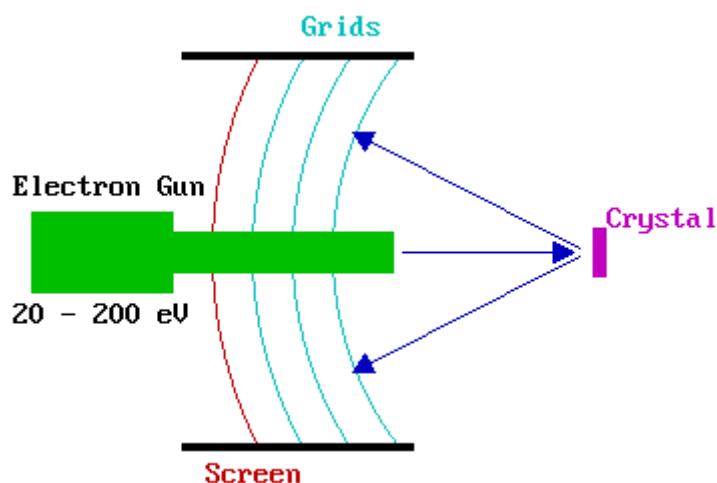


Figure 4.108. Experimental scheme of LEED method

The scheme of the method implementation is presented in Fig. 4.108. The source directs on the investigated crystal a beam of monoenergetic electrons, which are scattered on it both elastically and inelastically. Since only elastically scattered electrons contribute to the formation of the diffraction pattern, it becomes necessary to cut off electrons that are inelastically scattered on the

surface, as well as electrons scattered at some depth and lost energy as they pass through the object under study. For this, grids are mounted between the crystal under investigation and the electron detector, which stops the electrons with an energy different from the energy of the electrons of the primary beam.

Consider the formation of a diffraction pattern. As an example, let us consider diffraction by an atomic layer with mutual atom arrangement corresponding to the (110) plane of the face centered cubic structure. Schematically, the arrangement of the atoms is shown in Figure 4.108. (left panel). To calculate the resulting diffraction pattern, it is sufficient to use the standard condition for the interference amplification of electrons scattered by different atoms. Doing this, it is easy to get the diffraction pattern shown in the right panel of Fig. 4.109.

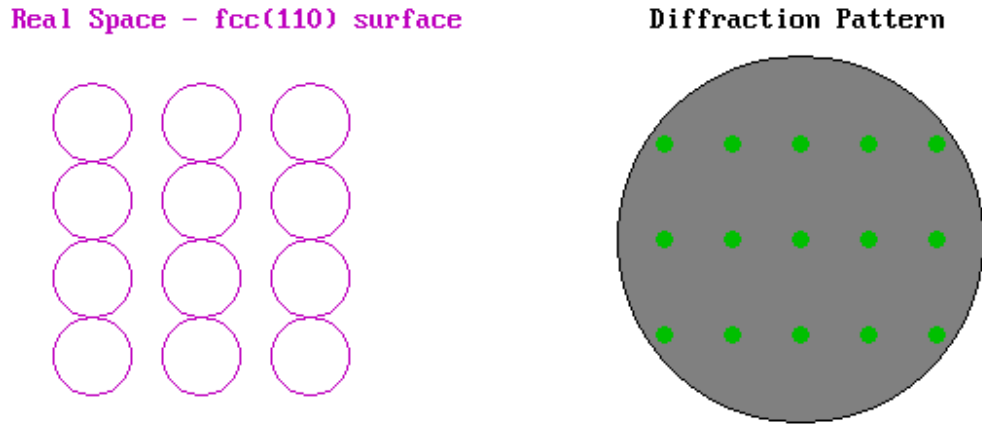


Figure 4.109. The diffraction pattern for electron diffraction at two-dimensional lattice with atom arrangement corresponding to the (110) plane of face centered cubic structure

As in the case of diffraction by 3D structures, the use of the concept of the reciprocal lattice is very convenient and useful for describing diffraction on two-dimensional objects. For two-dimensional structures, the inverse lattice is constructed as follows. If  $\mathbf{a}_1$  and  $\mathbf{a}_2$  are basis vectors of a direct lattice, then the reciprocal lattice vectors  $\mathbf{a}_1^*$  and  $\mathbf{a}_2^*$  are determined by the relations

$$\mathbf{a}_1 \cdot \mathbf{a}_2^* = \mathbf{a}_2 \cdot \mathbf{a}_1^* = 0, \quad (4.111)$$

$$\mathbf{a}_1 \cdot \mathbf{a}_1^* = \mathbf{a}_2 \cdot \mathbf{a}_2^* = 1. \quad (4.112)$$

It is easy to see that the generated diffraction pattern is a scaled image of the reciprocal lattice. This is illustrated in Figures 4.110. and 4.111.



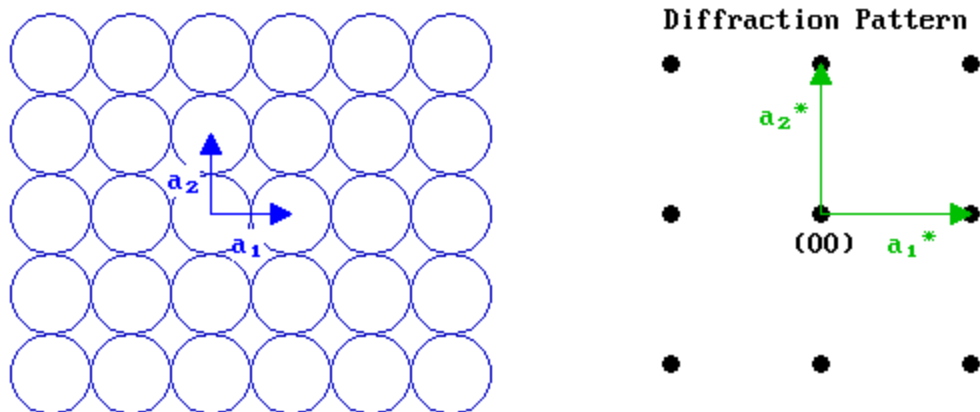


Figure 4.110. Diffraction pattern in the case of diffraction on a two-dimensional cubic lattice

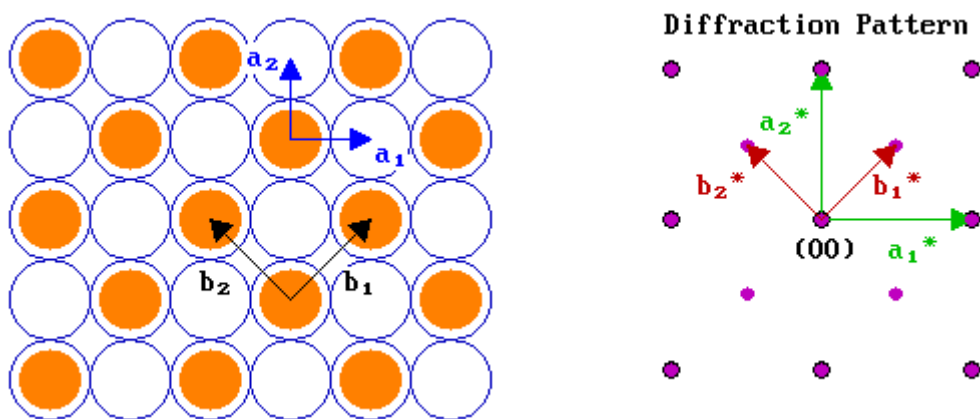


Figure 4.111. Diffraction pattern in the case of diffraction on a two-dimensional cubic lattice with a surface  $c(2 \times 2)$  structure

Thus, the simplest application of the low energy electron diffraction is the revealing the type, symmetry, and interatomic distances for surface structures. However, with the help of this method it is possible to solve even more complicated problems. For example, from broadening of the reflections, one can determine the size of the regions of coherent scattering, the background level reflects the concentration of point defects of surface structures.

#### 4.8.8 References

1. Eu modified  $\text{Cu}_2\text{O}$  thin films: significant enhancement in efficiency of photoelectrochemical processes through suppression of charge Carrier recombination / S. Shyamal, P. Hajra, H. Mandal, A. Bera, D. Sariket, A.K. Satpati, M.V. Malashchonak, A.V. Mazanik, O.V. Korolik, A.I. Kulak, E.V. Skorb, Ajun Maity, E.A. Streltsov, C. Bhattacharya // Chemical Engineering Journal. – 2018. – Vol. 335. – P. 676 – 684.
2. Layer-type BiOBr photoelectrodes / M.E. Kozyrevich, M.V. Malashchonak, E.A. Streltsov A.V. Mazanik // to be submitted.

3. Microstructural properties and enhanced photocatalytic performance of Zn doped CeO<sub>2</sub> nanocrystals / M. A. Majeed Khan, Wasi Khan, Maqusood Ahamed, Abdulaziz N. Alhaza // Scientific Reports 7, Article number: 12560 (2017).
4. Photoelectrochemical and Raman characterization of In<sub>2</sub>O<sub>3</sub> mesoporous films sensitized by CdS nanoparticles / N.V. Malashchonak, S.K. Poznyak, E.A. Streltsov, A.I. Kulak, O.V. Korolik, A.V. Mazanik // Beilstein Journal of Nanotechnology. – 2013. – Vol. 4. – P. 255 – 261.
5. Giant Incident Photon-to-Current Conversion with Photoconductivity Gain on Nanostructured Bismuth Oxysulfide Photoelectrodes under Visible-Light Illumination / E.A. Bondarenko, E.A. Streltsov, M.V. Malashchonak, A.V. Mazanik, A.I. Kulak, E.V. Skorb // Adv. Mater. – 2017, DOI: 10.1002/adma.201702387.

## **5 EMC, EMI AND RELIABILITY**

## 5.1 Introduction to EMC

### 5.1.1 Basic concepts

#### 5.1.1.1 EMC and EMI

EMC is an abbreviation for Electromagnetic Compatibility. Two devices are electromagnetically compatible when the operation of the first device does not have an impact on the operation of the second device and vice versa.

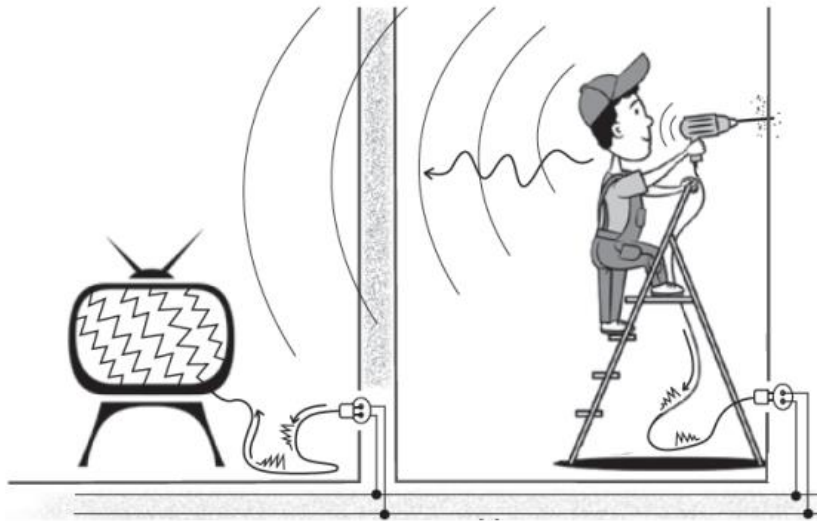


Figure 5.1. Electromagnetic Interference EMI

Electromagnetic interference (abbreviated as EMI) occurs when two devices are operating in each others proximity and when the operation of one device has an effect on the operation of the other device (see Figure 5.1 adapted from source: <http://www.electronic.nu/en/2015/06/10/what-is-emiemc>). Examples of EMI are:

- The use of a mobile phone (GSM: Global System for Mobile Communications) which disturbs the operation of the airbag, the central locking system of a car or the car radio.
- The use of the mobile phone which disturbs the operation of a pacemaker.
- The use of a mobile phone might disturb the electronic devices in the airplane. Due to this reason, the use of mobile phones (and other electronic equipment) is forbidden during the takeoff and the landing of the airplane.
- The use of a mobile phone might disturb the operation of several devices in a hospital. Due to this reason it is often forbidden or discouraged to use a mobile phone in the hospital (source of Fig. 5.2: <http://www.rachaelcampbell.com/blog/its-time-to-switch-your-phone-off/>).



Figure 5.2. Forbidden to use a mobile phone

#### 5.1.1.2 The CE marking

On a lot of electronic devices, the CE marking is available and this marking is required to sell a product in the European Economic Area (CE = Conformité Européenne = European Conformity) . The CE marking can also be found on a lot of products which are sold outside the European Economic Area because they are manufactured in the European Economic Area or because they are intended to be sold in the European Economic Area.

The CE marking is a declaration of the manufacturer that the device meets the requirements of the EC directives. More precisely, the device satisfies:

- The machinery safety directive MSD
  - The machinery safety directive describes a number of safety criteria for devices. The main goal is to reduce safety risks as much as possible. When the reduction of these risks is not possible or limited, additional protection of the user is required or additional warnings to the user are needed. For instance, rotating parts can't be touched by the user. If rotating parts can still be touched, warning signs are used. For instance when sawdust is released, this sawdust will be sucked up to protect the health of the users.
- The low voltage directive
  - The low voltage directive describes electrical safety guidelines in case of voltage levels between 50 V and 1000 V which must be respected by designers. For instance, due to appropriate covering, no danger for electrocution occurs when the device is used in a correct way.
- The EMC-directives

The manufacturer of the device is allowed to control whether the device satisfies the directives and is allowed to fix the CE label himself. This means the manufacturer is responsible when accidents occur due to EMI problems associated with the device.

#### 5.1.2 The EMC directives

##### 5.1.2.1 EMC

“You will not disturb and you will not be disturbed” is the main principle of the EMC directives. EMC is often defined as (International Electrotechnical Vocabulary):

*The ability of a device, equipment or system to function satisfactory in its electromagnetic environment without introducing intolerable electromagnetic disturbance to anything in that environment.*

Due to technical evolutions, it becomes harder to satisfy this EMC philosophy. In a lot of devices, the number of electronic parts has increased steadily. This evolution will continue in the future. Figure 5.3. visualises a car which contains a large amount of electronica parts (including sensors). Each electronic part (possibly) causes an emission of disturbances and care must be taken in order to prevent failure of other parts (source of Figure 5.3: <http://www.sensorsmag.com/test-measurement/test-considerations-automotive-sensors-14929>).

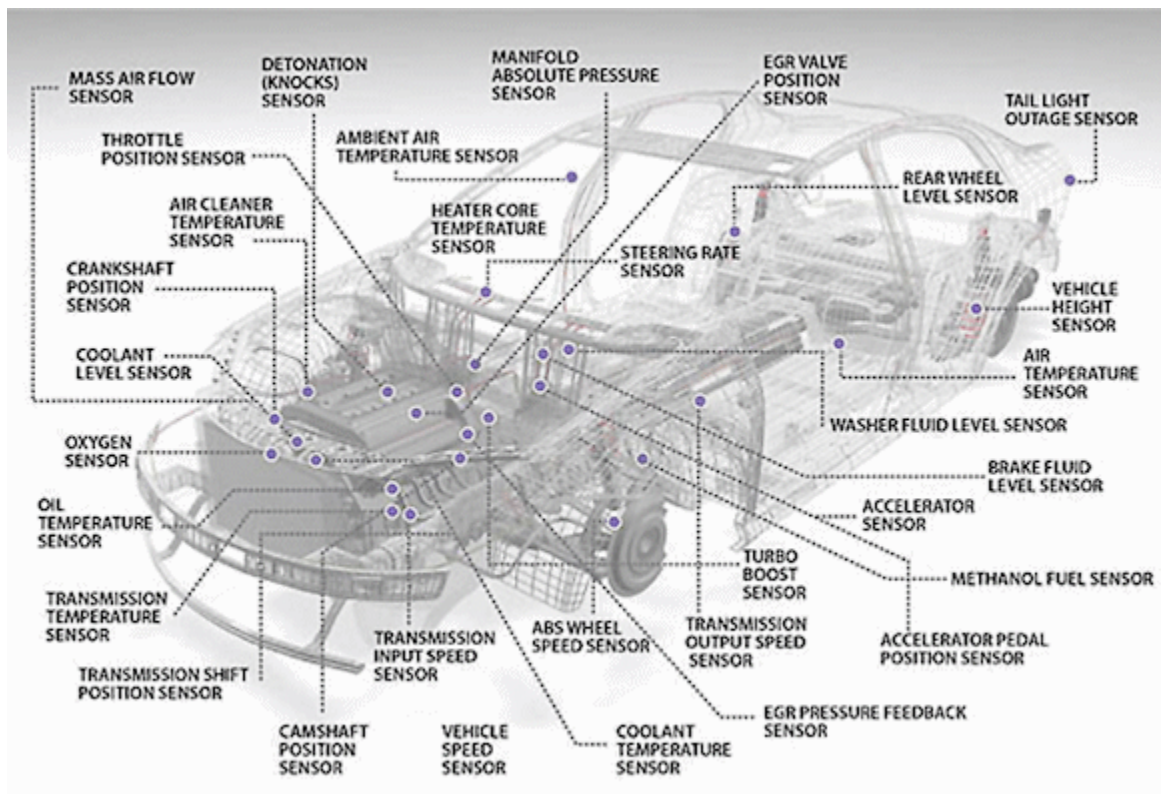


Figure 5.3. Car containing a lot of electronic parts/sensors

Due to technological evolutions, people more and more depend on electronic applications to guarantee their safety. Figure 5.4 visualises an “automated guided vehicle” (AGV) which realises the transport of materials in an industrial factory (source of Figure 5.4: [https://en.wikipedia.org/wiki/Automated\\_guided\\_vehicle](https://en.wikipedia.org/wiki/Automated_guided_vehicle)). When disturbing the operation of such an AGV, dangerous situations might occur. Figure 5.5 visualises an autonomous car (driverless car; self-driving car); also here it is important the electronic parts are not disturbed (source of Figure 5.5: <http://www.clubexus.com/forums/car-chat/767628-self-driving-vehicles-news-thread.html>).





Figure 5.4. Automated Guided Vehicle



Figure 5.5. Driverless car

#### 5.1.2.2 Emission and immunity

More and more, electronic devices intend to emit disturbances. The emission level increases. For instance the traditional incandescent lamp only accounts for a very limited emission of disturbances. Notice however the rise of power electronics, clocked processors and also mobile phones which account for much more electromagnetic emissions.

At the other hand side, electronic devices are more and more susceptible to disturbances i.e. their immunity level has decreased. Indeed, the transition from vacuum tubes to semiconductor transistors, the transition from 5V technology to 3V technology imply electronic devices are more vulnerable to disturbances.

Figure 5.6 visualises, during the last century, the natural increase of the emission level. Fig. 5.6. also visualises, during the last century, the natural decrease of the immunity level. In case the emission level exceeds the immunity level, one device will disturb the proper operation of another device.

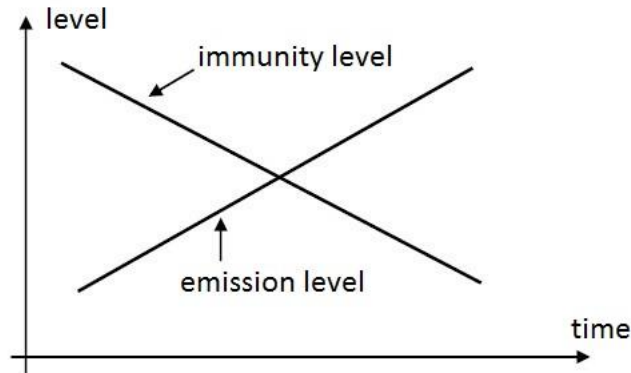


Figure 5.6. Natural evolution of emission and immunity levels

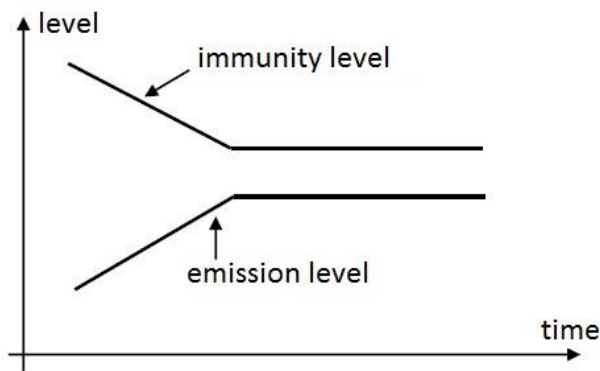


Figure 5.7. EMC-directives limit emission and immunity levels

The evolution visualised in Figure 5.6 illustrates that the emission level increases as a function of time and that the immunity level decreases as a function of time. At a certain moment of time, both curves intersect and (undesired) EMI occurs. The EMC-directives intend to avoid that both curves in Fig. 5.6. intersect by

- keeping the emission level sufficiently low,
- keeping the immunity level sufficiently high.

The EMC-directives intend to realise the situation visualised in Figure 5.7. The emission level is not allowed to exceed the immunity level.

Until now, we mainly emphasized that a first device is not allowed to disturb the proper operation of a second device. This is correct, but a device is also not allowed to disturb its own operation. Although this ‘intra-system EMC’ has not been regulated by EMC-directives, it is also very important.



### 5.1.3 Emission

In this paragraph, we will discuss the most important types of disturbances which can be emitted by an electronic device. When talking about emission, it is important to make a distinction between conducted emission and radiated emission.

#### 5.1.3.1 Conducted emission

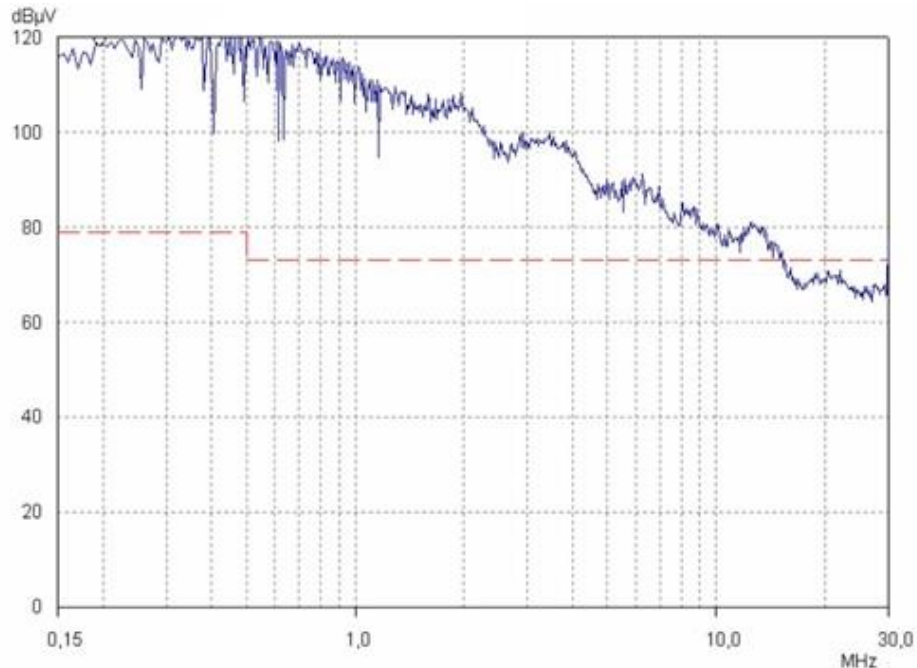


Figure 5.8: Conducted emission measurement

An electronic device can emit disturbances using the conductors of the electrical power grid. It is possible to make a distinction between

- harmonics (low frequent, multiples of the 50 Hz grid frequency),
- high frequent disturbances having frequencies higher than 9 kHz (there exist directives in the frequency range from 150 kHz to 30 MHz),
- changes in the voltage level of the power grid and flicker.

Here, we will mainly discuss conducted emission of high frequent disturbances in the frequency range from 150 kHz to 30 MHz. Figure 5.8 visualises an emission level measurement originating from a frequency converter. The measurement device performs a frequency sweep between 150 kHz and 30 MHz. Due to the high frequent currents of the frequency converter there are also high frequent voltages and these high frequent voltages are measured.

Since the voltage levels vary over a very broad amplitude range, Figure 5.8 uses a logarithmic scale for the vertical axis (the horizontal frequency axis also uses a logarithmic scale). The measured voltage  $U$  is compared with a reference voltage which equals  $1 \mu V$  implying the voltage level is expressed in  $dB\mu V$  by using the expression

$$20 \log \left( \frac{U}{1 \mu V} \right) \quad (5.1)$$

The maximum allowed emission level in Fig. 5.8. is indicated in red and this emission limit is valid in an industrial environment (class A, group 1). Between 150 kHz and 500 kHz, an emission level of 79 dB $\mu$ V is allowed (QP-measurement) and between 500 kHz and 30 MHz an emission level of 73 dB $\mu$ V is allowed (QP-measurement). The measured emission level in Fig. 5.8. is indicated in blue and this emission level is much higher than the allowed emission level. Actions are needed to reduce the emission level. For instance, this can be done by mounting an EMC filter between the frequency converter and the electrical power grid.

### 5.1.3.2 Radiated emission

The conductors and other components behave as antennas causing radiated emission in the environment (the atmosphere, the ether). Directives are available in the frequency range of 150 kHz to 1000 MHz. In order to measure the intensity of these electromagnetic waves, in the frequency range from 150 kHz to 30 MHz mainly the magnetic field strength is measured. In the frequency range from 30 MHz to 1000 MHz mainly the electrical field strength is measured.

When measuring the radiated emission level of an electronic device, it is important the radiated emission level of only this device has been measured. The measurement is not allowed to be influenced by the emissions of other devices (also not by the emission of for instance radio broadcasting stations). Therefore, such a measurement is performed in a cage of Faraday i.e. in an electromagnetically shielded room.

Figure 5.9. visualises such a cage of Faraday. The electronic device (Equipment Under Test = EUT; Device Under Test = DUT) is situated on a wooden table. Such a wooden table is used since wood does not reflect electromagnetic waves. The receiving antenna is placed in the cage of Faraday and a distance of 3 meter is available between the DUT and the antenna. The receiving antenna measures the emission levels and these measurement data are sent to an EMI receiver outside the cage of Faraday. Using a PC, these measurement data are visualised which allows an interpretation of the results.

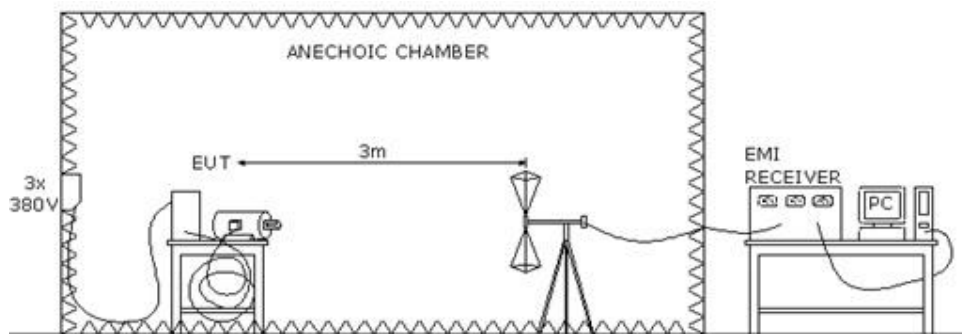


Figure 5.9. Radiated emission measurement in a cage of Faraday

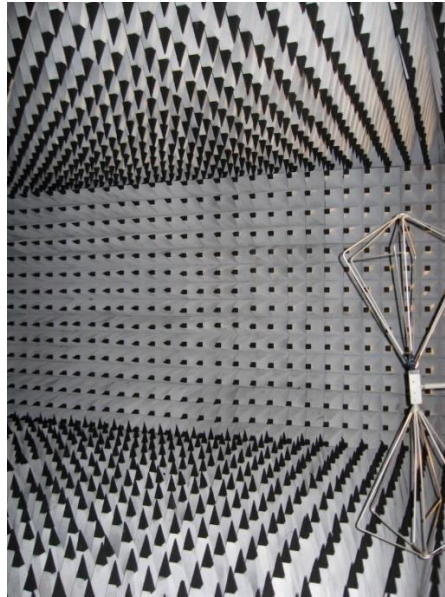


Figure 5.10. Cage of Faraday covered with cones inside

Notice in Figure 5.9 that the walls, the floor and the ceiling of the cage are covered with cones (see also Figure 5.10). These cones absorb the electromagnetic waves avoiding reflections. In this way, an open electromagnetic space is simulated. Figure 5.10 visualises the cage of Faraday from the inside containing these absorbing cones. Notice also the biconical receiving antenna.

Figure 5.11. visualises the inside of another cage of Faraday. Here, the walls are not covered with cones. This implies reflections occur when electromagnetic waves reach the walls. Notice also the biconical antenna in Fig. 5.11. Notice the biconical antenna is mounted on a wooden support (due to the wood, no electromagnetic waves are reflected).



Figure 5.11. The inside of a cage of Faraday without cones on the walls

Figure 5.12 visualises a radiated emission measurement which has been obtained in the cage of Faraday visualised in Figure 5.10 using the measurement setup of Figure 5.9. The measurements

are performed in a frequency range between 30 MHz to 200 MHz (using a logarithmic frequency axis). The measured electrical field strengths are expressed in  $V/m$ . Also for the vertical axis, a logarithmic scale has been used. This logarithmic scale is used because of the large range of measured field strengths. The measured field strength  $E$  is compared with a reference of  $1 \mu V/m$ . This implies the field strength is expressed in  $dB\mu V/m$  using the expression

$$20 \log \left( \frac{E}{1 \mu V/m} \right) \quad (5.2)$$

Figure 5.12. visualises the measurements (using blue) in case the cage does not contain a DUT causing radiation (an emission level is never zero). In green, the emission level of an operating frequency converter (VSD = Variable Speed Drive) is visualised. Notice the frequency converter causes a quite high emission level in comparison with the background noise. Indeed, a frequency converter contains fast switching IGBTs which are able to switch quite high voltages. The three phase cable between the inverter output and the induction motor causes the emission of electromagnetic waves.

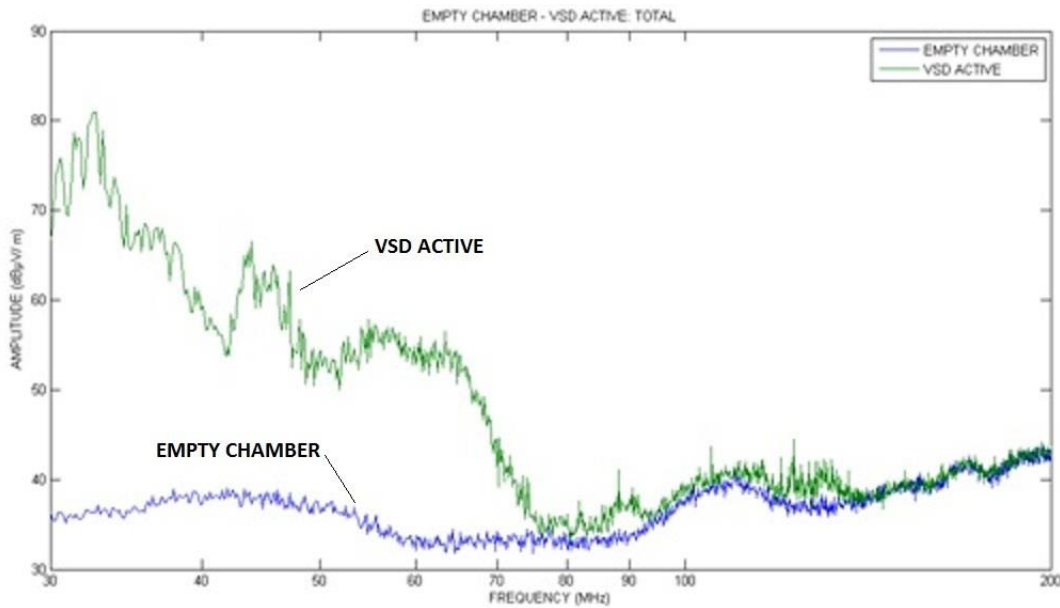


Figure 5.12. Radiated emission measurement

#### 5.1.4 Immunity

When considering immunity, it is also possible to make a distinction between disturbances which propagate using the electrical grid and disturbances which propagate using the atmosphere.

The disturbances which occur in the electrical power grid are for instance: voltage harmonics, fast transient/burst, surge, voltage dips.

The disturbances which propagate in the atmosphere

- elektromagnetic waves (e.g. 80 MHz to 1000 MHz)

- magnetic fields

Notice also immunity against an Electro-Static Discharge ESD is important.

#### 5.1.4.1 The importance of immunity

Electromagnetic immunity indicates to what extent an electronic device is able to withstand the influence of disturbances. In case the proper operation of the device is disturbed, there are a number of possibilities. Possibly, the proper operation is temporarily disturbed and the device behaves normal once the disturbances have disappeared.

The situation can be worse. Possibly, the device will not behave normal once the disturbances have disappeared. The device restarts to behave normal after resetting or switching off and on the device. The user must interact to restore the normal behaviour.

The worst situation occurs when the disturbances damage the device permanently i.e. a normal reset or switching the device off and on does not solve the problem. In such a situation, the device must be repaired or replaced by another one.

Another criterion is the question to what extent the undesired behaviour of the device harms the safety of people (e.g. consider medical devices in a hospital). Does the EMC related problem cause material damage and what is the importance of this damage?

#### 5.1.5 References

1. H. Balzer, P. Fischer, M. Lutz, J. Kappert, EMC: Immunitiestesten en meetmethoden, De Muiderkring, Weesp, 1993.
2. J. J. Goedbloed, Elektromagnetische compatibiliteit, analyse en onderdrukking van stoorproblemen, Kluwer Technische Boeken B.V., Deventer – Antwerpen, 1990.
3. J. Peuteman, T. Verbeerst, P. Vansielegheem, J. Knockaert, D. Pissort, J.-J. Vandenbussche, Reducing Electromagnetic Emitted Disturbances of an Adjustable Speed Drive System, Annual Journal of Electronics, Volume 6, Number 1: ISSN 1314-0078, pp. 116-119 (paper presented at the XXIst International Scientific and Applied Conference: Electronics – ET 2012, Sozopol, Bulgaria, September 19-21, 2012).
4. L. Tihanyi, Electromagnetic Compatibility in Power Electronics, IEEE Press, J.K. Eckert & Company Inc., Butterworth-Heinemann Ltd., Oxford, 1995.

## 5.2 Conducted emission measurements

### 5.2.1 Introduction

An electronic device can emit disturbances using the conductors of the electrical power grid. The present text will focus on conducted emission of high frequent disturbances in the frequency range from 150 kHz to 30 MHz.

#### 5.2.1.1 Frequency converters

In an industrial environment, frequency converters are an important source of such high frequent conducted emission. A frequency converter (also called a variable speed drive VSD) is a power electronic device which converts the fixed voltage of the three phase AC power grid (e.g. having a line voltage of 400 V and a frequency of 50 Hz) to a voltage having an adjustable amplitude and an adjustable frequency. When this adjustable output voltage is used to supply a three phase induction motor, a smooth adjustment of the speed of this induction motor is possible.



Figure 5.13. Frequency converter feeding an induction motor

The working principle of a frequency converter is visualised in Figure 5.14. On the left, the sinusoidal grid voltage (having a fixed amplitude and a fixed frequency) is rectified giving a constant DC voltage. Using an inverter (i.e. a DC/AC converter), the DC voltage is converted into a PWM signal (PWM = Pulse Width Modulation) which realises an AC voltage having an adjustable amplitude and an adjustable frequency.

The inverter contains semiconductor switches (e.g. IGBTs = Insulated Gate Bipolar Transistors) which allow to realise voltage pulses having an adjustable width. By changing the widths of these pulses in a proper way, the PWM voltage in Figure 5.15 is a good approximation of a sinusoidal voltage. The voltage in Figure 5.15 is a so-called bus voltage which is available between the nodes 0 and A in Figure 5.14 (also between nodes 0 and B and between nodes 0 and C).

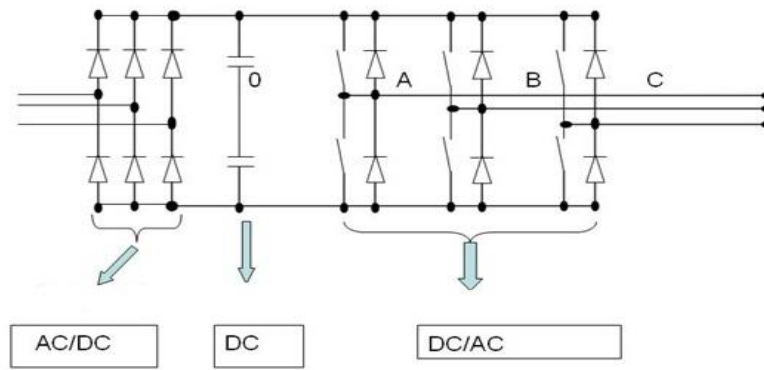


Figure 5.14. Working principle of a frequency converter (source: Deconinck et al.)

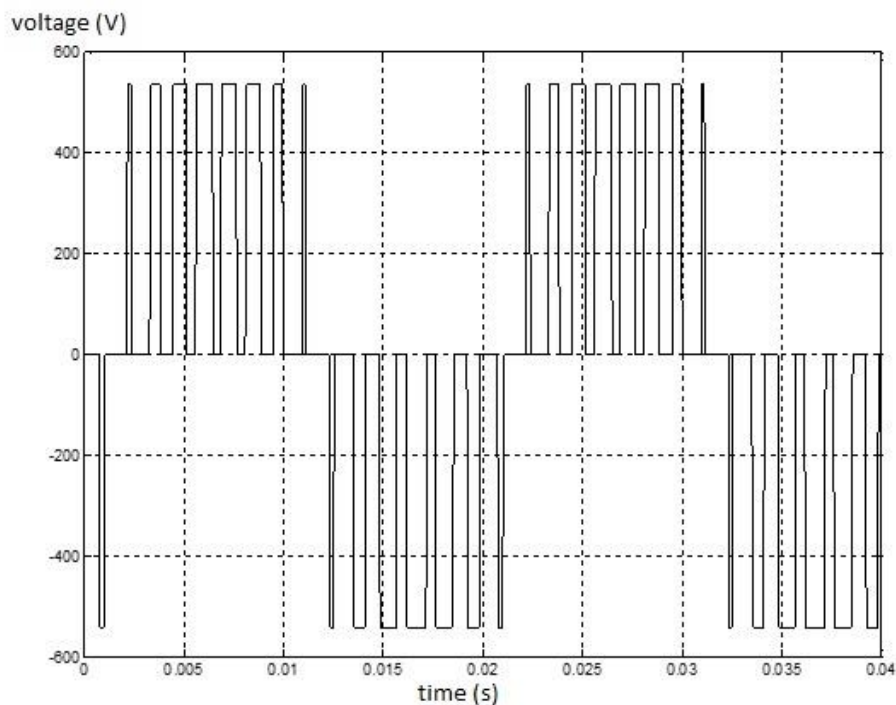


Figure 5.15. PWM output voltage (source: Deconinck et al.)

In general, the IGBT's are switched on and off using mainly switching frequencies of a few kHz (e.g. between 1 kHz and 20 kHz). In reality, the IGBTs switch very fast in order to limit the energy losses in the inverter. When considering a single PWM pulse in Figure 5.15, rise times and fall times between 80 ns and 200 ns are very realistic. Due to these small rise times and fall times, the PWM voltage in Figure 5.15 contains a lot of high frequent components (also in the range from 150 kHz to 30 MHz). Indeed, using Fourier analysis it is possible to calculate the frequency spectrum of the practical time signal visualised in Figure 5.15.



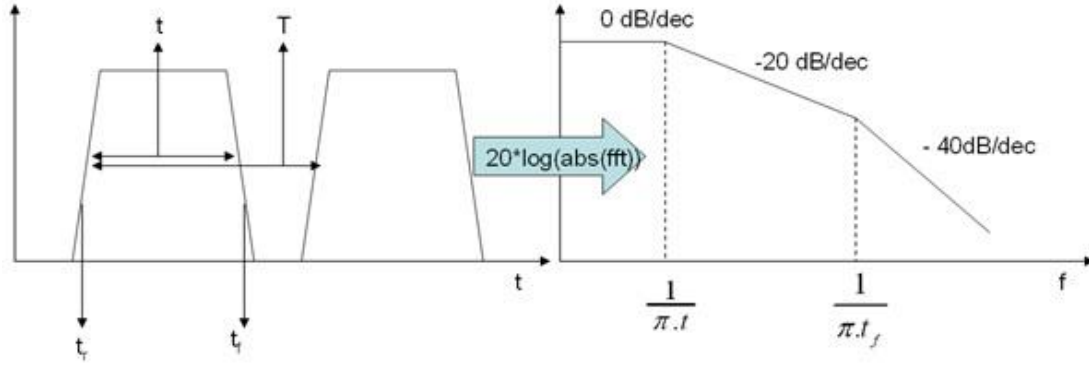


Figure 5.16. Fourier analysis of the PWM voltage (source: Deconinck et al.)

On the left, Figure 5.16 visualises two PWM voltage pulses each having a rise time  $t_r$  and a fall time  $t_f$  (the single pulse has a width  $t$  and there is a period  $T$ ). Very often,  $t_f \cong t_r$ . On the right, not the frequency spectrum but its envelope has been visualised. The higher the frequency, the smaller the frequency components.

When considering frequency components between  $1/(\pi t)$  and  $1/(\pi t_f)$ , each decade the envelope decreases with  $20 \text{ dB}$ . This means that when the frequency is 10 times larger, the amplitude is 10 times smaller. When considering frequency components higher than  $1/\pi t_f$ , each decade the envelope decreases with  $40 \text{ dB}$ . This means that when the frequency is 10 times larger, the amplitude is 100 times smaller.

In case of rather large  $t_f$  values, the importance of the high frequent components is limited. However, when using fast switching IGBTs small  $t_f \cong t_r$  values imply the presence of rather important high frequent voltage components in the spectrum. For instance in case  $t_f = t_r = 100 \text{ ns}$ , the corner frequency  $1/(\pi t_f) \cong 3 \text{ MHz}$ .

Especially due to these high frequent voltage compontents in the PWM voltage, the frequency converter is an important source of high frequent currents in the grid conductors.

#### 5.2.1.2 High frequent voltages

The electrical power grid mainly behaves as a Thévenin equivalent circuit as visualised in Figure 5.17. Due to the high frequent currents, high frequent voltages appear across the grid impedance which cause high frequent voltages in the grid voltage due to the voltage law of Kirchoff.

Suppose in Figure 5.17 the original grid voltage  $u_{f1}(t)$  is a sine voltage having a frequency of  $50 \text{ Hz}$  (in reality the voltage grid is often a three phase grid, but for simplicity consider a single phase grid). The current consumed by the frequency converter equals  $i_{f1}(t)$ . In an ideal situation, this current  $i_{f1}(t)$  is a sine but in reality this current will also contain high frequent components between  $150 \text{ KHz}$  and  $30 \text{ MHz}$ . This implies also high frequent voltage components across the



ohmic-inductive grid impedance in the frequency range from 150 kHz to 30 MHz. Using the voltage law of Kirchoff, this implies voltage components in the frequency range from 150 kHz to 30 MHz in the grid voltage  $u_{grid}(t)$ .

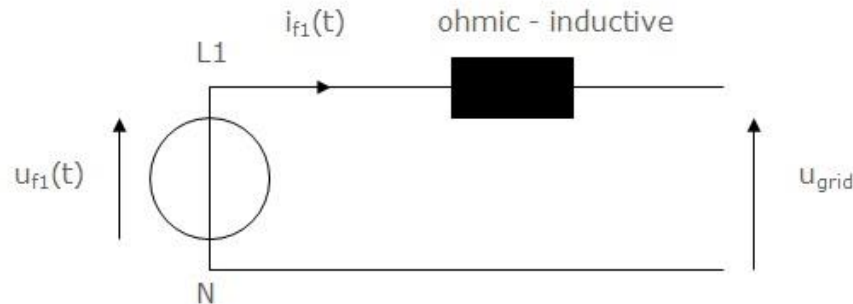


Figure 5.17. Thevenin equivalent circuit of the power grid

## 5.2.2 Conducted emission measurement

### 5.2.2.1 Measurement setup

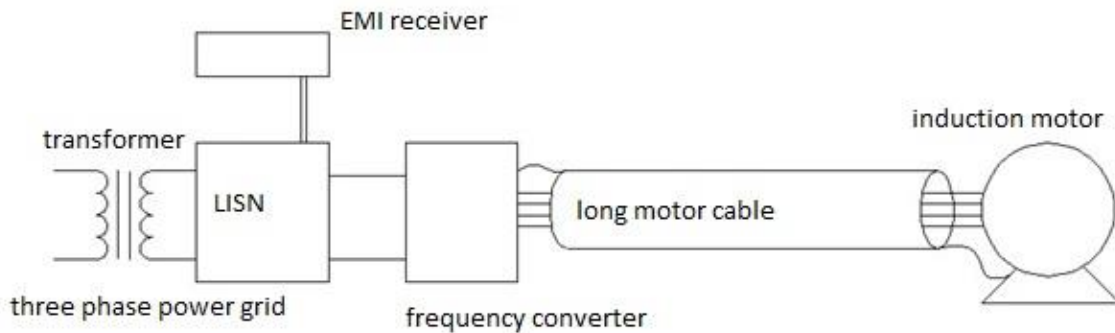


Figure 5.18. Measurement setup of a conducted emission measurement

Figure 5.18. visualises the measurement setup needed to measure the high frequent conducted emission levels of an electronical device (the DUT = Device Under Test). Notice the frequency converter where the PWM voltage supplies a long motor cable and an induction motor. Motor cables, having a length of for instance 30 m, occur in case the induction motor is operating in a harsh environment and the frequency converter is situated in another room having more moderate temperatures and environmental conditions in general. The entire drive system, i.e. the frequency converter, the motor cable and the induction motor, is the DUT.

The measurement setup also contains an EMI receiver (on the left in Fig. 5.19.) and a LISN (LISN = Line Impedance Stabilisation Network) (on the right in Fig. 5.19). The transformer between the power grid and the LISN avoids the residual-current circuit breaker to open. But especially this

LISN is an important device. Due to this LISN, the high frequent grid impedance (see Fig. 5.17) is known in the frequency range from 150 kHz to 30 MHz.



Figure 5.19. EMI receiver and LISN

#### 5.2.2.2 Line Impedance Stabilisation Network

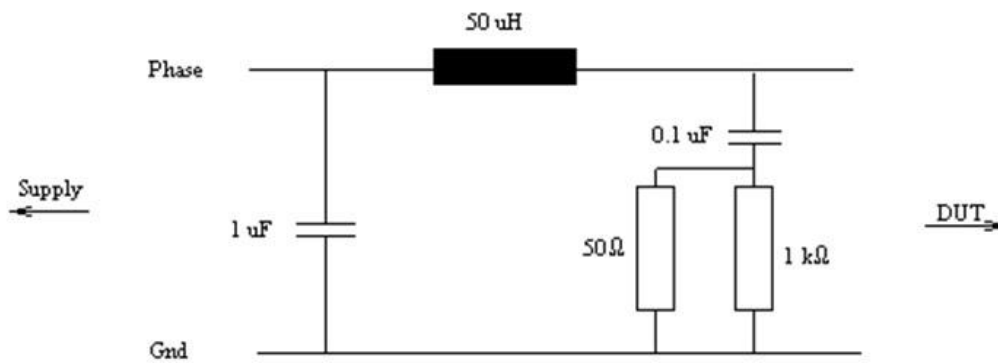


Figure 5.20. Internal behaviour of a LISN (source: Deconinck et al.)

Fig. 5.20 visualises the internal behaviour of the LISN. There exist single phase and there exist three phase LISNs but due to symmetry reasons also a three phase LISN can be modeled as a single phase system. On the left, the 50 Hz voltage of the power grid is applied which allows to supply the DUT (Device Under Test) on the right. For the considered frequency range between 150 kHz and 30 MHz, the impedances of the capacitors are small implying a high frequent grid impedance of  $50 \mu H$  in parallel with  $1 k\Omega$  and  $50 \Omega$ . This  $50 \Omega$  impedance is the impedance of the EMI-receiver.

Due to this normalised high frequent grid impedance, the same DUT accounts for the same high frequent currents and for the same high frequent voltages. Even when performing measurements at different locations and at different moments of time, the same measurement results are obtained. Since the  $50 \Omega$  resistor in Figure 5.20 models the EMI-receiver, the EMI-receiver measures the voltage across the  $1 k\Omega$  resistor and the  $50 \mu H$  inductor.

### 5.2.2.3 Conducted emission measurement

Figure 5.21 visualises a conducted emission measurement using the measurement setup of Figure 5.18 where the DUT is a frequency converter feeding a long motor cable and an induction motor. The measurements are performed in a frequency range from 150 kHz to 30 MHz and the voltages are expressed in  $\text{dB}\mu\text{V}$ .

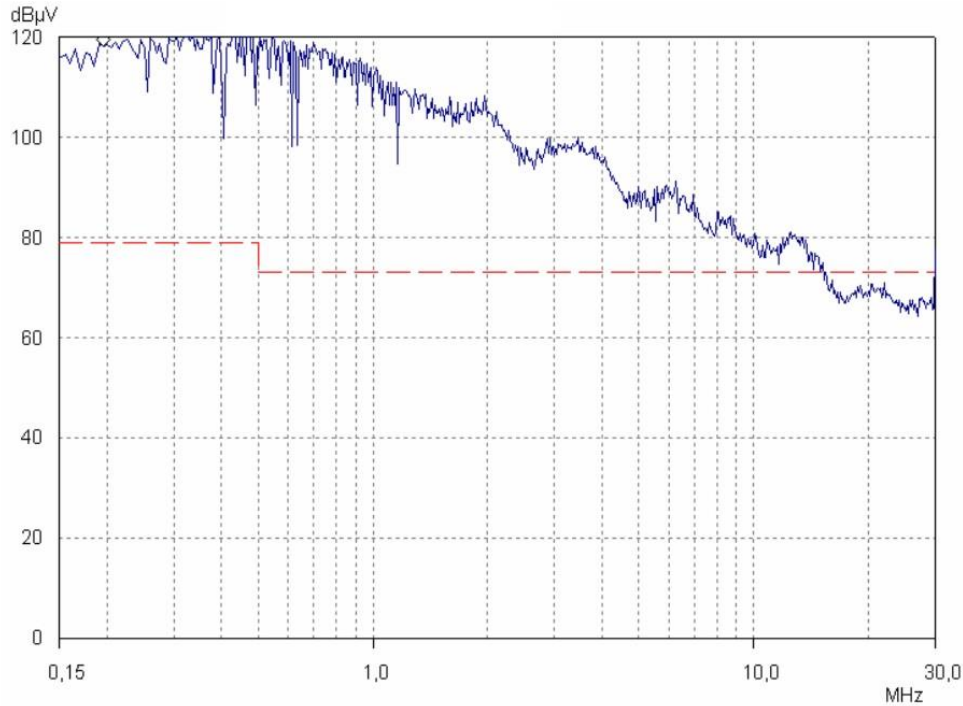


Figure 5.21. Conducted emission measurement

Since the voltage levels vary over a very broad amplitude range, Figure 5.21 uses a logarithmic scale for the vertical axis (the horizontal frequency axis also uses a logarithmic scale). The measured voltage  $U$  is compared with a reference voltage which equals  $1\mu\text{V}$  implying the voltage level is expressed in  $\text{dB}\mu\text{V}$  by using the expression:  $20 \log \left( \frac{U}{1\mu\text{V}} \right)$ .

The maximum allowed emission level in Figure 5.21 is indicated in red and this emission limit is valid in an industrial environment (class A, group 1 according to : CISPR 11 – EN55011 : International Electrotechnical Commission : Industrial, scientific and medical (ISM) radio-frequency equipment – Electromagnetic disturbance characteristics – Limits and methods of measurement). Between 150 kHz and 500 kHz, an emission level of 79  $\text{dB}\mu\text{V}$  is allowed (QP-measurement) and between 500 kHz and 30 MHz an emission level of 73  $\text{dB}\mu\text{V}$  is allowed (QP-measurement). The measured emission level in Figure 5.21 is indicated in blue and this emission level is much higher than the allowed emission level. Actions are needed to reduce the emission level. For instance, this can be done by mounting an EMC filter between the frequency converter and the electrical power grid.

### 5.2.3 Reducing the emission levels

When considering Figure 5.21, the conducted emission levels are significantly higher than the emission levels allowed by the EMC directives (CISPR 11). These high frequent signals have an impact (or can have an impact) on the behaviour of other devices fed by this polluted electrical power grid. Moreover, the cables of the power grid behave as undesired transmitting antennas causing the radiation of electromagnetic waves in the frequency range between 150 kHz and 30 MHz. These electromagnetic waves can have an impact on the behaviour of other devices.

The measurement setup visualised in Figure 5.18 commonly disturbs a long wave and a medium wave radio receiver. When considering long wave radio emissions, very often frequencies between 148.5 kHz and 283.5 kHz are considered (in case of a frequency lower than 300 kHz, the wavelength is larger than 1000 m). When considering medium wave radio emissions, very often frequencies between 526.5 kHz and 1606.5 kHz are considered (more general, frequencies between 300 kHz and 3000 kHz). This is easily demonstrated by placing an AM radio receiver (AM = Amplitude Modulation), using the long wave and the medium wave frequency bands, more or less in the vicinity of the measurement setup. Even a battery fed AM radio receiver is heavily disturbed.

#### 5.2.3.1 The use of an EMC filter

The radiated emission level can be reduced by mounting an EMC filter between the power grid and the frequency converter as visualised in Figure 5.22. Such an EMC filter mainly behaves as a low pass filter i.e. the filter passes the 50 Hz signal but blocks the high frequent components as much as possible.

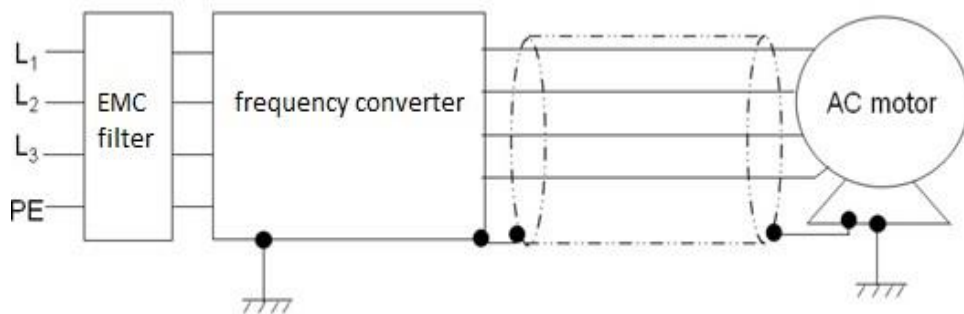


Figure 5.22. The use of an EMC filter to reduce the emission levels

Such an EMC filter is a passive filter and can be built as visualised in Fig. 5.23. although a lot of other configurations are possible. It is important the EMC filter blocks high frequent common mode currents and high frequent differential mode currents.

### 5.2.3.2 Differential mode and common mode currents

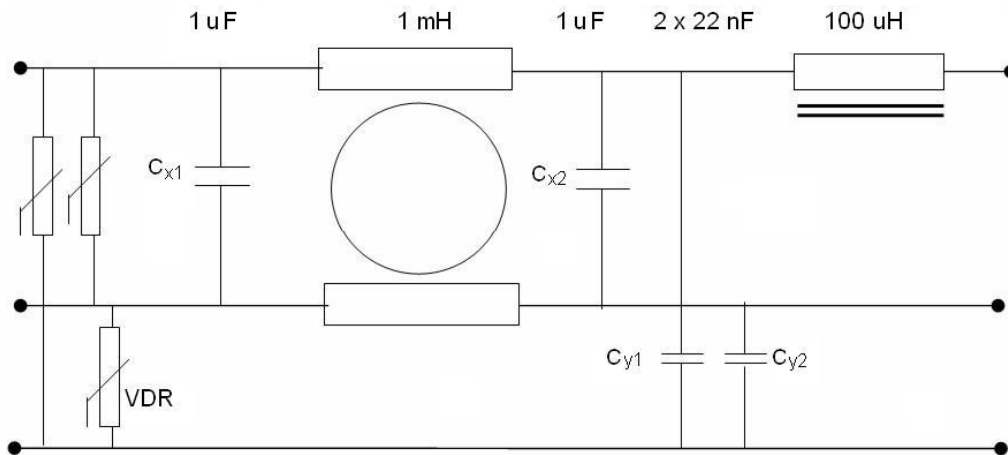


Figure 5.23. Internal circuit of a passive EMC filter (source: Deconinck et al.)

A differential mode current  $I_{DM}$  is flowing to the load using a grid conductor and the currents returns using another conductor. For instance when considering a single phase power supply, the normal 50 Hz current is a differential mode current. A common mode current  $I_{CM}$  is flowing twice in the same direction from source to the load. Parasitic capacitances are needed to allow the common mode currents to flow back to the earthed source. Normally, common mode currents are high frequent currents since the parasitic capacitances are small.

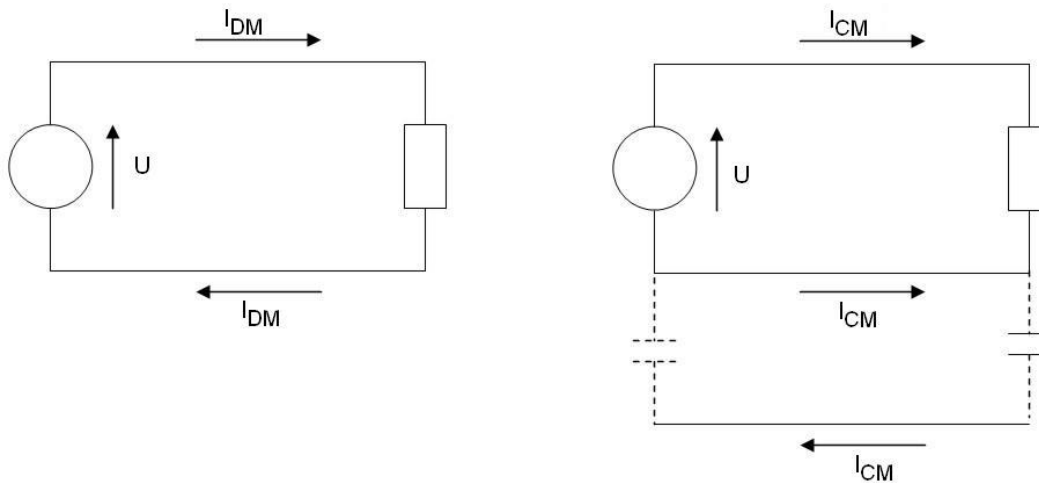


Figure 5.24. Differential mode and common mode currents (source: Deconinck et al.)

In general, the common mode currents cause larger problems than the differential mode currents. When considering a differential mode current, the current is flowing to the load and flowing back to the source using two conductors which are generally close to each other. This implies the magnetic fields generated by both conductors mainly cancel each other implying a relatively small resulting magnetic field.

The common mode currents are flowing back to the source using parasitic capacitances and different metal parts (e.g. using water pipes, reinforcements of concrete, ...). The current flowing to the load and the current flowing back to the source use paths at different locations implying the magnetic fields do not compensate each other. A larger resulting magnetic field is obtained.

#### 5.2.3.3 Internal structure of the EMC filter

When considering the EMC filter in Fig. 5.23, notice a number of components:

- capacitors  $C_X$  between active conductors (phase conductor and neutral conductor),
- capacitors  $C_Y$  between an active conductor and the earthing,
- common mode chokes (1 mH in Fig. 5.23).

In some situations, the filter also contains:

- differential mode chokes (100  $\mu$ H in Fig. 5.23.),
- resistors  $R$ ,
- Voltage Dependent Resistors (VDR) to protect the installation against voltage peaks originating from the power grid.

#### 5.2.3.4 Mounting the EMC filter

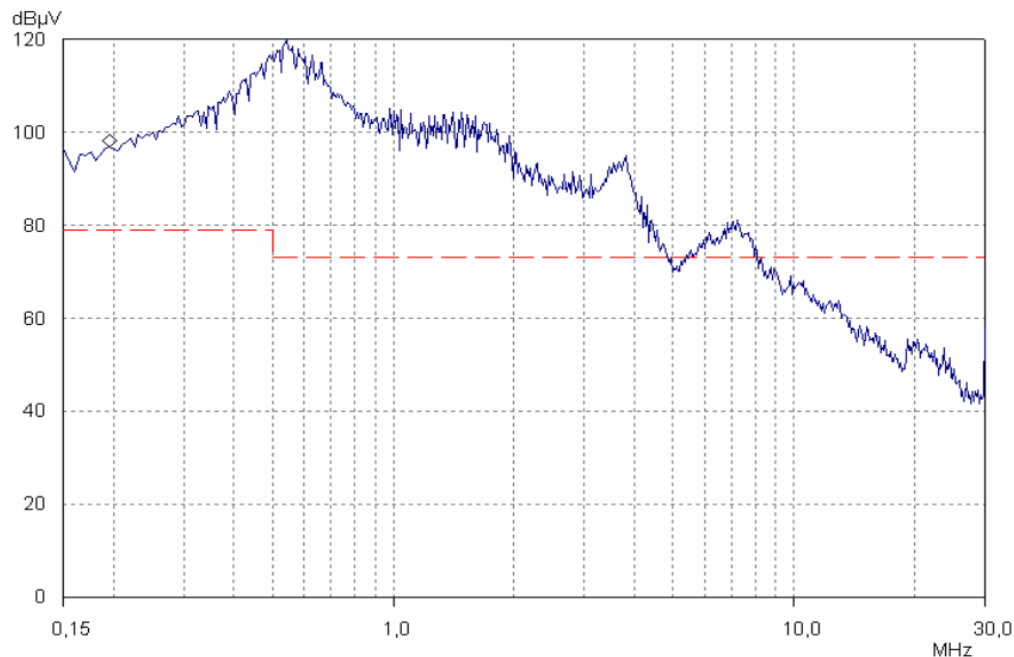


Figure 5.25. Conducted emission level when using an EMC filter without a proper high frequent earthing

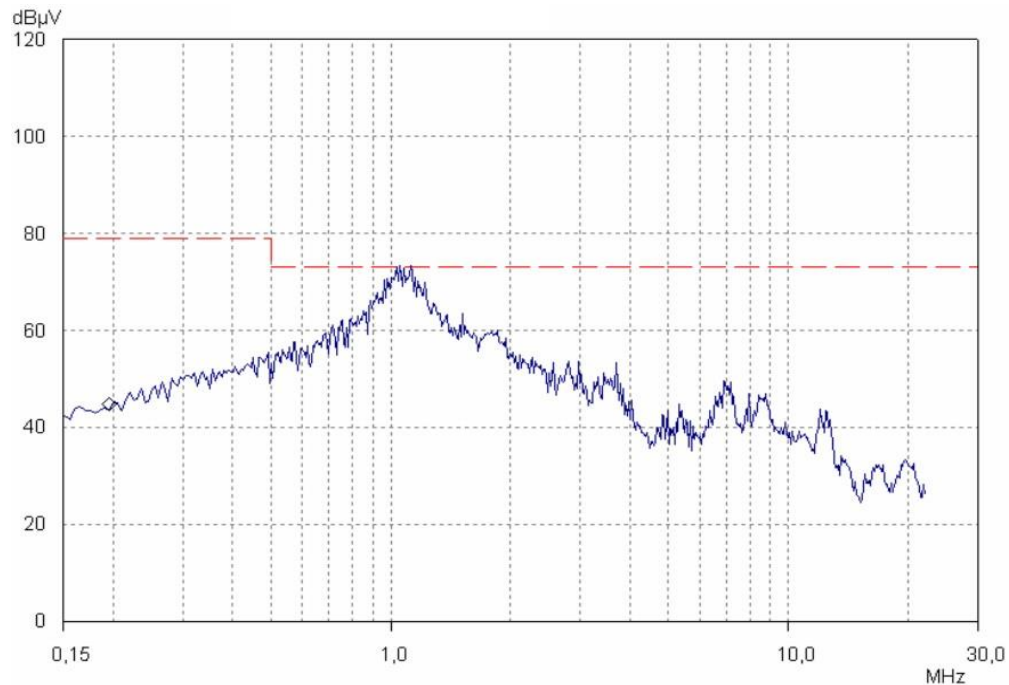


Figure 5.26. Conducted emission level when using a properly earthed EMC filter

An EMC filter is earthed using a PE conductor (Protective Earthing). However, to obtain a good high frequent earthing, this PE conductor alone is not satisfactory. This high frequent earthing is obtained by connecting the back of the EMC filter and the back of the frequency converter on a common conductive metal sheet (realising a good electrical contact by means of an entire plane).

Figure 5.25 visualises a conducted emission measurement by mounting an EMC filter between the frequency converter and the power grid without a proper high frequent earthing. When comparing with the emission measurement visualised in Figure 5.21, the emission level has decreased but the emission levels are still higher than the allowed emission limits.

In case the EMC filter is properly earthed realising a good high frequent earthing, a further decrease of the emission levels is obtained. In Figure 5.26, the properly earthed EMC filter reduces the conducted emission implying the emission limits are lower than the emission limits allowed by the EMC directives.

#### 5.2.4 References

1. Deconinck G., Saey P. (ed.), *Elektrotechniek en automatisering: Capita Selecta*, Academia Press, Gent, 2009-2010, section 9: *EMC en elektrische aandrijvingen* (Peuteman J., Knockaert J.), pp. 279 - 302.
2. Mohan N., *Power Electronics*, New York, John Wiley & Sons, 1995.
3. Paul, C.R., *Analysis of multiconductor transmission lines*, New York, John Wiley & Sons, 1994.

4. Persson, E., Transient effects in applications on PWM inverters to induction motors, in: *IEEE Transactions on Industrial Applications*, volume 28, 1992, p. 1095-1101.
5. Peuteman J., Devlies P., Knockaert J., Vandecasteele G., Verbeerst T., Vervenne I., Laboratory setup for teaching reflection phenomena between a PWM inverter and an induction motor, in: *Annual Journal of Electronics*, volume 4, number 1: ISSN 1313-1842, p. 172-175.
6. Schaffner catalogue sine eand sine plus filters: [www.schaffner.com](http://www.schaffner.com).
7. Schaffner catalogue filters: [www.schaffner.com](http://www.schaffner.com).
8. Von Jouanne, A., Rendusara, D.A., Enjeti, P.N. en Gray, J.W., Filtering techniques to minimize the effect of long motor leads on PWM inverter-fed AC motor drive systems, in: *IEEE Transactions on Ind. Appl.*, volume 32, number 4, 1996, p. 919-926.
9. Williams T., *EMC en productontwikkeling*, Elektuur – Segment B.V., 1999.



## 5.3 The practical use of a Faraday cage

### 5.3.1 Introduction and basic principle

A Faraday cage is well known from basic courses on physics. A Faraday cage is an enclosure made from a conductive material which blocks electrical and magnetic fields from outside. This allows to protect human beings and animals inside the cage, but also electrical and electronic devices can be protected.

The way of construction of such a Faraday cage can be very diverse and allows to realise a very broad range of applications. In the present text, a number of well known applications are studied and the use of a Faraday cage in the realm of EMC (EMC = Electromagnetic Compatibility) to verify normalisations on emission and immunity is discussed.

A Faraday cage is a metal enclosure, often copper or iron, which protects the inside against electrical and/or magnetic fields from outside. Figure 5.27 (source: M. Hartmann et al., Handboek Bliksembeveiliging voor gebouwen en installaties, MYbusinessmedia: EMC-reeks, Deventer, 2008) visualises a typical educational laboratory setup. Figure 5.27 visualises an unearthed Faraday cage which is loaded with static electricity. The electroscope connected with the Faraday cage reveals the charged condition of the Faraday cage. A second electroscope, mounted inside the Faraday cage, does not detect any charges. Indeed, there is no electrostatic field inside the Faraday cage; all charges are situated on the outside.

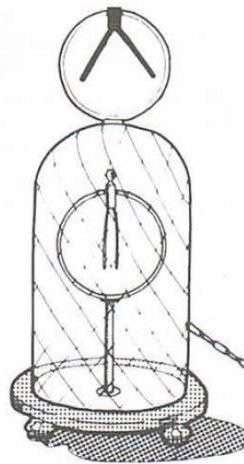


Figure 5.27. Educational Faraday cage

In general, a Faraday cage does not have the shape visualised in Fig. 5.27. For instance, an airplane, a train or a car behave as a Faraday cage due to the metal bodywork. A Faraday cage, which is loaded with electrical charges, does not cause electrical field strengths on the inner side. This phenomenon protects the passengers inside the vehicle against lightning strikes as visualised in Fig. 5.28. (source: <http://www.appelogen.be/2009/06/06/top-gear-test-de-kooi-van-faraday/>). During a lightning strike, (in the present situation simulated in a high voltage laboratory) the passengers are unharmed.



Figure 5.28. Lightning strikes on a car

### 5.3.2 Protection against lightning strikes

Benjamin Franklin developed around 1753 the first lightning rod. A lightning rod is mainly a metal rod where the lightning strikes. Using a conductor, the charges are drained to the earth. Figure 5.29 (source: <http://www.lightningconductor-ese.com/protection1.htm>) visualizes a lightning rod which protects a pilone.

In case entire buildings, including the electrical and electronic devices, must be protected against lightning strikes, it is possible to use a cage shaped structure. Also here, metal rods are pointing outside to catch the lightning strikes. Using conductors, the electrical charges are drained to the ground but the mesh shaped structure behaves as a Faraday cage which covers the walls and the roof in order to protect the building (as visualised in Fig. 5.30; source: <http://www.lightningconductor-ese.com/protection1.htm>) as it is the case of the car in Fig. 5.28.

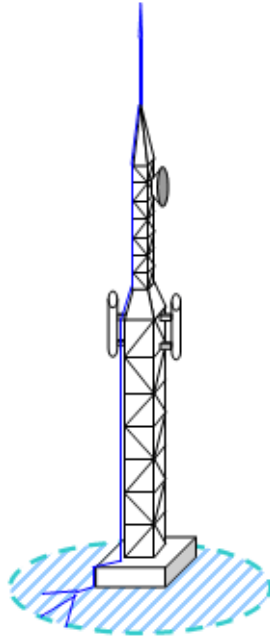


Figure 5.29. Lightning rod

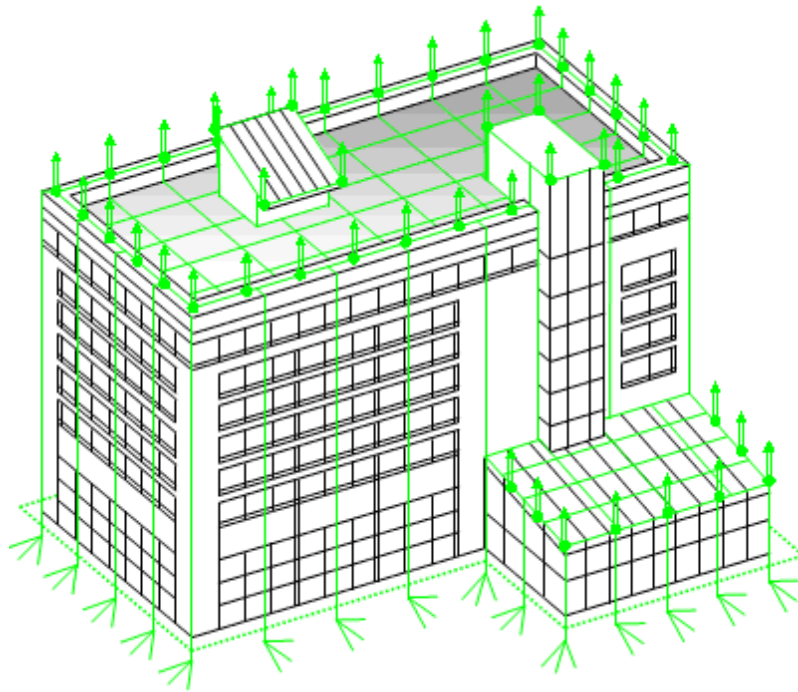


Figure 5.30. Lightning rod with a cage shaped structure

### 5.3.3 Electrical and magnetic fields

In case a constant voltage is applied between two electrically isolated conductors, a constant electrical field is obtained. This constant magnetic field has a field strength expressed in  $V/m$ . The

larger the applied voltage level and the shorter the distance between the isolated conductors, the higher the electrical field strengths.

In case a constant current is flowing in a conductor, a constant magnetic field is obtained. The magnetic field has a field strength expressed in  $A/m$ . The larger the current, the larger the magnetic field strength. Especially close to the conductor, high magnetic field strengths are obtained.

When considering alternating current (AC), there will always be a voltage and a current at the same time implying a changing electrical field and a changing magnetic field at the same time. According to Maxwell's equations, a change in the electrical field causes a magnetic field and a change in the magnetic field causes an electrical field.

In case an AC-voltage implies an AC-current, an electromagnetic wave is obtained. This electromagnetic wave propagates in the ether as a combination of electrical and magnetic fields. The propagation speed equals the propagation speed of light (since light is an electromagnetic wave). In vacuum and in air, this speed approximately equals  $3 \cdot 10^8$  m/s or 300 000 km/s.

In reality, there is a distinction between the so-called near field and the so-called far field. The transition from near field to far field is a gradual transition. When considering the near field, either the electrical field or the magnetic field can be dominant. This dominance depends on the type of antenna which is used. A transmitting antenna can be an antenna which is constructed intentionally (a dipole antenna, a biconical antenna, ...). But also the most diverse electrical and electronic components often behave as a transmitting antenna.

Close to the source of the electromagnetic radiation (the antenna), i.e. in the near field, this source determines the shape and the strength of the electrical field and the magnetic field. Sometimes when considering the near field, the magnetic field will dominate and sometimes the electrical field will dominate. More remote from the antenna, i.e. in the far field, the behaviour of the electromagnetic wave becomes less complex. The electrical field strength is perpendicular with the magnetic field strength. A plane shaped wave is obtained where the electrical field strength and the magnetic field strength are both perpendicular with the direction of propagation.

The wavelength of an electromagnetic wave is an important parameter. The wavelength is the distance travelled by this electromagnetic wave during one single period  $T$ . When considering a frequency  $f$  (expressed in Hz), this period  $T = 1/f$  which implies a wavelength

$$\lambda = c T = \frac{c}{f}. \quad (5.3)$$

Here,  $c$  is the propagation speed of the electromagnetic wave. The wavelength strongly determines where the near field and the far field are located. In case the distance to the antenna is smaller than  $\lambda/2\pi$ , a near field behaviour is obtained. In case the distance to the antenna is larger than  $\lambda/2\pi$ , a far field behaviour is obtained (although there is a transition region between the near field and the far field). Here, the assumption has been made that the dimensions of the transmitting antenna are sufficiently small. When considering antennas with larger dimensions, other formulas are needed.

### 5.3.4 Shielding using a Faraday cage

As visualised in Figure 5.27, a Faraday cage protects against static electricity. However, a Faraday cage also protects against incident electromagnetic waves originating from outside. The incident wave (here represented by its electrical component  $E_{IN}$  but the conclusion also applies for the magnetic component) is partially reflected implying a reflected wave  $E_{R1}$ . The part  $E_{MED1}$  which is not reflected propagates through the wall and attenuates to  $E_{MED2}$ . A part of  $E_{MED2}$  will be reflected giving  $E_{R2}$  and finally only  $E_{UIT}$  enters the cage.

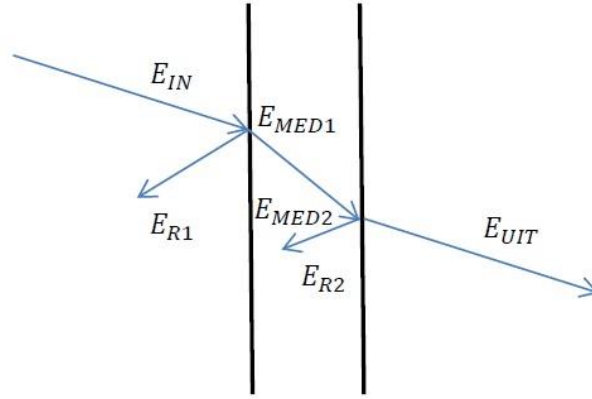


Figure 5.31. Shielding using a wall of the Faraday cage

In case the Faraday cage realises an appropriate shielding,  $E_{UIT}$  is much smaller than  $E_{IN}$ . Often, this shielding effectiveness is expressed using dBs (decibel). More precisely, a shielding effectiveness

$$SE \text{ (dB)} = 20 \log_{10} \left( \frac{E_{IN}}{E_{UIT}} \right) \quad (5.4)$$

is obtained. In case the Faraday cage has a shielding effectiveness of 100 dB, this means that  $E_{UIT}$  is 100000 times smaller than  $E_{IN}$ . In case of a shielding effectiveness of 40 dB, then  $E_{UIT}$  is 100 times smaller than  $E_{IN}$ .

The working principle visualised in Figure 5.31 applies to incident waves in the near field and to incident waves in the far field. Not only the amplitudes of the reflected waves are important. Also the attenuation from  $E_{MED1}$  to  $E_{MED2}$  is important. The higher the frequency of the incident wave, the larger the attenuation. Also the electrical conductivity and the magnetic conductivity of the wall in Figure 5.31 are important. The higher these conductivities of the wall, the higher the attenuation. This attenuation also increases as the thickness of the wall increases.

In general, when considering incident electromagnetic waves in the far field, it is quite easy to obtain a high shielding effectiveness SE. Shielding incident waves in the near field, which mainly consist of an electrical field, is also not that difficult. Shielding incident waves, which mainly consist of a magnetic field, is much more difficult. This is especially the case when the frequency is (relatively) low. Indeed, realising a shielding effectiveness against low frequent magnetic fields is quite difficult.

When considering a steel wall, the magnetic conductivity is much higher than the magnetic conductivity of a copper wall. In case also for low frequent magnetic fields a considerable attenuation from  $E_{MED1}$  to  $E_{MED2}$  is required, than a steel wall having a sufficient thickness is needed. However, shielding against low frequent magnetic fields remains difficult to realise. A Faraday cage does not protect against the magnetic field of the Earth (having a magnetic induction of approximately 30 to 60 micro Tesla). Also inside a Faraday cage, a compass will indicate the direction of the North.

#### 5.3.5 Faraday cage having walls with apertures

A Faraday cage can be built using metal sheets (for instance made of copper or steel) but a Faraday cage can also be built using a metal mesh. A Faraday cage is allowed to have apertures in the walls as long as these apertures are not too large. A rule of thumb states that the cage is ‘impermeable’ for electromagnetic waves as long as the dimensions of the apertures are smaller than one tenth of the wavelength of the waves which must be eliminated.



Figure 5.32. Test chamber containing a door

A test chamber is a Faraday cage which is often useful to perform EMC measurements. Test chambers (as visualised in Figure 5.32, source: <http://shieldingsystems.eu>) are able to realise shielding effectiveness SE of more than 100 dB. The walls of these test chambers are built using solid walls but often the walls even contain two solid walls. The walls are made of “sandwich”-panels made of wood and steel. The walls contain on both sides a steel wall with a wooden support in between.

Figure 5.33 (source: <http://shieldingsystems.eu>) visualises the shielding effectiveness SE (expressed in dB) of such a test chamber as visualised in Figure 5.32. This shielding effectiveness depends on the frequency. The high shielding effectiveness is only obtained in case the frequency of the incident wave is not too high nor too low.

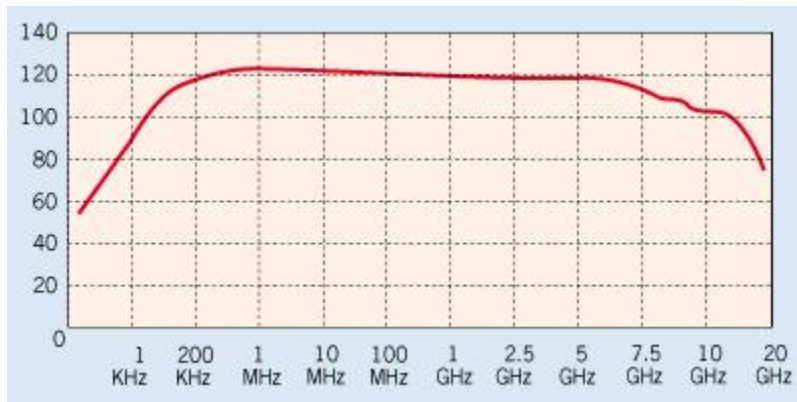


Figure 5.33. Frequency dependent shielding effectiveness

In practice, the walls of the test chamber contain a number of apertures. Apertures are not only needed to allow ventilation (to allow fresh air entering the chamber or to eliminate heat from the chamber) but also measurement cables and power cables must pass the walls. It is important to limit the number and the sizes of apertures as much as possible to keep the shielding effectiveness as high as possible. For instance, ventilation is realised using a honeycomb structure.

Especially when considering a test chamber, it is important to avoid rectangular apertures since especially the long side of the aperture has an impact on the reduction of the shielding effectiveness. Since a test chamber (as visualised in Figure 5.32) always needs a door, it is important that the borders of this door do not realise a split. As visualised by Fig. 5.34, a lot of attention focuses towards an appropriate sealing of the door using electromagnetic seals called gaskets.

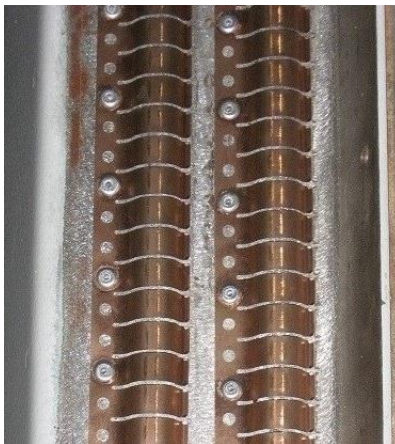


Figure 5.34. Gaskets sealing the door

A test chamber or a Faraday cage can be constructed using a mesh made of metal wires but in such a situation, the shielding effectiveness is much lower (for instance only 40 dB).

### 5.3.6 Miscellaneous

A Faraday cage shields the interior of the test chamber from electromagnetic waves from outside. For instance, a mobile phone is not able to receive messages from an external antenna when this mobile phone is situated inside a well shielding Faraday cage. The same cage of Faraday will also protect the outer world from electromagnetic waves which are generated inside the Faraday cage. Using a mobile phone, it is not possible to send electromagnetic waves to an external GSM antenna.

In case a room behaves unintentionally as a Faraday cage (for instance an underground subway station where the reinforcement of concrete behaves as a Faraday cage) and it is important to give passengers the possibility to use their mobile phones, it is required to mount GSM-antennas inside this shielded room.

In case transmitters need to be tested which transmit emergency signals, these tests need to be done in a Faraday cage. It is important to avoid external people from receiving these test emergency signals and react.

The shielding principle realised in Fig. 5.31 does not require that the Faraday cage is earthed. In reality, a Faraday cage is almost always earthed (for instance the test chambers of Fig. 5.38, Fig. 5.40, Fig. 5.43, Fig. 5.44) to protect people against indirect contact of electrical active parts. It is also important to protect people against static electricity (due to the Faraday cage principle, this earthing is only required to protect people against static electricity when touching the cage from the outside). Notice however, no earthing is needed to shield electromagnetic waves.

As already explained, trains and cars behave as a Faraday cage. When considering GSM frequencies of approximately 900 MHz, the wavelength equals 33 cm which implies the windows are sufficiently large to let the GSM frequencies pass. Indeed, a mobile phone can be used in a train or in a car.

A microwave oven actually also behaves as a Faraday cage. The food is heated using microwaves in a metal cage (the oven) which keeps the electromagnetic waves inside (this avoids energy losses and takes care other electronic devices are not disturbed). In order to make the door transparent and allow the cook to see the food, the door contains a metal coating with a sufficient number of apertures. The microwaves have a wavelength of approximately 12 cm (a frequency of 2.45 GHz is used). In case the dimensions of the apertures are smaller than 1.2 cm the microwaves remain inside the oven.

In hospitals, sensitive medical devices are often installed inside a Faraday cage. In this way, mobile phones in the corridor, or the presence of transformers, do not harm EMG measurements (electromyogram = measuring electrical activity of muscles) or ECG measurements (electrocardiogram = measuring activity of the heart muscle = an electromyogram of the heart).

A Faraday cage also takes care that medical devices do not disturb their environment. For instance, in a MRI scan (Magnetic Resonance Imaging = imaging using magnetic resonance) very strong electromagnetic waves are generated (having magnetic inductions of a few tesla) and it is important



to avoid disturbances of other devices. Fig. 5.35 (source: <http://www.hollandshielding.com>) visualises such a MRI scanner.



Figure 5.35. MRI scanner in a shielded environment

#### 5.3.7 EMC measurements

It is a technological evolution during the last decades that electronic devices have the tendency to:

- become more and more sensitive to disturbances originating from electromagnetic waves emitted by other devices i.e. the immunity level of the devices decrease,
- emit an increased level of electromagnetic waves and to disturb other devices i.e. the emission level of the devices increase,

As Figure 5.36 visualises, from a specific moment in time the emission level is higher than the immunity level which means a first device will harm the proper operation of another device. In such a situation, electromagnetic interference (EMI) occurs. Several EMC techniques have been developed to prevent a further decrease of the immunity level and to prevent a further increase of the emission level.

As Figure 5.37 visualises, using appropriate normalisations engineers try to keep the emission level lower than the immunity level which normally implies a first device does not disturb another device. The devices are electromagnetically compatible (EMC = Electromagnetic Compatibility).

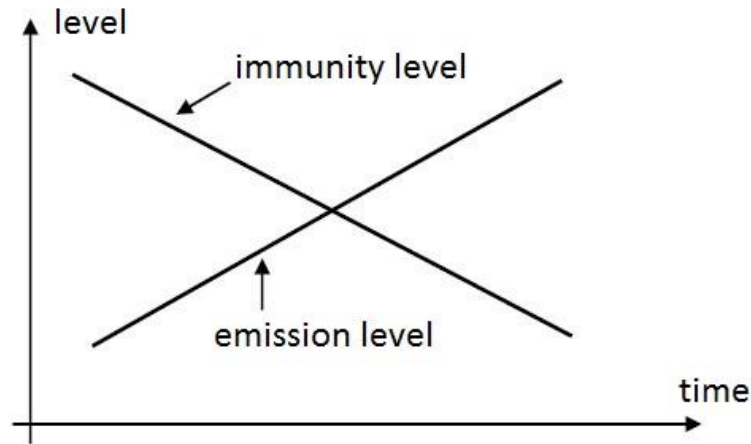


Figure 5.36. EMI related problems

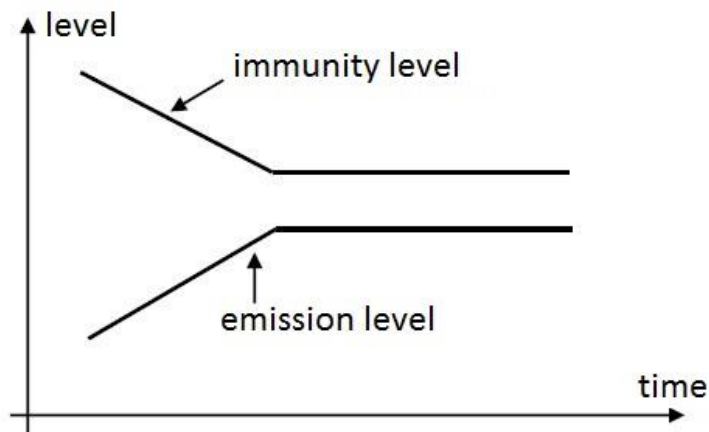


Figure 5.37. EMC philosophy

To verify whether radiated emissions do not exceed the allowed emission limits, the emission level must be measured using a receiving antenna. When performing such a measurement, it is important to measure whether the device (the DUT = Device Under Test = the EUT = Equipment Under Test) radiates. The measurement result is not allowed to be disturbed by other devices. Therefore, it is important to perform such an emission measurement in a Faraday cage.

Several directives require that electronic devices are not disturbed by electromagnetic waves having a specified intensity. To verify these requirements, immunity tests are needed. Using a transmitting antenna, disturbances are emitted and one investigates whether the device still functions properly. It is important the DUT is only influenced by the controlled and known electromagnetic waves emitted by the transmitting antenna and not by additional unknown and uncontrollable disturbances. The immunity measurements are performed inside a Faraday cage.



Figure 5.38. Faraday cage

### 5.3.8 Anechoic chamber

The Faraday cage in Figure 5.38 (photograph taken in the research laboratory at the KU Leuven Research Group ReMI at Ostend) has metal walls which shield the inner part of the cage and the outer world. Notice however these metal walls also cause reflections. In case the inner walls of the Faraday cage are covered with absorbers, then incident electromagnetic waves are absorbed and reflections are avoided. By avoiding these reflections, the behaviour of an open space having infinitely large dimensions is obtained. Figure 5.39 visualises such pyramid shaped absorbers which absorb electromagnetic waves in a large frequency range. For instance, ETS-LINDGREN sells absorbers which absorb electromagnetic waves in a frequency range from 30 MHz to 100 GHz (<http://www.ets-lindgren.com/EHP-MicrowaveAbsorber>).

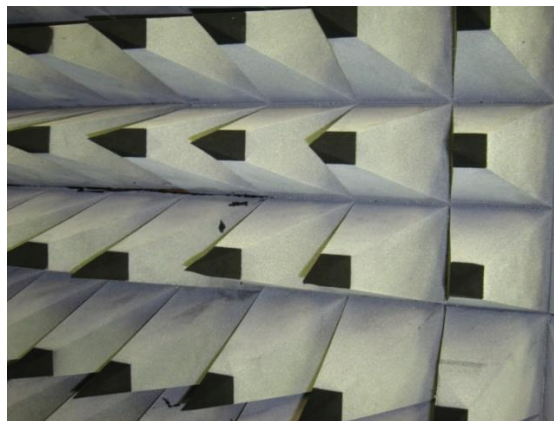


Figure 5.39. Pyramid shaped absorbers

In case the walls, the ceiling and the floor are covered with absorbers, in theory all reflections are avoided (Fig. 5.40; photograph taken in the research laboratory at the KU Leuven Research Group ReMI at Ostend). Sometimes, only the walls and the ceiling are covered with absorbers implying

the floor still accounts for reflections (semi anechoic chamber). This situation mimics measurements performed in a free open space where also a floor accounts for reflections.

The absorbing material at the inner sides of the Faraday cage need to absorb waves as much as possible and irrespective of the angle of incidence. Pyramid shaped absorbers, as visualized in Figure 5.39, need a limited electric conductivity. In practice, such absorbers are made of foam impregnated with a well chosen concentration of carbon and iron.

The length of such pyramid shaped cones is determined by the minimal frequency which needs to be absorbed and the desired absorption rate. The lower the frequencies which need to be absorbed, the larger the required length. The pyramid shaped cones absorb maximally when the incident wave incides perpendicular with the wall of the cage and the cones have a length of approximately  $\lambda/4$ . Here,  $\lambda$  is the wavelength of the incident wave in the free air.

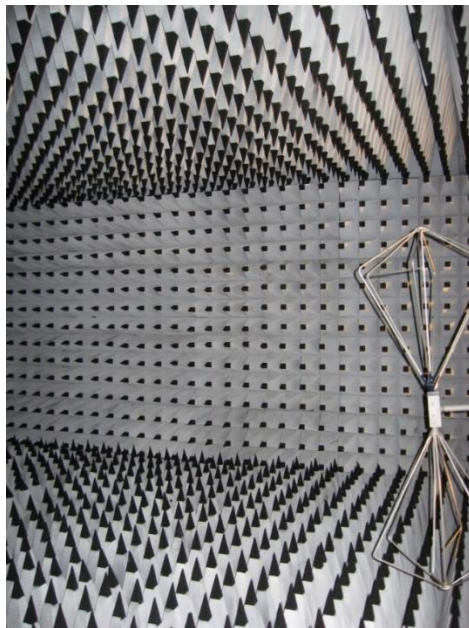


Figure 5.40. Anechoic chamber covered with pyramid shaped absorbers

Not only pyramid shaped absorbers are used. Very often, also tiles made of a ferrite material are used. These ferrite tiles absorb incident waves having frequencies ranging from 30 MHz to 1000 MHz. When considering higher frequencies, their behaviour reduces. Such ferrite tiles are easy to attach to the conductive walls.

The energy of the absorbed waves is converted into heat. This amount of heat can be large in case a transmitting antenna is situated close to a wall covered with absorbers. Especially when an antenna emits a lot of power in a limited space, the temperature of the absorbers can rise. This local temperature rise is not allowed to cause fire hazard implying the absorbers contain fire retardant materials. Moreover, a Faraday cage can contain ventilators to eliminate the dissipated heat. Some Faraday cages contain smoke detectors.

### 5.3.9 Performing measurements in an anechoic chamber

Figure 5.41 visualises an anechoic chamber inside a Faraday cage which is used to perform emission measurements. The device which must be tested (EUT = Equipment Under Test) is fed by a three phase (filtered) power supply. The EUT is placed at a wooden table (as visualised in Fig. 5.42). There is a distance of three meters between the receiving antenna and the EUT. The receiving antenna is connected with an EMI receiver. Using a personal computer, the measurement data is stored and visualised. The EMI receiver and the personal computer are situated outside the cage in order to prevent them from disturbing the measurements.

In case an antenna or a DUT (DUT = Device Under Test = EUT = Equipment Under Test) emits electromagnetic waves, the distinction between the near field and the far field becomes important. To perform measurements in the far field, the test chamber must be sufficiently large since the distance between the emitting DUT and the receiving antenna must exceed the  $\lambda/2\pi$  distance.

For instance, when considering measurements in the frequency range from 30 MHz to 200 MHz, the wavelength  $\lambda$  varies between 10 m and 1.5 m. This implies  $\lambda/2\pi$  varies between 1.59 m and 0.24 m. The antenna is situated in the far field by respecting the 3 m distance.

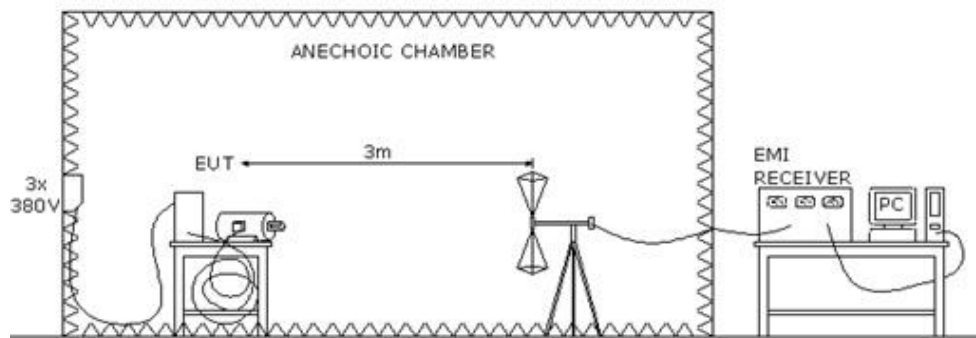


Figure 5.41. Emission measurements in the anechoic chamber

Reflections in a test chamber are not only avoided by mounting absorbers on the walls. In case the EUT is placed on a table, it is a wooden table and not a metal table in order to avoid reflections. For the same reason, the antenna is mounted on a wooden support.



Figure 5.42. EUT on a wooden table inside a Faraday cage

#### 5.3.10 Electromagnetic reverberation chamber

A reverberation chamber is actually the opponent of an anechoic chamber. In an anechoic chamber, reflections are avoided as much as possible whereas in a reverberation chamber the occurrence of reflections is stimulated. Due to the reflections, i.e. when having as less absorption as possible, it is possible to realise a high field strength with a minimum of emitted energy. Notice in Figure 5.43 the metal walls which cause reflections. In order to have (almost) everywhere field strengths which are approximately the same, additional reflectors are needed. These additional reflectors are rotating.

Fig. 5.43. visualises an electromagnetic reverberation chamber in the Otto-von-Guericke University at Magdenburg in Germany (source: [http://en.wikipedia.org/wiki/Electromagnetic\\_reverberation\\_chamber](http://en.wikipedia.org/wiki/Electromagnetic_reverberation_chamber)). In the corner on the right, notice the additional rotating reflectors. The DUT is a motorcycle. Indeed, motorcycles and cars contain an increasing number of electronic parts and EMC measurements are needed to obtain a reliable operation.





Figure 5.43. Electromagnetic reverberation chamber

Notice that also the earlier mentioned microwave oven has metal walls which implies reflections of the electromagnetic waves. In this microwave oven, it is important to realise a maximum field strength to heat the food using a minimum power.

#### 5.3.11 Semi anechoic chamber

Between the full anechoic chamber (Figure 5.40 and Figure 5.41 where the ceiling, the walls and the floor are covered with absorbing materials) and the reverberation chamber (Figure 5.43 lacking all absorbing materials), a semi anechoic chamber is situated.

Figure 5.44 (photograph taken in the research laboratory at the KU Leuven Research Group ReMI at Ostend) visualises a semi anechoic chamber where absorbing pyramid shaped cones are mounted on well chosen locations. Figure 5.45 (source: <http://www.e-ohtama.jp/english/lab/tokyo.html>) visualizes a semi anechoic chamber where the ceiling, and the walls are covered with absorbing material whereas the floor has reflective properties. This allows to simulate an open space having a reflective floor.



Figure 5.44. Semi anechoic chamber



Figure 5.45. Semi anechoic chamber having a reflective floor





Figure 5.46. Demonstration using a Faraday cage

#### 5.3.12 References:

1. M. Hartmann et al., Handboek Bliksembeveiliging voor gebouwen en installaties, MYbusinessmedia: EMC-reeks, Deventer, 2008.
2. K.L. Kaiser, Electromagnetic Compatibility Handbook, CRC Press, 2005.

## 5.4 Radiated emission

### 5.4.1 Introduction

Radiated emission occurs whenever electromagnetic energy is emitted by an electronic device. It is possible to make a distinction between intentionally and unintentionally emitted electromagnetic energy. Figure 5.47 visualises a long wave transmitting antenna (source: <http://ieccetech.org/issue/2011-11/Keeping-copper-in-the-loop>) realising an intentional emission of electromagnetic waves.



Figure 5.47. Long wave transmitting antenna

Electronic devices (and parts of these electronic devices) also unintentionally emit electromagnetic waves. Actually emission occurs whenever an AC current is flowing and whenever an AC voltage is available. To prevent EMI, it is important the intensity of the radiated emission is sufficiently low. Standards are available which describe the allowed emission levels.

### 5.4.2 Electrical and magnetic fields

In case a constant voltage is applied between two electrically isolated conductors, a constant electrical field is obtained. This constant electrical field has a field strength expressed in  $V/m$ . The larger the applied voltage level and the shorter the distance between the isolated conductors, the higher the electrical field strengths.

In case a constant current is flowing in a conductor, a constant magnetic field is obtained. The magnetic field has a field strength expressed in  $A/m$ . The larger the current, the larger the magnetic field strength. Especially close to the conductor, high magnetic field strengths are obtained.

When considering alternating current (AC), there will always be a voltage and a current at the same time implying a changing electrical field and a changing magnetic field at the same time. According to Maxwell's equations, a change in the electrical field causes a magnetic field and a change in the magnetic field causes an electrical field.

In case an AC-voltage implies an AC-current, an electromagnetic wave is obtained. This electromagnetic wave propagates in the ether as a combination of electrical and magnetic fields. The propagation speed equals the propagation speed of light (since light is an electromagnetic wave). In vacuum and in air, this speed approximately equals  $3 \cdot 10^8$  m/s or 300 000 km/s.

#### 5.4.2.1 Near field and far field

A transmitting antenna can be an antenna which is constructed intentionally (a dipole antenna, a biconical antenna, ...), but also the most diverse electrical and electronic components often behave as a transmitting antenna. When considering the emitted electromagnetic field, there is a distinction between the so-called near field and the so-called far field. The transition from near field to far field is a gradual transition.

Close to the source of the electromagnetic radiation, i.e. in the near field, this source determines the shape and the strength of the electrical field and the magnetic field. When considering the near field, either the magnetic field will dominate or the electrical field will dominate. More remote from the source of the electromagnetic radiation, i.e. in the far field, the behaviour of the electromagnetic wave becomes less complex. The electrical field strength is perpendicular with the magnetic field strength. A plane shaped wave is obtained where the electrical field strength and the magnetic field strength are both perpendicular with the direction of propagation.

#### 5.4.2.2 The wavelength

The wavelength of an electromagnetic wave is an important parameter. The wavelength is the distance travelled by this electromagnetic wave during one single period  $T$ . When considering a frequency  $f$  (expressed in Hz), this period  $T = 1/f$  which implies a wavelength

$$\lambda = c T = \frac{c}{f}. \quad (5.5)$$

Here,  $c$  is the propagation speed of the electromagnetic wave. The wavelength strongly determines where the near field and the far field are located. In case the distance to the antenna is smaller than  $\lambda/2\pi$ , a near field behaviour is obtained. In case the distance to the antenna is larger than  $\lambda/2\pi$ , a far field behaviour is obtained (although there is a transition region between the near field and the far field). Here, the assumption has been made that the dimensions of the transmitting antenna are sufficiently small. When considering antennas with larger dimensions, other formulas are needed to distinguish between the near field and the far field.

The wavelength is also an important parameter when designing an antenna. Figure 5.48 visualises the experimental transmission station used by Marconi in 1901 to send the letter S (using Morse code: dot dot dot) from Poldhu (Cornwall, United Kingdom) across the Atlantic Ocean to Newfoundland. Marconi used long waves (electromagnetic waves having a low frequency and therefore a large wavelength) since these long waves are, due to reflections, able to bridge large

distances. However, using these long waves requires the use of sufficiently large transmitting antennas as visualised in Fig. 5.48 (source: <http://www.mds975.co.uk/Content/ukradio.html>).

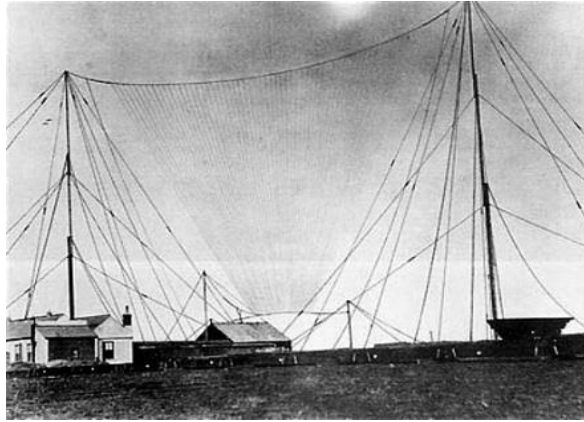


Figure 5.48. Long wave transmitting antennas used by Marconi

#### 5.4.3 Measuring unintentionally emitted electromagnetic waves

All conductors and other components in an electronic device behave as antennas and cause radiated emission when an AC current is flowing. Standards are available in the frequency range from 150 kHz to 1000 MHz. In order to measure the intensity of these electromagnetic waves, in the frequency range from 150 kHz to 30 MHz mainly the magnetic field strength is measured. In the frequency range from 30 MHz to 1000 MHz mainly the electrical field strength is measured.

When measuring the radiated emission level of an electronic device, it is important the radiated emission level of only this device has been measured. The measurement is not allowed to be influenced by the emissions of other devices (for instance also not by the emission of radio broadcasting stations). Therefore, such a measurement is performed in a cage of Faraday i.e. in an electromagnetically shielded room.

Figure 5.49 visualises such a cage of Faraday. The electronic device (Equipment Under Test = EUT; Device Under Test = DUT) is situated on a wooden table. The receiving antenna is placed in the cage of Faraday and a distance of 3 meter is available between the DUT and the antenna. The receiving antenna measures the emission levels and the measurement data are sent to an EMI receiver outside the cage of Faraday. Using a PC, the measurement data are visualised which allows an interpretation of the results.

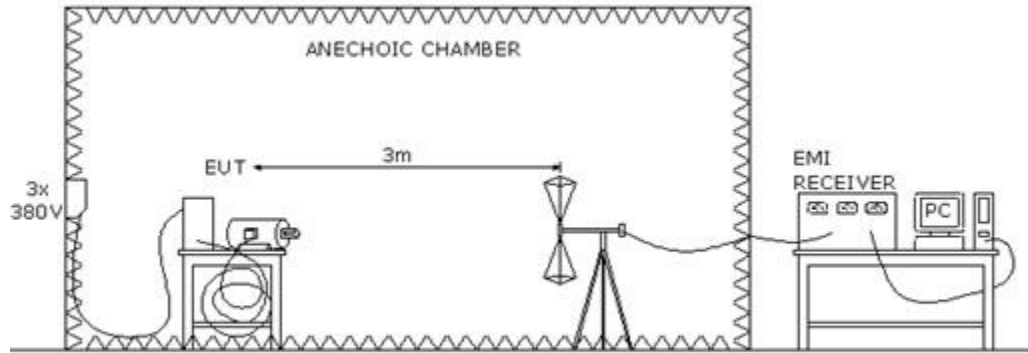


Figure 5.49. Radiated emission measurement in a cage of Faraday

Notice in Fig. 5.49 that the walls, the floor and the ceiling of the cage are covered with cones. These cones absorb the electromagnetic waves avoiding reflections. In this way, an open electromagnetic space is simulated. Notice also the biconical receiving antenna.

#### 5.4.3.1 Radiated emission measurements

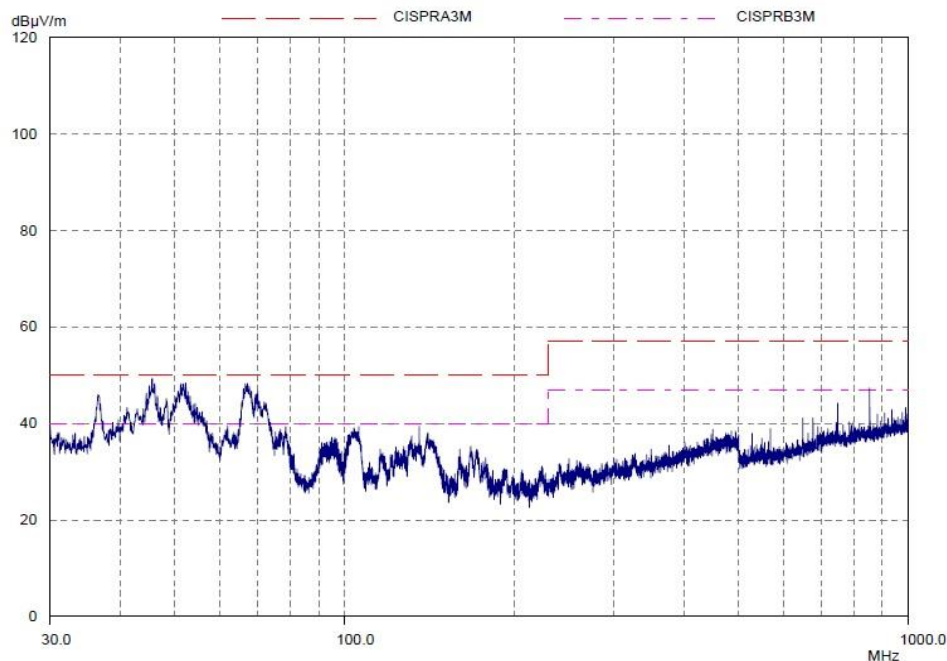


Figure 5.50. Radiated emission measurement

Figure 5.50 visualises a radiated emission measurement (blue coloured) which has been obtained in the cage of Faraday visualised in Fig. 5.49. The measurements are performed in a frequency range between 30 MHz to 1000 MHz (using a logarithmic frequency axis). The measured electrical field strengths are expressed in  $V/m$ . Also for the vertical axis, a logarithmic scale has been used. This logarithmic scale is used because of the large range of measured field strengths. The measured

field strength  $E$  is compared with a reference of  $1 \mu V/m$ . This implies the field strength is expressed in  $dB\mu V/m$  using the expression

$$20 \log \left( \frac{E}{1 \mu V/m} \right). \quad (5.6)$$

#### 5.4.3.2 The decibel

The radiated emission level in Figure 5.50 is expressed in  $dB\mu V/m$  using the expression (5.6).

Such an approach can be generalised starting from a quantity  $X$  expressed in the unity  $y$ . More precisely,

$$X(dBy) = 20 \log(X(y)) \quad (5.7)$$

Not only the notation  $dBy$  but also the notation  $dB(y)$  is often used. When considering the unity  $y$ , a number of habits exist. For instance when considering a voltage, often the unity  $\mu V$  is used giving  $dB\mu V$ . When considering a current, often the unity  $\mu A$  is used giving  $dB\mu A$ . An electrical field strength is often expressed using  $dB\mu V/m$ . For instance a field strength of  $30 dB\mu V/m$  corresponds with a field strength of  $32 \mu V/m$ .

When studying emission limits mentioned in standards for frequencies lower than  $30 MHz$ , one would expect the unity  $dB\mu A/m$  since magnetic field strenghts are measured. But also for frequencies lower than  $30 MHz$ , often the unity  $dB\mu V/m$  is used which corresponds with an electrical field strength. Using the expression  $H(dB\mu A/m) = H(dB\mu V/m) - 51.5(dB\Omega)$ ,

the magnetic field strength can be calculated using  $dB\mu A/m$  in case it is given using  $dB\mu V/m$ . Here,  $51.5 dB\Omega$  equals  $20 \log(Z_0)$  where  $Z_0 = 377\Omega$  which is the impedance of a plane wave.

#### 5.4.3.3 Antennas

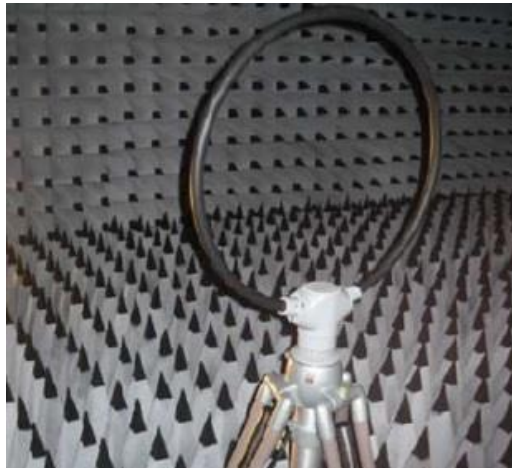


Figure 5.51. Loop antenna

When measuring in the frequency range from 150 *kHz* to 30 *MHz*, commonly the loop antenna visualised in Fig. 5.51 is used. When measuring frequency components higher than 30 *MHz*, commonly the biconical antenna visualised in Fig. 5.52 is used.

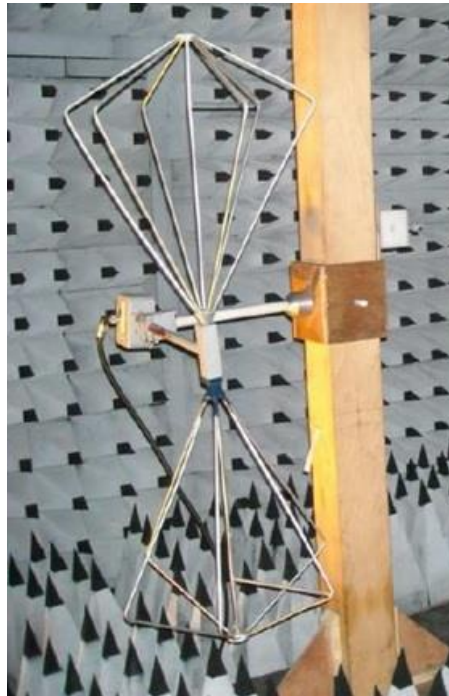


Figure 5.52. Biconical antenna

#### 5.4.4 EMC standards

Mentioning all emission standards is far beyond the scope of the present text. We restrict ourselves to a few relevant examples.

When considering the CISPR 11 standard (CISPR stands for Comité International Spécial des Perturbations Radio) formulates emission limits for ISM equipment. Here, ISM stands for Industrial, Scientific and Medical equipment. The standard separates the equipments in two different groups:

- “Group 1 ISM equipment contains all ISM equipment in which there is intentionally generated and/or used conductively coupled radio-frequency energy which is necessary for the internal functioning of the equipment itself.”
- “Group 2 ISM equipment contains all ISM equipment in which radio-frequency energy is intentionally generated and/or used in the form of electromagnetic radiation for the treatment of material, and EDM (electrodischarge machine) and arc welding equipment.”

In each group, a distinction is made between Class A and Class B ISM equipment:



- “Class A ISM equipment is equipment suitable for use in all establishments other than domestic and those directly connected to a low voltage power supply network which supplies buildings used for domestic use.”
- “Class B ISM equipment is equipment suitable for use in domestic establishments and in establishments directly connected to a low voltage power supply network which supplies buildings used for domestic use.”

More information is available in the CISPR 11 standard (and the EN 50 011 standard). The allowed emission limits are summarized in Table 5.1 for Group 1 equipment.

Table 5.1. Radiated emission limits according to the CISPR 11 standard

Frequency (MHz)	Group 1, class A measured using a distance of 10 m ( $dB\mu V/m$ )	Group 1, class B measured using a distance of 10 m ( $dB\mu V/m$ )
0.15 – 30	under consideration	under consideration
30 – 230	40	30
230 - 1000	47	37

When considering frequencies between 150  $kHz$  and 30  $MHz$ , the magnetic component of the electromagnetic waves are measured using a loop antenna. When considering frequencies between 30  $MHz$  and 1  $GHz$ , the electrical component of the electromagnetic waves are measured using a dipole antenna (or a biconical antenna or a log periodic antenna).

Notice that the emission limits of class B equipments are more restrictive than when considering class A equipments. In a domestic environment, the emission limits are more restrictive.

The emission limits in Table 1 are valid in case the distance between the DUT and the antenna equals 10  $m$ . In case the distance between the DUT and the antenna only equals 3  $m$ , the emission level is 10  $dB$  larger (since  $20 \log(10/3) = 10 \text{ dB}$ ). The measurement in Figure 4 has been performed with a distance of 3  $m$  between the DUT and the antenna giving emission limits of 40  $dB$  and 47  $dB$  in case of class B equipment and emission limits of 50  $dB$  and 57  $dB$  in case of class A equipment.

#### 5.4.4.1 Measuring the magnetic field strength

When considering frequencies lower than 30  $MHz$ , the standards consider the magnetic field strength. A distance of 10  $m$  is mentioned between the DUT and the antenna. In principle, the DUT and the antenna are placed in an open field. This means the DUT must be located in a wooden cabin in order to protect this DUT against rain and other hostile weather conditions. A wooden cabin is used since a wooden cabin has (almost) no impact on the electromagnetic field emitted by



the DUT. It is also important there are no reflecting objects in the vicinity of the DUT. During the measurements, the DUT is rotated until the direction with the maximum radiation is obtained. Secondly, the measuring antenna is rotated along a vertical axis until a maximum measurement is obtained.

In reality, it is often a problem that the intensity of the disturbances due to external causes (e.g. due to radio broadcasting stations) is tens of dBs higher than the emission limit allowed by the standard. This implies a useful measurement is practically impossible.

Due to this reason, the antenna is very often placed closer to the DUT (where the field strengths of the radiated waves are higher). Using this measurement, the field strength at a distance of 10 m is calculated giving an approximation of the real value.

#### 5.4.4.2 Measuring the electrical field strength

When considering frequencies higher than 30 MHz, always the electrical field strength will be measured. The measurement setup (in the open field) is visualised in Fig. 5.53. (source: J. J. Goedbloed, Elektromagnetische compatibiliteit, analyse en onderdrukking van stoorproblemen, Kluwer Technische boeken B.V., Deventer – Antwerpen, 1990). The DUT is placed on a turning table at an appropriate distance (for instance 10 m) from the antenna. The turning table has a height of 1 m and a conducting ground plane is required. First, the DUT is rotated until the maximum intensity of the radiation is obtained. Secondly, the antenna is moved upwards and downwards until a maximum measurement is obtained. This maximum is obtained when the direct radiation and the reflected radiation have the same phase which implies that they are added to each other. The measurement must be performed for the vertically and for the horizontally polarized field.

In the official requirements, the measurement must be performed using a  $\lambda/2$ -dipole antenna. In reality, almost nobody uses such a dipole antenna since the antenna must be adjusted for each harmonic of the clock oscillator which is a very time consuming approach. For this reason, it is allowed to use a broadband antenna (e.g. a biconical antenna).

In order to be independent of weather influences and the electromagnetic noise from the environment, the measurements are almost always performed in a cage of Faraday. This cage of Faraday contains absorbing cones which implies an anechoic chamber is obtained as visualised in Figure 5.49.

The absorbers are used to obtain an anechoic chamber in a specific frequency range (in case of emission measurements for instance from 30 MHz). In case the DUT is used on Earth, the floor is often reflecting (without absorbing cones). In case the DUT is used in the air or in space, the floor also contains absorbers avoiding floor reflections.

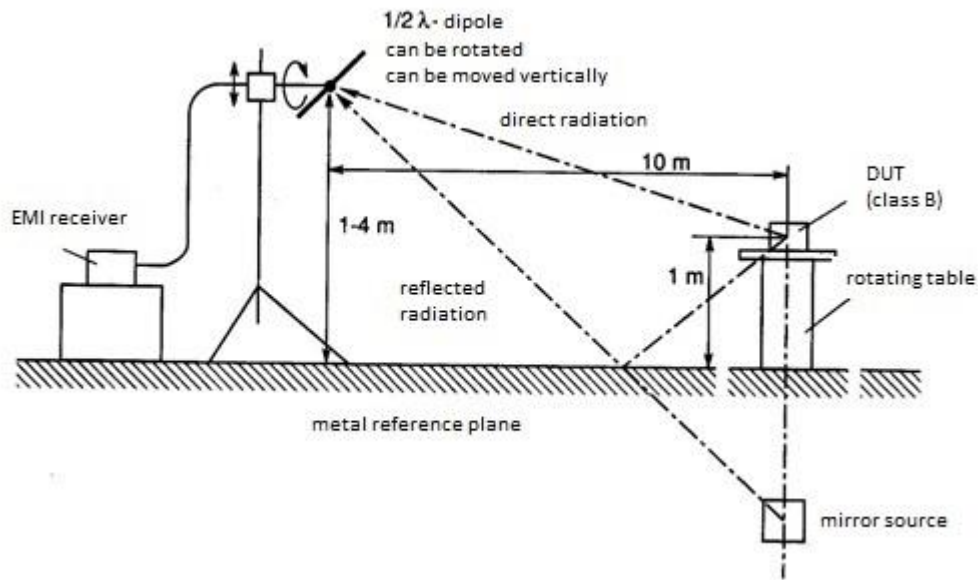


Figure 5.53. Radiated emission measurement in an open field (source: Goedbloed)

An anechoic chamber is not only used to measure the radiated emission of an electronic device. An anechoic chamber is also used to realise immunity tests. Using a generator and a transmitting antenna, electromagnetic waves are sent to the DUT to verify whether it still functions in a normal way.

#### 5.4.5 The use of shielding

The use of shielding has two effects. A shielding around an electronic device can be used to limit its radiated emission (i.e. by mounting the device in a cage of Faraday). But a shielding does not only reduce the radiated emission, it also increases the immunity against electromagnetic waves coming from outside.

In case an electromagnetic wave incides a conductive obstacle, then a part of the incident wave will be reflected ( $R_1$ ). Another part will propagate throughout the shielding and will partially be absorbed implying only a wave having a lower intensity remains ( $A_1$ ). A second reflection ( $R_2$ ) occurs implying only a much smaller wave enters the enclosure. This situation is visualised in Figure 5.54. On the right, the evolution of the electrical field strength is given but a similar evolution is obtained for the magnetic field strength.

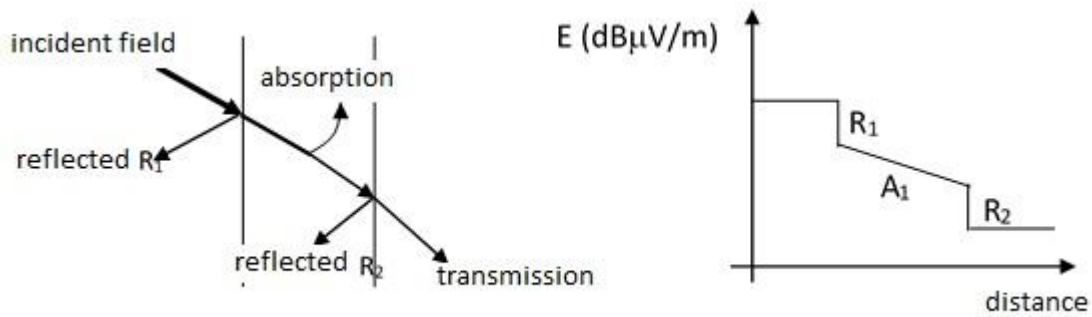


Figure 5.54. Shielding

The resulting intensity of the wave is smaller when there is more reflection and when the absorption level is higher. The amplitude of the wave propagating throughout the shielding and entering the enclosure are different when comparing electrical and magnetic fields. Table 5.2 indicates that the reflection depends on

- the frequency of the incident wave,
- the used material (e.g. aluminum, copper or steel),
- the incident wave (electrical field strength or magnetic field strength).

Table 5.2. Reflection loss occurring in case of aluminum or steel walls

Frequency	Electrical field (dB)		Magnetic field (dB)	
	aluminum	steel	aluminum	steel
60 Hz	280	240	20	0
1 kKz	240	200	30	10
150 kHz	180	130	50	20
15 MHz	115	80	75	40
100 MHz	90	60	80	60

In Table 5.2, when the absorption and the multiple reflections are not taken into consideration, the ratio between the incident field strength ( $E_i$  or  $H_i$ ) and the transmitted field strength ( $E_t$  or  $H_t$ ) is compared. More precisely,

$$20 \log \left( \frac{E_i}{E_t} \right) \quad (5.8)$$

or

$$20 \log \left( \frac{H_i}{H_t} \right) \quad (5.9)$$

is given.

When considering electrical fields, as the frequency decreases the reflection loss increases i.e.  $E_t$  decreases in case of a same  $E_i$ . It is easier to block low frequent electrical fields, it is more difficult to block high frequent electrical fields.

When considering magnetic fields, as the frequency increases, the reflection loss increases i.e.  $H_t$  decreases in case of a same  $H_i$ . It is easier to block high frequent magnetic fields, it is more difficult to block low frequent magnetic fields.

Also the absorption (expressed in  $dB/mm$ ) depends on the frequency and the used material (e.g. aluminum or steel) (magnetic). The higher the frequency, the higher the absorption.

Table 5.3. Absorption of the magnetic component of a wave

Frequency	Absorption ( $dB/mm$ )	
	aluminum	steel
60 Hz	0.02	0.33
1 kHz	0.08	1.4
150 kHz	1	17
15 MHz	10	106
100 MHz	26	137

Combining the results of Table 5.2 and Table 5.3, it is really difficult to block low frequent magnetic fields.

Magnetic fields having a low frequency are not attenuated using electrically conducting materials like aluminum (aluminum is a paramagnetic material). In such a situation, materials having a high relative permeability  $\mu_r$  are needed (for instance: Hipernom (80% Ni, 20% Fe) having a  $\mu_r = 400000$ ; Hipernik (90% Ni, 10% Fe) having a  $\mu_r = 75000$ , silicon iron having a  $\mu_r = 5000$ ).

In an ideal situation, the shielding avoids all interaction between the internal electronic device and the outer world. In reality, a high but finite shielding effectiveness is obtained or should be obtained. Moreover, a real world shielding often contains apertures in order to allow air circulation and the associated cooling. Additionally, the shielding very often contains apertures to allow power supply cables to enter the device.

In a next paragraph, the shielding of cables will be studied. Such a shielding does not only reduce the radiated emission (and has a positive effect on the radiated immunity), but it also avoids or

reduces inductive and capacitive crosstalk. Avoiding inductive and capacitive crosstalk requires an appropriate grounding of the shielding.

#### 5.4.6 Shielding of cables

##### 5.4.6.1 Crosstalk

In general, cables are an important source of disturbances due to electromagnetic waves. A cable (but the same applies for a trace on a PCB) behaves as a transmitting antenna causing radiated emission. The cable or the trace on a PCB also behaves as a receiving antenna implying also immunity issues are important.

Moreover, two conductors also cause crosstalk due to inductive and capacitive couplings. Both phenomena will be studied here.

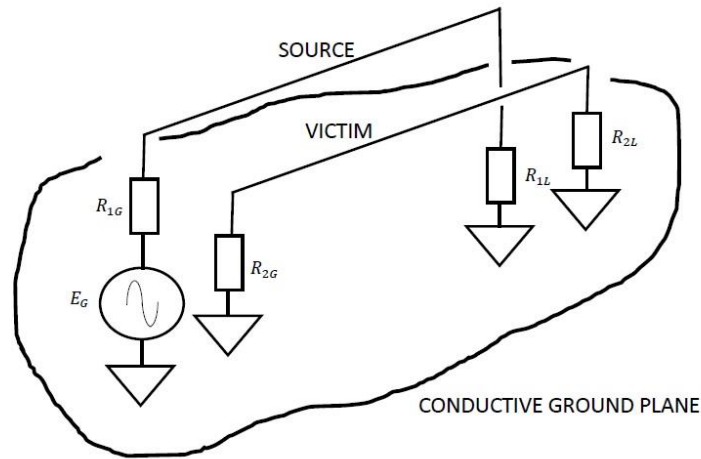


Figure 5.55. Crosstalk

Figure 5.55 shows a conductor having on one side a voltage source ( $E_G$  and  $R_{1G}$ ) and on the other side a load resistance  $R_{1L}$ . This conductor has been indicated as “source” in Figure 5.55 and causes both an electrical field and a magnetic field. These fields have an impact on the second conductor. In Figure 5.55, this second conductor has been indicated as “victim”.

Due to the current in the first conductor, there is a magnetic field. Changes in this magnetic field induce voltages in the second conductor. This behaviour is modeled using the equivalent circuit of Figure 5.56 by means of a mutual coupling  $M$  between these conductors. This magnetic coupling has an impact on the “victim”. When there is a current change  $d i(t)/dt$  in the first conductor, a voltage

$$M \frac{d i(t)}{d t} \quad (5.9)$$

is induced in the second conductor.

Besides the magnetic coupling, there is also a capacitive coupling between both conductors. This capacitive coupling originates from the capacitances  $C$  in Figure 5.56. Indeed, both conductors behave as the electrodes of a parasitic capacitance. This capacitive coupling has an impact on the second conductor i.e. on the “victim”. When there is a voltage change  $d v(t)/dt$  on the first conductor, a current

$$C \frac{d v(t)}{d t} \quad (5.10)$$

will flow.

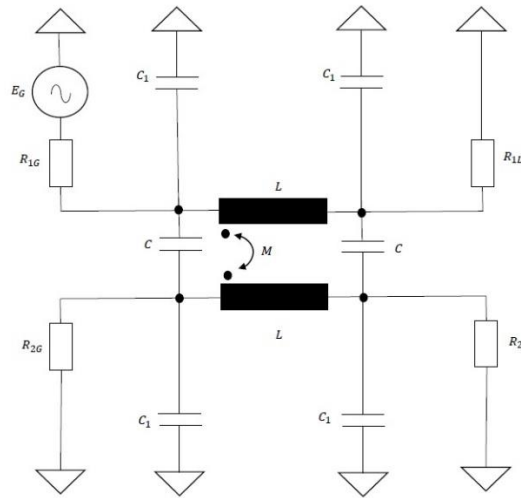


Figure 5.56. Modeling crosstalk

Notice in Figure 5.56:

- the voltage source  $E_G$ ,
- the resistors  $R_{1L}$  and  $R_{1G}$ ,
- the parasitic capacitances  $C_1$  between the “source” and the conductive ground plane,
- the inductor  $L$  of the “source” and the inductor  $L$  of the “victim”,
- the mutual inductance  $M$  between the “source” and the “victim”,
- the parasitic capacitances  $C$  between the “source” and the “victim”,
- the parasitic capacitances  $C_1$  between the “victim” and the conductive ground plane,
- the resistors  $R_{2G}$  and  $R_{2L}$ .

### 5.4.6.2 Reducing crosstalk

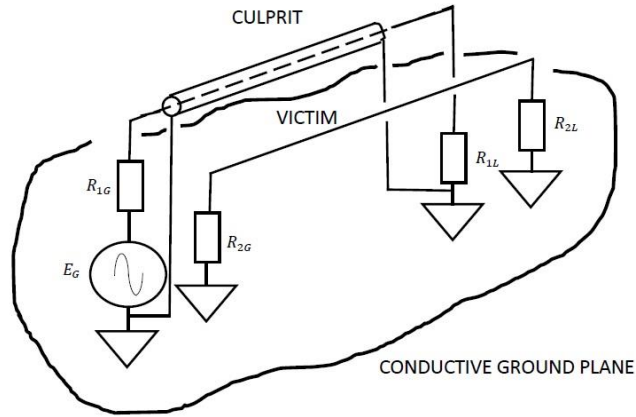


Figure 5.57. Reducing crosstalk

A first approach used to reduce crosstalk, is mounting both conductors sufficiently far removed from each other. Due to the larger distance between both conductors, the mutual coupling  $M$  and the parasitic capacitances  $C$  become smaller in Fig. 5.56.

Another approach is the use of a grounded shielding (grounded on both sides) around the first conductor (the culprit). This situation is visualised in Figure 5.57. The equivalent circuit changes from Fig. 5.56 to Fig. 5.58. Notice the additional shielding having an inductor  $L_S$  and the connection with the conductive ground plane which accounts for twice a resistance  $R_{SH}/2$ . Notice the parasitic capacitances  $C_1$  and  $C$  between “source”, shielding and “victim”.

When there is a voltage change  $d v(t)/dt$  on the first conductor, a current  $C_1 \frac{d v(t)}{d t}$ .

will flow to ground due to a small  $R_{SH}$ . This avoids (or reduces) a voltage change of the shielding implying  $C$  causes no current in the “victim” circuit. Since the shielding is connected to the conductive ground plane on both sides (implying twice a small resistance  $R_{SH}/2$ ), a current will flow in the shield which (mainly) cancels the magnetic field due to the “source” conductor. This eliminates (or reduces) the induced voltage in the “victim” conductor.

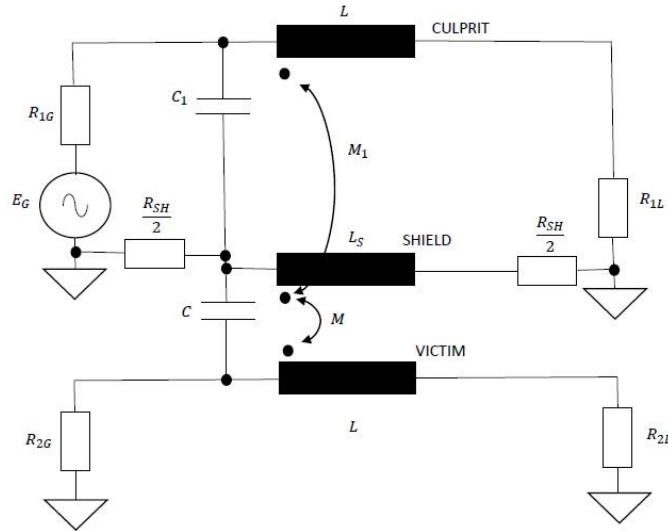


Figure 5.58. Modelling the impact of the shielding

It is also possible to realise a shielding around the “victim” instead of around the “source”. In this way, an equivalent circuit similar with the equivalent circuit of Figure 5.58 is obtained.

Here, we consider grounded shieldings in order to reduce the magnetic and the electrical coupling. Magnetic coupling can also be reduced by twisting the conductors. This twisting implies that each induced voltage component is mainly compensated by an induced voltage component having another polarity.

#### 5.4.6.3 The grounding of the shielding

In the previous subsection (Fig. 5.57.), we assumed the shielding was grounded on both sides. In reality, sometimes the shielding is only grounded on one single side. Sometimes the advice is given to ground the shielding on one single side in case of low frequent disturbances and to ground the shielding on both sides in case of high frequent disturbances. Notice however, this rule of thumb is not always correct. In order to know whether it is better to ground the shielding on one single side or on both sides, it is important to know what type of disturbance is faced.

In case the length of the shielding is smaller than  $\lambda/4$  of the disturbing signal, then one makes a distinction between the magnetic and the electrical fields. In order to obtain a shielding against electrical fields (avoiding a capacitive coupling), it is better to ground the shielding on one single side. In order to obtain a shielding against magnetic fields (avoiding an inductive mutual coupling), it is better to ground the shield on both sides which allows currents flowing in this shielding. Notice that the situation where the length of the conductor is smaller than  $\lambda/4$ , will mainly occur in case of low frequent signals. Indeed, these low frequencies account for large wavelengths  $\lambda$ .



If the length of the shielding is larger than  $\lambda/4$ , it is better to ground the shielding on both sides. It can be useful to realise two shieldings. A first shielding is grounded on one single side whereas the other shielding is grounded on both sides.

#### 5.4.7 References

1. K. Armstrong, EMC Design Techniques for electronic engineers, Nutwood UK Ltd, Cornwall, 2010.
2. J. J. Goedbloed, Elektromagnetische compatibiliteit, analyse en onderdrukking van stoorproblemen, Kluwer Technische Boeken B.V., Deventer – Antwerpen, 1990.
3. CISPR, CISPR 11: Industrial, scientific and medical (ISM) radio-frequency equipment – Electromagnetic disturbance characteristics – Limits and methodes of measurement, fourth editio, International Electrotechnical Commission IEC, 2003.
4. J. Knockaert, Inleiding tot EMC, course material KHBO.
5. H.W. Ott, Electromagnetic Compatiblity, John Wiley & Sons, 2009.
6. C.R. Paul, Introduction to Electromagnetic Compatiblity, John Wiley & Sons, 2006.
7. J. Peuteman, T. Verbeerst, P. Vansieleghe, J. Knockaert, D. Pissoort, J.-J. Vandenbussche, Reducing Electromagnetic Emitted Disturbances of an Adjustable Speed Drive System, Annual Journal of Electronics, Volume 6, Number 1: ISSN 1314-0078, pp. 116-119 (paper presented at the XXIs International Scientific and Applied Conference: Electronics – ET 2012, Sozopol, Bulgaria, September 19-21, 2012).
8. L. Tihanyi, Electromagnetic Compatibility in Power Electronics, IEEE Press, J.K. Eckert & Company Inc., Butterworth-Heinemann Ltd., Oxford, 1995.

## 5.5 Reliability and functional safety

### 5.5.1 Introduction

Due to the increasing number of electronic devices which are operating in each others vicinity, avoiding EMI is a hard task. By respecting the EMC standards on emission and immunity, the risk of EMI is significantly reduced. Notice however, the risk of EMI is never eliminated entirely.

When these electronic devices are used in e.g. health care or mobility, a failure of an electronic device accounts for serious dangers for the health and even the life of many people. This failure may originate from EMI but also other types of failures are encountered in real life. Due to this reason, it is important to integrate a functional safety approach.

A functional safety approach tries to reduce the risk of a failure. Moreover, in case of a failure the functional safety approach tries to reduce the negative impact. Unfortunately, realising a zero risk situation is impossible.

Several technical solutions exist but when considering electronic devices, the use of redundancy, the use of diversity, the use of watchdog circuits, the use of temporal separation, the use of error detection and error correction can be an option when realising this functional safety approach.

### 5.5.2 Functional safety

#### 5.5.2.1 Examples

Electric and electronic devices, but actually all devices, may cause harm to people, animals, equipment or the environment. By using appropriate safety mechanisms, it is possible to reduce the risks to a sufficiently low level.

As already mentioned, especially medical devices need to be very reliable (it is important the number of failures is sufficiently low) and the consequences of a failure must be limited. For instance in case of a malfunction of an infusion pump, an alarm must be activated to warn the medical staff. Moreover, the pump can be deactivated to avoid an overdose to the patient.

When considering mobility applications like a train or a car, functional safety is also important. It is important the doors close before the train departs and the doors are not allowed to open while the train is moving. When considering a car with an airbag, the airbag must be activated in case of a crash to protect the passengers. The airbag is not allowed to be activated when driving the car in a normal situation, even not in case of a sudden braking.

In an industrial situation, consider a roller bridge which is made to lift a maximum load. A load limiter detects a load which is too heavy and prevents from lifting this heavy load. This approach protects the entire installation against an operator which tries to lift a load which is too heavy. When using an industrial robot in a cubicle, sensors are used to detect someone entering this cubicle where the moving robot might injure him. By shutting down the robot, a protection of people is obtained.

### 5.5.2.2 The origin of failures

Failures in a device originate from different sources. In literature, a distinction is made between failures and faults. A fault is a defect within a system. Due to this fault, a failure can occur i.e. the system fails to perform the required function. It is always a challenge to design a so-called fault tolerant system. A system is fault tolerant in case a fault (or even a number of faults) do not imply a failure of the system.

To give an idea, some examples are given:

- There are software and hardware faults. These faults can occur in a systematic way or in a random way. For instance a programming error can occur in a systematic way. Hardware faults of electronic components often occur in a more random way. When considering hardware faults, aging can cause a hardware fault since the properties of the electronic parts are slowly changing with respect to time.
- Some failures originate from a human error e.g. a driver who causes a car accident or an operator who tries to lift a too heavy load with a roller bridge. A well developed tool protects the user against his own errors or the errors from another human being.
- Some faults and the associated failures originate from the environment. For instance due to high or low temperatures, the properties and the behaviour of electronic components change. The changing behaviours of the components have an impact on the behaviour of the entire electronic device.
- EMI is also an important source of failures for electronic devices. Although efforts are made to reduce emission levels and to increase immunity levels, EMI related problems are still possible. Aging of devices can have an impact but also the presence of multiple disturbances at the same time is not faced in most EMC standards.

The origins of faults, failures and errors are very diverse and the associated consequences are also very diverse. This diversity implies also the solutions are very diverse. In the next paragraphs, we will discuss a number of solutions to avoid or to reduce the malfunctioning of a device.

### 5.5.3 Reducing the malfunctioning of an electronic device: redundancy

#### 5.5.3.1 Basic approach

Suppose a digital module contains hardware components and due to random hardware faults, a module failure occurs. More precisely, an output which equals 0 is obtained in case the correct output equals 1 and an output which equals 1 is obtained in case the correct output equals 0.

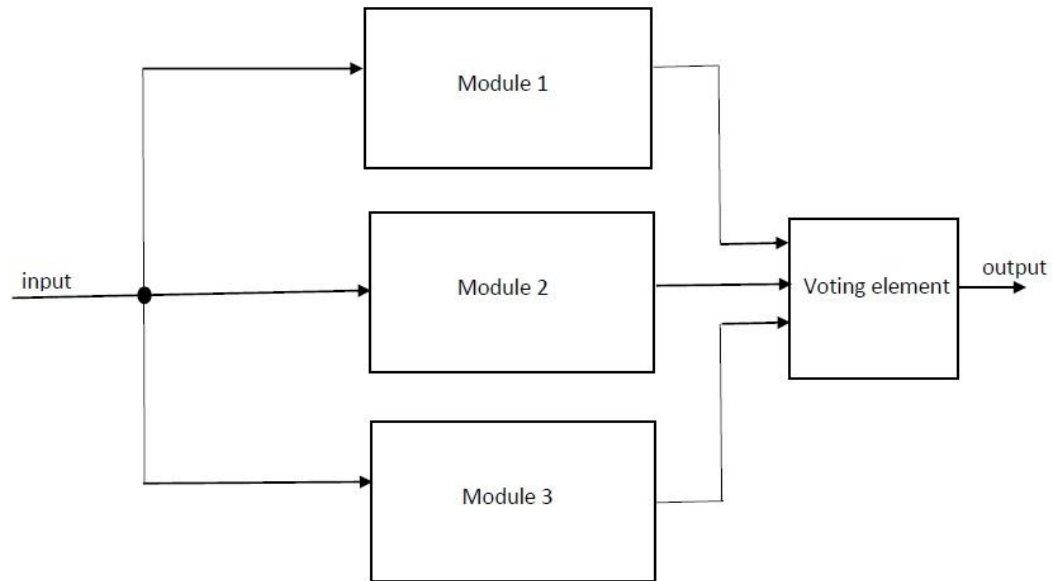


Figure 5.59. A triple modular redundant system

Static redundancy is obtained by using three identical modules in parallel each having the same input as visualised in Figure 5.59. Figure 5.59 visualises a so-called triple modular redundant system. The outputs of the three modules are sent to a voting element giving an output which corresponds with the majority of the module outputs. More precisely, the truth table of the voting element is given in Table 5.4.

Table 5.4. Truth table of the voting element

Output module 1	Output module 2	Output module 3	Output voting element
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

The approach of Figure 5.59 and Table 5.4 prevents that the failure of one single module causes a failure of the output of the voting element i.e. of the entire system. Notice this approach prevents

a failure of the entire system in case of a random fault in one single module. In case of a failure in two or three modules, the output of the voting element is incorrect. In case of a systematic fault (e.g. a software fault due to a bug in the identical software of the identical modules), the same incorrect output is obtained for all three modules implying an incorrect output of the voting element.

#### 5.5.3.2 The reliability of the redundant system

Consider the triple modular redundant system visualised in Figure 5.59. If we ignore the faults which can occur in the voting element and if we assume the input of the modules is always correct, the probability of a correct operation of the entire system can be calculated if the probability of a correct working of the individual modules are known.

The probability the entire system is working correctly equals the sum of

- the probability all three modules are working correctly,
- the probability only module 1 is failing,
- the probability only module 2 is failing,
- the probability only module 3 is failing.

Suppose the probability a module is working correctly equals  $R$ , the probability a module is failing equals  $1 - R$ . The probability the entire system is working correctly equals

$$R^3 + (1 - R)R^2 + R(1 - R)R + R^2(1 - R) = 3R^2 - 2R^3. \quad (5.11)$$

Suppose a single module has a reliability  $R = 0.9$ , the entire system has a reliability 0.972 which is significantly higher than the reliability of a single module. Notice however, a more complex system is needed which contains more hardware and consumes more power.

In case a single module has a reliability  $R = 0.3$ , the entire system has a reliability 0.216 which is lower than the reliability of a single module. This illustrates that introducing fault tolerance not always implies a larger reliability. When considering Fig. 5.59, only in case the reliability of a single module is larger than 0.5 the overall reliability increases.

#### 5.5.3.3 Generalization of the redundancy principle

The redundancy principle of Figure 5.59 can be generalised by using  $N$  modules instead of three. The reliability of the entire system can also be increased by using different modules in parallel. The approach of Figure 5.59, using identical modules, protects against random faults. Using identical modules, no protection against systematic faults is obtained since the same fault will occur in all parallel modules at the same time.

To avoid the same systematic fault in all parallel modules at the same time, the use of different modules is recommended (diversity). In case no fault occurs, the same behaviour must be obtained for all parallel modules. For instance, these behaviours can be obtained by using a first module containing only hardware and other modules containing processors which execute programs. It is possible to use modules containing different processor types. Another approach is the use of the

same processors in combination with different software (e.g. a different algorithm giving the same result when no faults occur).

Notice failures do not only originate from the modules. The sensor providing the input signal can also face a fault implying a failure of the system. It can be useful to use multiple sensors giving the same input signals in case no fault occurs. Also here, systematic faults can be avoided by using different sensor technologies. Fig. 5.60 visualises a redundant system using multiple sensors.

Since also faults can occur in the voting element (in general a voting element is not that complicated implying a high reliability), it can be useful to use multiple voting elements as visualised in Fig. 5.60. A failure of one single voting element will imply one single output failure and its effect should be removed.

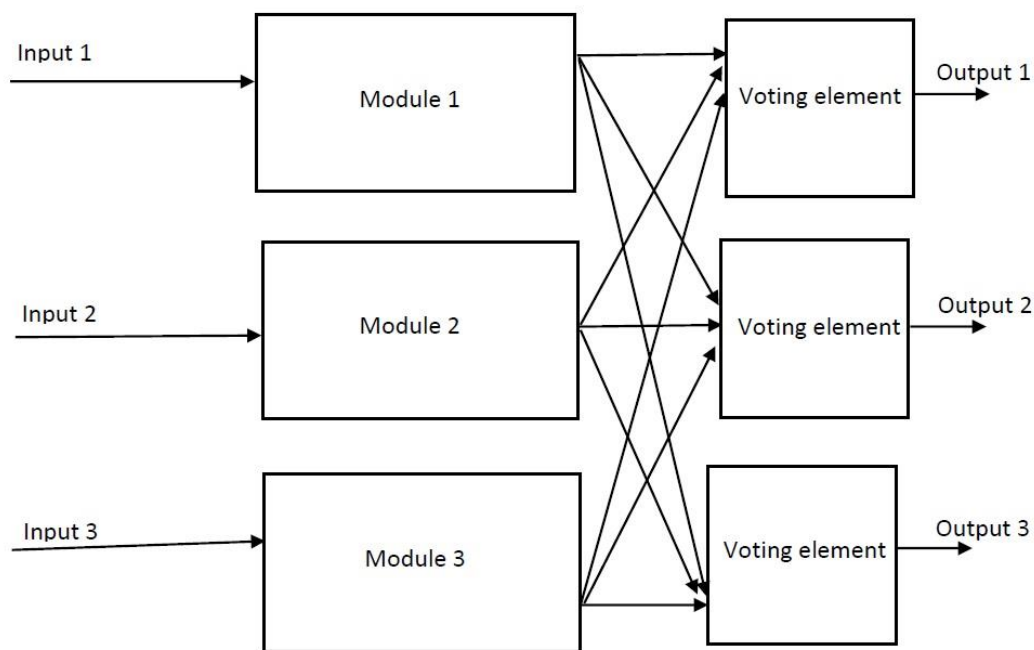


Figure 5.60. A redundant system containing redundant sensors and voting elements

#### 5.5.4 Reducing the malfunctioning of an electronic device: watchdog timer

In a normal situation, a processor executes the instructions it fetches from the memory (for instance the program counter points to the instruction which must be fetched and executed). In case the program counter contains an incorrect address, a bit combination is fetched which is not a valid instruction or the bit combination is an existing instruction which corresponds with an illegal instruction. A similar situation occurs in case code in the memory has been overwritten with data and when this data has been fetched, decoded and executed. In such situations, a 'crash of the processor' is obtained. Due to such a 'crash', the program may appear to freeze or illegal instructions are performed.

Using a watchdog timer, the impact of such a ‘crash of the processor’ can be limited. While the processor is functioning correctly, the watchdog timer is periodically loaded with an appropriate value. The watchdog timer repeatedly decrements this value. In a normal situation, the watchdog timer is reloaded with this appropriate value before the value in the watchdog timer reaches zero. This implies the watchdog timer does not send a reset instruction to the processor and the processor continues to function properly.

In case of a ‘crash of the processor’, such an appropriate value is not loaded anymore in the watchdog timer. Since the watchdog timer repeatedly decrements its value, its value will reach zero and in such a situation the watchdog timer will reset the processor. This limits the impact of the ‘crash of the processor’.

Such an approach, i.e. using a watchdog timer, has its limitations. Possibly the processor will continue to operate as long as the watchdog value did not reach a value which equals zero. During this time interval, the operation of the processor is unpredictable and hazardous situations may occur (even in a few milliseconds hazardous situations can occur). In case the ‘crash of the processor’ does not prevent the processor from loading the appropriate value in the watchdog timer repeatedly, the watchdog timer will not reset the processor.

#### 5.5.5 Reducing the malfunctioning of an electronic device: a parity bit

When digital communication takes place between for instance a CPU and a memory or a peripheral device, errors can occur. When considering a parallel bus used to transport 8 bit of data, a 0 can change into a 1 or a 1 can change into a 0. It is important to detect an error and when possible it is important to correct the error.

By using an additional parity bit, single bit errors can be detected. A distinction exists between even parity and odd parity. When using even parity, the total number of 1-bits is an even number. When using odd parity, the total number of 1-bits is an odd number. Figure 5.61 visualises an example of 8 data bits and one single parity bit using even parity. The situation is given in case no errors occur. Figure 5.61 also visualises two situations in case a single bit error occurs. First a data bit error occurs and then an error with the parity bit occurs. In both cases, the errors are detected since the total number of 1-bits is odd. Notice the errors can be detected but it is not possible to correct the errors. Notice also it is not possible to know whether an error has occurred with a data bit or with the parity bit.

Correct		1	0	0	1	1	1	0	0	0
Data bit error		1	0	1	1	1	1	0	0	0
Parity bit error		1	0	0	1	1	1	0	0	1

errors

Figure 5.61. Error detection using an even parity bit

It is important the receiver uses the same type of parity check as the transmitter (i.e. even parity or odd parity). Notice it is not possible to correct the error using a parity bit since it is not possible to detect where the error occurs (which data bit or the parity bit). As visualised in Figure 5.62, in case two bit errors occur at the same time, no error will be detected. Even parity is used and both in the correct and in the erroneous situation the number of 1-bits is even.

Correct		1	0	0	1	1	1	0	0	0
Two data bit errors		1	0	1	1	1	0	0	0	0

errors

Figure 5.62. No error detection in case of two bit errors

## 5.5.6 Reducing the malfunctioning of a device: the Hamming's algorithm

### 5.5.6.1 The Hamming distance

When adding more than one single bit in an appropriate way, it is possible to detect more complicated errors and/or to correct these errors. Consider the situation where  $m$  data bits are used and where  $r$  check bits are used (in Figure 5.61 and Figure 5.62,  $m = 8$  and  $r = 1$ ). When combining the data bits and the check bits, a codeword having a length of  $m + r$  bits is obtained.

When considering two codewords, the Hamming distance between these codewords equals the number of bit positions in which the codewords differ. For instance, Figure 5.63 visualises two code words having a Hamming distance which equals 3. By using an exor-function, a high output 1 is obtained at each bit position where the codewords differ. The number of exor-outputs which equal 1 is the Hamming distance.



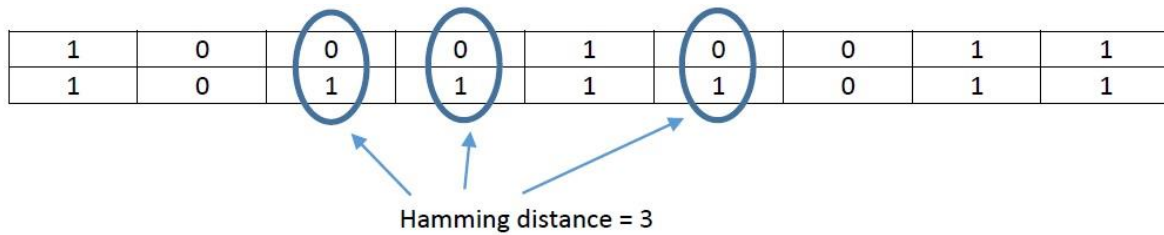


Figure 5.63. Hamming distance

### 5.5.6.2 Error detection and error correction

Due to a single bit error, a Hamming distance which equals 1 is obtained. Using a parity bit, it is possible to detect an error implying a Hamming distance which equals 1. However, it is not possible to detect errors implying larger Hamming distances.

Consider a codeword containing  $m$  data bits and  $r$  check bits giving  $n = m + r$ . There are  $2^m$  correct codewords in a total of  $2^n$  bit combinations. In order to detect an error containing  $d$  error bits, at least a Hamming distance of  $d + 1$  is needed between two valid codewords (Figure 5.64).

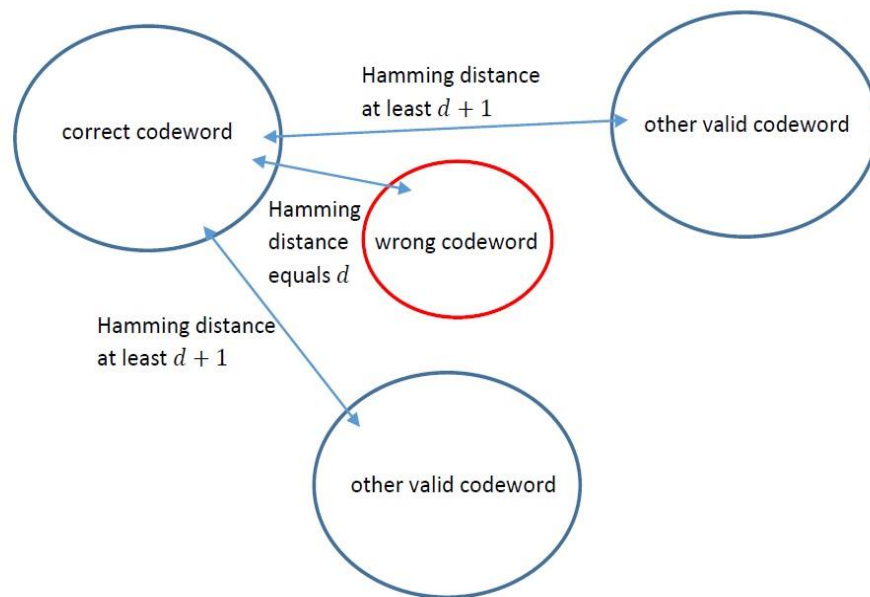


Figure 5.64. Error detection

In order to correct (and not only detect) a  $d$  error bit, a Hamming distance of at least  $2d + 1$  between two correct codewords is needed. As Visualised in Figure 5.65, when “codeword\_1” changes to “codeword\_3”, the Hamming distance between “codeword\_1” and “codeword\_3” is smaller than the Hamming distance between “codeword\_3” and “codeword\_2”. This implies “codeword\_3” can be corrected to “codeword\_1”.

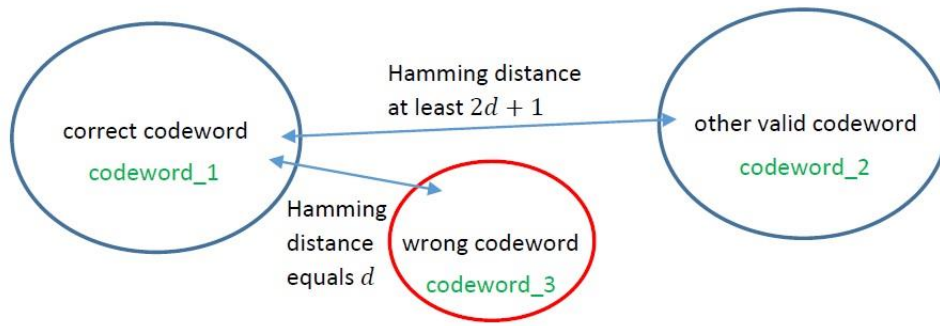


Figure 5.65. Error correction

### 5.5.6.3 The number of check bits

As an example, suppose there are four correct codewords:

00000 00000  
 00000 11111  
 11111 00000  
 11111 11111

The Hamming distances between different codewords equals at least 5. This corresponds with  $2d + 1 = 5$  and  $d = 2$ . Suppose a two bit error occurs on 00000 00000 to 11000 00000. In such a situation, it is clear 11000 00000 needs to be corrected to 00000 00000. Notice however, if a three bit error occurs on 00000 00000 to 11100 00000. In such a situation, it is clear 11100 00000 will be corrected to 11111 0000 which is wrong. Although this approach allows to detect and correct 2 bit errors, there exist much more elegant ways to realise this error detection and error correction (i.e. requiring less bits).

Consider a codeword containing  $m$  data bits. The number  $r$  of check bits which is needed to detect and correct single bit errors will be determined. There are  $2^m$  correct codewords in a total of  $2^n = 2^{m+r}$  possible bit combinations. As visualised in Figure 5.66, there are  $n$  codewords having a one single bit deviation. There are  $2^m$  such groups of  $n + 1$  bit combinations. In order to obtain disjunct groups, it is important

$$(n + 1) 2^m \leq 2^n. \quad (5.12)$$

This condition can be rewritten as

$$(m + r + 1) 2^m \leq 2^{m+r} \quad (5.13)$$

or equivalently

$$(m + r + 1) \leq 2^r. \quad (5.14)$$

For instance in case  $m = 8$ , this condition is satisfied by choosing  $r = 4$ . In order to detect and correct a single bit error in 8 data bits, 4 check bits are needed (implying a single parity bit is indeed not sufficient to correct a single error bit).

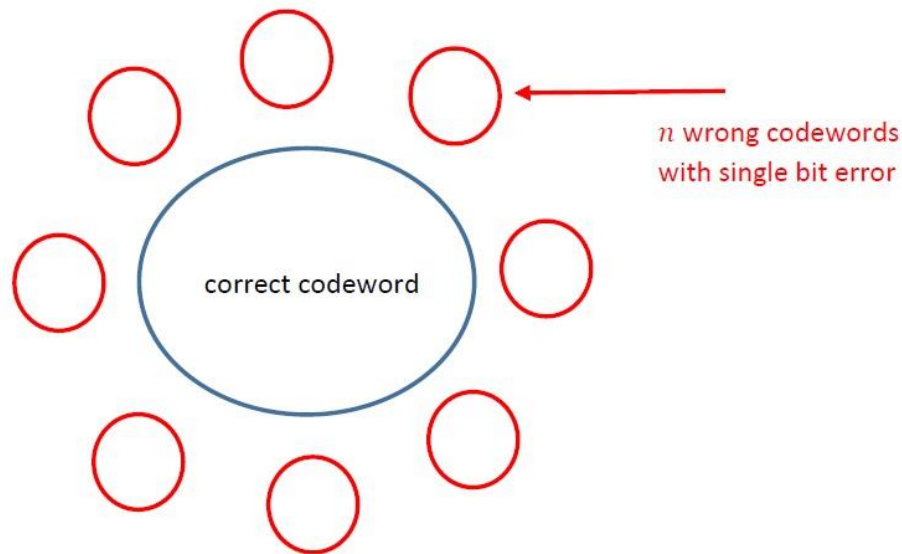


Figure 5.66. A correct codeword with  $n$  single bit errors

In case of  $m = 8$  data bits,  $r = 4$  check bits are needed to detect and correct a single bit error giving an overhead of 50%. In case  $m$  increases, the overhead decreases ( $m = 16$  requires  $r = 5$  giving an overhead of 31%;  $m = 32$  requires  $r = 6$  giving an overhead of 19%).

#### 5.5.6.4 The check bits using the Hamming's algorithm

Notice however, it is important to make an appropriate choice for these  $r$  check bits. Suppose there are  $m = 8$  data bits and all data bits are copied requiring an  $r = m = 8$ . This approach allows to detect a single bit error but it is not possible to correct such a single bit error. As illustrated in Fig. 5.67, the error can occur in the data bits but also in the additional check bits.

	Data bits								Check bits							
Correct	1	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0
Error in the data bits	0	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0
Error in the check bits	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0

Figure 5.67. Unappropriate choice of the check bits

Using the Hamming's algorithm, a much better solution exists. Consider a codeword containing  $m = 16$  data bits and  $r = 5$  check bits. The data bits are located as indicated in Figure 5.68 with a “d”. The check bits are located as indicated in Figure 5.68 with a “c”. More precisely, the check bits are located at positions numbered as  $1 = 2^0$ ,  $2 = 2^1$ ,  $4 = 2^2$ ,  $8 = 2^3$  and  $16 = 2^4$ .

c	c	d	c	d	d	d	c	d	d	d	d	d	d	d	c	d	d	d	d	d
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

Figure 5.68. Codeword containing data bits and check bits

The check bit numbered as 1 contains a 0 or a 1 in order to obtain an even number of 1-bits when considering bits numbered as 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 (even parity). The check bit numbered as 2 contains a 0 or a 1 in order to obtain an even number of 1-bits when considering bits numbered as 2, 3, 6, 7, 10, 11, 14, 15, 18, 19 (even parity). The check bit numbered as 4 contains a 0 or a 1 in order to obtain an even number of 1-bits when considering bits numbered as 4, 5, 6, 7, 12, 13, 14, 15, 20, 21 (even parity). The check bit numbered as 8 contains a 0 or a 1 in order to obtain an even number of 1-bits when considering bits numbered as 8, 9, 10, 11, 12, 13, 14, 15 (even parity). The check bit numbered as 16 contains a 0 or a 1 in order to obtain an even number of 1-bits when considering bits numbered as 16, 17, 18, 19, 20, 21 (even parity).

When considering for instance the bit located at position  $17 = 16 + 1 = 2^4 + 2^0$ . The bit located at position 17 corresponds with the check bits located at 16 and 1. In case an error occurs at the bit located at position 17, a parity error will be detected using check bit 16 and check bit 1. When considering for instance the bit located at position  $21 = 16 + 4 + 1 = 2^4 + 2^2 + 2^0$ . The bit located at position 21 corresponds with the check bits located at 16, 4 and 1. In case an error occurs at the bit located at position 21, a parity error will be detected using check bits located at 16, 4 and 1.

c	c	d	c	d	d	d	c	d	d	d	d	d	d	d	c	d	d	d	d	d
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
*	*	1	*	1	1	1	*	0	0	0	0	1	0	1	*	0	1	1	1	0

Figure 5.69. Codeword containing 16 data bits

Figure 5.69 visualises a codeword containing the 16 data bits. The check bit numbered as 1 contains a 0 in order to obtain an even number of 1-bits when considering bits numbered as 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 (even parity). The check bit numbered as 2 contains a 0 in order to obtain an even number of 1-bits when considering bits numbered as 2, 3, 6, 7, 10, 11, 14, 15, 18, 19 (even parity). The check bit numbered as 4 contains a 0 in order to obtain an even number of 1-bits when considering bits numbered as 4, 5, 6, 7, 12, 13, 14, 15, 20, 21 (even parity). The check bit numbered as 8 contains a 0 in order to obtain an even number of 1-bits when considering bits numbered as 8, 9, 10, 11, 12, 13, 14, 15 (even parity). The check bit numbered as 16 contains a 1 in order to obtain an even number of 1-bits when considering bits numbered as 16, 17, 18, 19, 20, 21 (even parity). The codeword visualised in Figure 5.70 contains no errors.

c	c	d	c	d	d	d	c	d	d	d	d	d	d	d	c	d	d	d	d	d
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	0	1	0	1	1	1	0	0	0	0	0	1	0	1	1	0	1	1	1	0

Figure 5.70. Codeword containing no errors

Suppose an error occurs at the bit numbered as 17. In this situation, a parity error will be detected using check bit 16 and check bit 1. No parity error will be detected using the other check bits 8, 4 and 2. This allows to conclude the error is indeed located at the bit numbered as 17 and since the location is known the error can be corrected. Indeed, a single bit error can be detected and corrected.

In case only one single check bit indicates a parity error, then an error occurred with this check bit. This allows to correct the check bit.

### 5.5.7 Highly Accelerated Lifetime Testing

When designing and manufacturing electronic devices, the reliability of the system is important. Not only the immunity against electromagnetic disturbances is important, also the immunity against temperature variations is important. The immunity against mechanical shocks and vibrations is also taken into account. When using an environmental chamber also the impact of humidity can be studied.

The abbreviation HALT stands for Highly Accelerated Lifetime Testing. As visualised in Figure 5.71, when realising a HALT test the EUT (Equipment Under Test) is subjected to temperature variations and mechanical shocks/vibrations far beyond the normal product operational specs. Although an electronic device has a number of product operational specs where the manufacturer guarantees a normal operation, in reality there is a larger operating margin which normally gives

no real problems. Notice however, when performing a HALT test the EUT is subjected to stress far beyond these operating margins. In case of a HALT test, the stress on the EUT is that large the EUT (or components of the EUT) are damaged/destroyed.

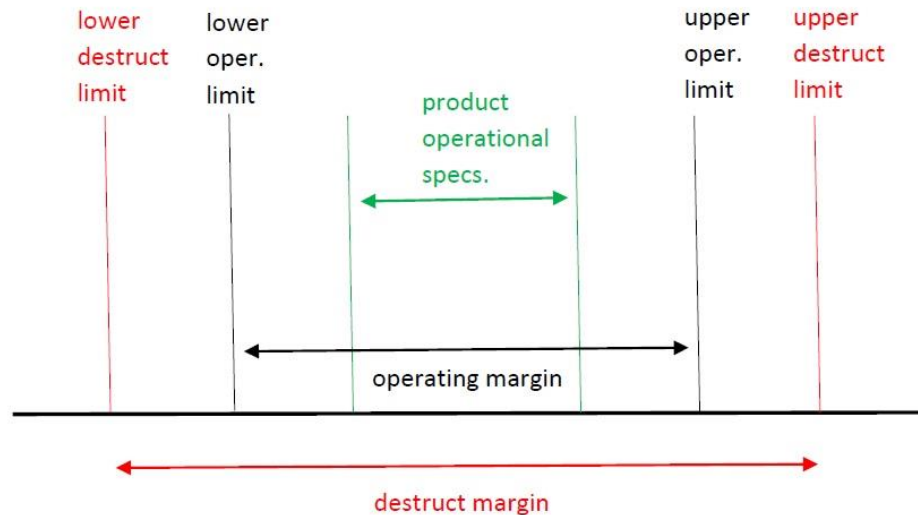


Figure 5.71. Product operational specifications, operating and destructive margins

Using a HALT chamber, an electronic device is exposed to thermal and mechanical stresses. In reality, mainly prototypes of an electronic device are exposed to stresses in order to improve the design. More precisely, the EUT is subjected to

- temperature variations,
- mechanical shocks and vibrations,
- combinations of temperature variations and mechanicals shocks/vibrations.

#### 5.5.7.1 The use of a HALT chamber

Fig. 5.72 visualises the HALT chamber at the KULeuven Technologicampus at Ostend (Belgium). Using such a HALT chamber, it is possible to realise combinations of thermal and mechanical stresses as visualised in Fig. 5.75. It is realistic to realise temperature variations between  $-50^{\circ}\text{C}$  and  $+120^{\circ}\text{C}$ . Using liquid nitrogen, it is possible to realise such low temperatures. Figure 5.73 visualises the liquid nitrogen tank at the KULeuven Technologicampus at Ostend (Belgium).





Figure 5.72. HALT chamber



Figure 5.73. Liquid nitrogen tank

Using the HALT chamber of Fig. 5.72 it is not only possible to realise temperature variations between  $-50^{\circ}\text{C}$  and  $+120^{\circ}\text{C}$ , the generation of random vibrations are obtained up to 25 Grms.

In such a HALT chamber, the EUT is subjected to a continuously varying pseudo random broad spectrum vibration. A typical time signal of the acceleration is visualised in Figure 5.74 (source: N. Doertenbach, The Calculation of Grms, QualMark Corp., retrieved from [http://www.dfrsolutions.com/uploads/services/HALT\\_grms\\_calculation\\_ndoertenbach.pdf](http://www.dfrsolutions.com/uploads/services/HALT_grms_calculation_ndoertenbach.pdf)). The

acceleration is expressed as a multiple of the acceleration of gravity  $g$ . In case this acceleration  $a(t)$  is measured using an accelerometer, the measured value can be squared at each time instant giving  $a^2(t)$ . By taking the square root of the average of this squared value gives the rms value of the acceleration (expressed in Grms).

Using the time signal of Fig. 5.74, the rms value of the acceleration can be calculated. Often, the rms value of the acceleration is also calculated using the Power Spectral Density Curve and using Parseval's Theorem stating that

$$\int_{-\infty}^{+\infty} a^2(t) dt = \int_{-\infty}^{+\infty} |A(f)|^2 df . \quad (5.15)$$

Here,  $a(t)$  is the acceleration as a function of time (expressed as a multiple of  $g$ ) and  $|A(f)|^2$  is the Power Spectral Density (expressed as  $g^2/Hz$ ).

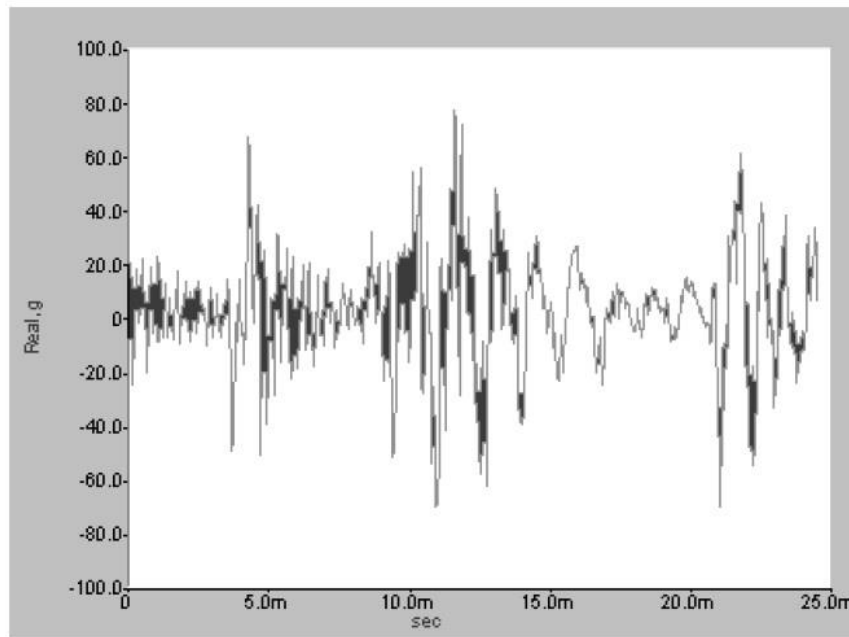


Figure 5.74. Acceleration time signal in a HALT chamber (source: Doertenbach)

#### 5.5.7.2 Performing the HALT test

A typical HALT test is performed as visualised in Figure 5.75. First, a cold step stress test is performed by performing a stepwise decrease of the temperature level. This cold step stress test is followed by a hot step stress test by performing a stepwise increase of the temperature level. The hot step stress test is followed by a number of rapid thermal transitions. During these three tests, the EUT is only subjected to temperature variations (blue lines in Figure 5.75).



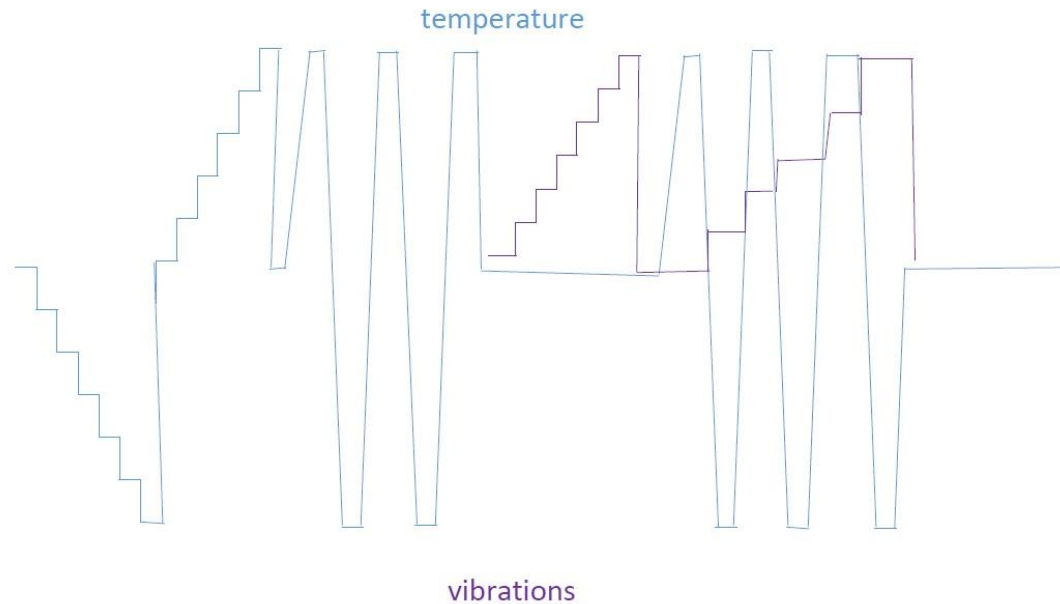


Figure 5.75. Performing the HALT test

Having finished the temperature stress tests, a vibration step stress test is performed by stepwise increasing the rms value of the vibrations (purple lines in Figure 5.75). During this first vibration step stress test, the temperature remains constant. Finally, a combination of rapid thermal transitions and a stepwise increasing rms value of the vibrations is applied.

Such an approach allows to detect the weak parts of the Equipment Under Test. For instance a lift of SMD components, a cracking of components, a breaking or loosening of the lead of a component (especially capacitors and electrolytic capacitors are vulnerable) can be detected. Using this information an improved design is obtained and this improved design will be tested again providing new information about the remaining weak parts. By repeatedly performing HALT tests and improving the design, a reliable final design is obtained.

#### 5.5.8 References

- 1.. Degraeve, personal communication.
2. K.R. Bhote, A.K. Bhote, World Class Reliability: Using Multiple Environment Overstress Tests to Make It Happen, American Management Association, New York, 2004.
3. N. Doertenbach, The Calculation of Grms, QualMark Corp., retrieved from website [http://www.dfrsolutions.com/uploads/services/HALT\\_grms\\_calculation\\_ndoertenbach.pdf](http://www.dfrsolutions.com/uploads/services/HALT_grms_calculation_ndoertenbach.pdf)
4. IEC, Functional Safety: Essential to overall safety, retrieved from website [http://www.iec.ch/about/brochures/pdf/technology/functional\\_safety.pdf](http://www.iec.ch/about/brochures/pdf/technology/functional_safety.pdf)
5. H.W. Mc Clean, HALT, HASS, and HASA explained: Accelerated Reliability Techniques, ASQ Quality Press, Milwaukee, USA, 2009.

6. N. Storey, Safety-Critical Computer Systems, Pearson – Prentice Hall, London, England, 1996.
7. A.S. Tanenbaum, T. Austin, Structured Computer Organization, Pearson, London, England, 6th edition, 2013.
8. I. Vervenne, G. Deconinck, Reliability assessment of electronic equipment on system level, Annual Journal of Electronics, Volume 5, ISSN 1313-1842 (paper presented at the XXth International Scientific and Applied Science Conference: Electronics – ET 2011, Sozopol, Bulgaria, September 14-16, 2011).

## **6 APPLIED SYSTEM THEORY**

## 6.1 System theory

### 6.1.1 Introduction

System theory is a science which describes and studies the behaviour of a very broad range of systems. In reality most systems are very complicated and very often only a number of aspects of the system will be studied. Actually models are used which approximate the behaviour of the real system. Fig. 6.1 visualises the approach where starting from the physical reality an idealized physical approximation is considered. Based on this physical approximation, the mathematical model is constructed which allows to calculate an approximation of the behaviour of the original physical reality. The behaviour of the physical reality can be approximated and predicted.

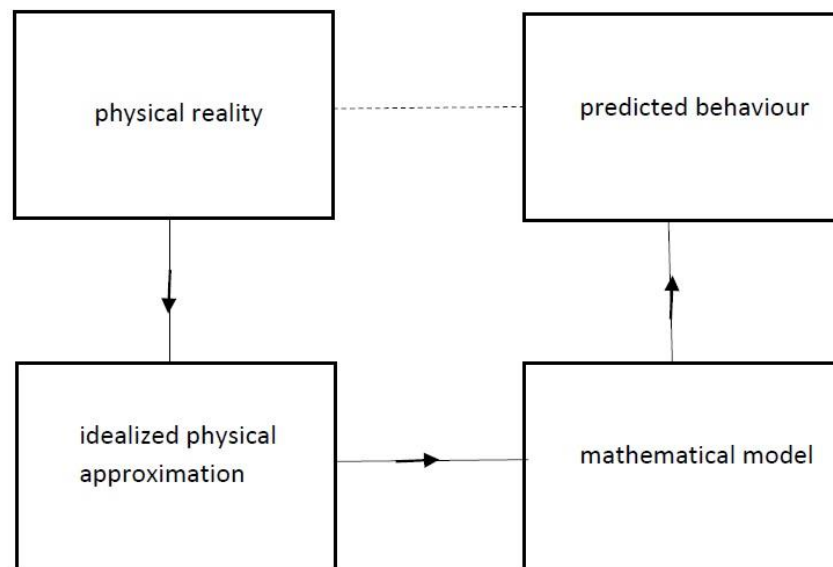


Figure 6.1. Modelling a physical reality

Quite often, a compromise is needed between accuracy and complexity. When increasing the accuracy of the model it can become more complicated. When reducing the complexity of the model a decrease of the accuracy can occur.

Using appropriate models, all types of systems can be studied. System theory is used in biology, physics, engineering, social sciences, economy, sociology,... When focusing on physics and engineering sciences, system theory is used to model mechanical systems, electrical systems, control systems, ...

### 6.1.2 The behaviour of a pendulum

Differential equations are very often used to describe the behaviour of a “physical reality”. To give an idea, consider the pendulum visualised in Fig. 6.2. A spherical bob having a mass  $m$  is attached at the end of a rod with a length  $l$ . Assume the rod is rigid and the rod has zero mass. The pendulum

is free to swing in the vertical plane and the pivot is fixed. The behaviour of the pendulum originates from the forces acting on the bob.

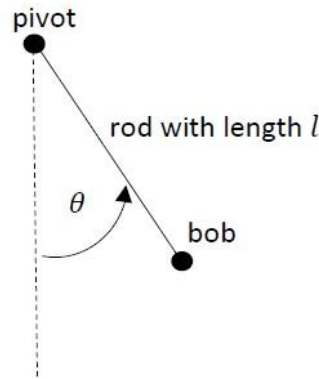


Figure 6.2. Pendulum

The gravitation exerts a vertical force  $mg$  on the bob ( $g$  is the acceleration due to gravity). There is a frictional force which is assumed to be proportional with the speed of the bob. The coefficient of friction equals  $k$ .

#### 6.1.2.1 Differential equation

Based on Newton's second law of motion in the tangential direction (perpendicular with the rod),

$$ml \ddot{\theta} = -mg \sin\theta - kl \dot{\theta} . \quad (6.1)$$

Here,  $\ddot{\theta}$  denotes the second derivative with respect to time of  $\theta(t)$  and  $\dot{\theta}$  denotes the first derivative with respect to time of  $\theta(t)$ . A positive  $\dot{\theta}$  implies the angle  $\theta$  increases i.e. the bob is moving upwards. In case  $mg \sin\theta > 0$  (with  $\theta > 0$  as visualised in Fig. 6.2) and  $kl \dot{\theta} > 0$  (in case  $\dot{\theta} > 0$  the bob is moving upwards), due to gravity and friction  $\ddot{\theta} < 0$  implying the angular velocity is decreasing. By considering the tangential direction (perpendicular with the rod), the force exerted by the rod does not appear in the equation.

The differential equation describing the behaviour of the pendulum is a nonlinear differential equation (due to the  $\sin\theta$  term). In general, solving nonlinear differential equations is hard. By linearizing the differential equation, a linear differential equation  $ml \ddot{\theta} = -mg \theta - kl \dot{\theta}$  is obtained. In case the angle  $\theta$  is small, the behaviour of the linear differential equations approximates the behaviour of the nonlinear differential equation. Solving a linear differential equation with constant coefficients (a linear time-invariant differential equation) is not really hard since for instance Laplace transforms can be used.

### 6.1.2.2 The use of a Laplace transform

By taking the Laplace transform of the differential equation  $ml \ddot{\theta}(t) = -mg \theta(t) - kl \dot{\theta}(t)$  one obtains the expression

$$ml [s^2 \theta(s) - s \theta(t_0) - \theta'(t_0)] = -kl [s \theta(s) - \theta(t_0)] - mg \theta(s). \quad (6.2)$$

Suppose the pendulum starts at  $t = t_0$  having a velocity zero (i.e.  $\theta'(t_0) = 0$ ) and a position  $\theta(t_0) = \theta_0$ . This implies the expression in the Laplace domain reduces to  $[ml s^2 + kl s + mg] \theta(s) = l(m s + k) \theta_0$

or equivalently

$$\theta(s) = \frac{l(m s + k) \theta_0}{ml s^2 + kl s + mg} \quad (6.3)$$

In case the coefficient of friction  $k = 0$ , one obtains that  $\theta(s) = \frac{s}{s^2 + \frac{g}{l}} \theta_0$ .

By taking the inverse Laplace transform, one obtains an undamped oscillation giving

$$\theta(t) = \theta_0 \cos\left(\sqrt{\frac{g}{l}}(t - t_0)\right) \mu(t - t_0). \quad (6.4)$$

Here,  $\mu(t - t_0)$  is a unity step which equals 1 in case  $t \geq t_0$  and which equals 0 in case  $t < t_0$ . Notice the pulsation of the oscillation depends on  $g$  and  $l$  but does not depend on the mass  $m$  of the bob.

In case the coefficient of friction  $k$  is not zero, after some calculations one obtains that

$$\theta(s) = \frac{\left(s + \frac{k}{2m}\right)\theta_0 + \frac{k}{2m}\theta_0}{\left(s + \frac{k}{2m}\right)^2 + \frac{g}{l} - \frac{k^2}{4m^2}}. \quad (6.5)$$

By defining

$$\omega^2 = \frac{g}{l} - \frac{k^2}{4m^2}, \quad (6.6)$$

one obtains that

$$\theta(s) = \frac{\left(s + \frac{k}{2m}\right)\theta_0}{\left(s + \frac{k}{2m}\right)^2 + \omega^2} + \frac{\frac{k}{2m}\theta_0 \omega}{\left(s + \frac{k}{2m}\right)^2 + \omega^2} \frac{1}{\omega}. \quad (6.7)$$

By taking the inverse Laplace transform, a damped oscillation is obtained. More precisely,

$$\theta(t) = \theta_0 e^{-\frac{k}{2m}(t-t_0)} \left( \cos(\omega(t-t_0)) + \frac{k}{2m} \frac{1}{\omega} \sin(\omega(t-t_0)) \right) \mu(t-t_0). \quad (6.8)$$

Verify this expression for  $\theta(t)$  is indeed a solution of the differential equation  $ml \ddot{\theta}(t) = -mg \theta(t) - kl \dot{\theta}(t)$ .

### 6.1.3 Electrical systems

In the previous paragraph, a pendulum (which is a mechanical application) has been studied. The behaviour has been described using a differential equation and in case of a linear time-invariant differential equation, the solution can be calculated using a Laplace transform. A similar approach also applies to electrical networks containing resistors, inductors and capacitors. First we discuss the behaviour of these basic network elements and then they will be combined to study a series RLC circuit.

#### 6.1.3.1 Behaviour of resistors, inductors and capacitors

Resistors, inductors and capacitors are the basic building components of electrical systems. When a voltage  $E_R(t)$  is applied to a resistor  $R$ , the relationship between  $E_R(t)$  and the current  $I(t)$  flowing through the resistor can be written as

$$E_R(t) = R I(t). \quad (6.9)$$

In order to study the behaviour of a component, also its description in the Laplace domain can be used. Here, for a resistor

$$E_R(s) = R I(s). \quad (6.10)$$

When a voltage  $E_L(t)$  is applied to an inductor  $L$ , the relationship between  $E_L(t)$  and the current  $I(t)$  flowing through the inductor can be written as

$$E_L(t) = L \frac{dI(t)}{dt}. \quad (6.11)$$

The behaviour in the Laplace domain of the inductor (with  $I(0^-) = 0$ ) can be written as  $E_L(s) = sL I(s)$ .

When a voltage  $E_C(t)$  is applied to a capacitor  $C$ , the relationship between  $E_C(t)$  and the current  $I(t)$  flowing through the capacitor can be written as

$$E_C(t) = \frac{1}{C} \int_{0^-}^t I(\tau) d\tau + E_C(0^-). \quad (6.12)$$

The behaviour in the Laplace domain of the capacitor (with  $E_C(0^-) = 0$ ) can be written as

$$E_C(s) = \frac{1}{sC} I(s). \quad (6.13)$$

It is important to realise that these expressions in the Laplace domain are only valid in case the resistor value  $R$ , the inductor value  $L$  and the capacitor value  $C$  are time invariant and linear. The  $R$ ,  $L$  and  $C$  values are linear when they do not depend on the voltage across it and they do not depend on the current flowing through it.

### 6.1.3.2 Series electrical systems

When considering a resistor  $R$ , it has an impedance  $Z(s) = R$  in the Laplace domain. An inductor  $L$  has an impedance  $Z(s) = sL$  and a capacitor has an impedance  $Z(s) = 1/sC$ . Each time, the relationship  $E(s) = Z(s) I(s)$  is valid.

When considering  $n$  series elements having impedances  $Z_i(s)$  with  $i \in \{1, \dots, n\}$ , the total impedance equals  $Z(s) = Z_1(s) + \dots + Z_n(s)$  and  $E(s) = Z(s) I(s)$ . Here,  $E(s)$  is the total voltage and  $I(s)$  is the current through the series circuit.

### 6.1.3.3 Behaviour in the Laplace domain of a series RLC circuit

For the series RLC circuit shown in Fig. 6.3, the total voltage drop  $E(t)$  can be written as

$$E(t) = E_R(t) + E_L(t) + E_C(t) = R I(t) + L \frac{dI(t)}{dt} + \frac{1}{C} \int_{0^-}^t I(\tau) d\tau + E_C(0^-). \quad (6.14)$$

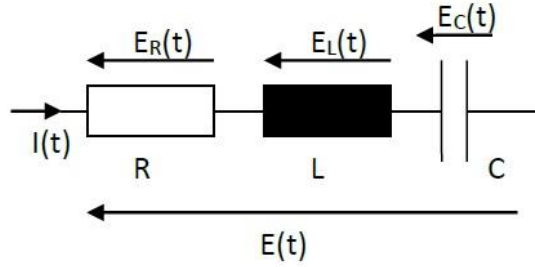


Figure 6.3. Series RLC circuit

By taking the Laplace transform (with  $I(0^-) = 0$  and  $E_C(0^-) = 0$ ),

$$E(s) = \left( R + sL + \frac{1}{sC} \right) I(s) = \left( s^2 L + sR + \frac{1}{C} \right) Q(s). \quad (6.15)$$

Here,  $Q$  is the charge available on the electrodes of the capacitor. Since the voltage  $E_C(0^-) = 0$ , also  $Q(0^-) = 0$ . In the time and Laplace domain, we can write  $Q(t) = \int_{0^-}^t I(\tau) d\tau$ ,  $Q(s) = \frac{I(s)}{s}$ .

The RLC series circuit can be seen as a second order system with input  $E(t)$  and output  $Q(t)$ . In the Laplace domain, the second order system is described by its transfer function

$$\frac{Q(s)}{E(s)} = \frac{1}{s^2 L + sR + 1/C}. \quad (6.16)$$

Notice the RLC circuit is a second order system since the system contains two “energy reservoirs” i.e. energy is stored in the capacitor when there is a voltage across this capacitor and energy is



stored in the inductor when a current is flowing through this inductor. Also the pendulum, which has been discussed earlier, is a second order system because a moving pendulum has potential energy (depending on the position of the bob) and kinetic energy (depending on the speed of the bob).

#### 6.1.3.4 Exercise

Consider the series RLC circuit of Figure 6.3 with zero initial conditions i.e.  $E_C(0^-) = 0$  (or equivalently  $Q(0^-) = 0$ ) and  $I(0^-) = 0$ . In case the externally applied voltage  $E(t) = \delta(t)$  ( $\delta(t)$  is a dirac impulse), calculate  $Q(t)$ .

Since  $E(s) = 1$ ,

$$Q(s) = \frac{1}{s^2 L + sR + \frac{1}{C}} = \frac{1}{L} \frac{1}{\left(s + \frac{R}{2L} - \frac{D}{2L}\right) \left(s + \frac{R}{2L} + \frac{D}{2L}\right)} \quad (6.17)$$

With  $D = \sqrt{D^2}$  and  $D^2 = R^2 - \frac{4L}{C}$ .

There are three important situations. In case  $D^2 = 0$ , the series RLC circuit is critically damped. This implies that

$$Q(s) = \frac{1}{L} \frac{1}{\left(s + \frac{R}{2L}\right)^2} \quad (6.18)$$

and that

$$Q(t) = \frac{1}{L} t e^{-\frac{R}{2L}t} \mu(t). \quad (6.19)$$

In case  $D^2 \neq 0$ , by partial fraction decomposition

$$Q(s) = \frac{1}{D} \left( \frac{1}{s + \frac{R}{2L} - \frac{D}{2L}} - \frac{1}{s + \frac{R}{2L} + \frac{D}{2L}} \right). \quad (6.20)$$

In case  $D^2 > 0$ , the series RLC circuit is overdamped. Here,  $D > 0$  and  $D$  is real valued. This implies that

$$Q(t) = \frac{1}{D} \left( e^{\frac{D-R}{2L}t} - e^{\frac{-D-R}{2L}t} \right) = \frac{2}{D} e^{\frac{-R}{2L}t} \sinh\left(\frac{D}{2L}t\right). \quad (6.21)$$

In case  $D^2 < 0$ , the series RCL circuit is underdamped. Here,  $D = \sqrt{D^2} = \sqrt{-|D^2|} = j\sqrt{|D^2|}$  and

$$Q(t) = \frac{2}{D} e^{\frac{-R}{2L}t} \sinh\left(\frac{j\sqrt{|D^2|}}{2L}t\right) = \frac{2j}{D} e^{\frac{-R}{2L}t} \sin\left(\frac{\sqrt{|D^2|}}{2L}t\right) = \frac{2}{\sqrt{|D^2|}} e^{\frac{-R}{2L}t} \sin\left(\frac{\sqrt{|D^2|}}{2L}t\right). \quad (6.22)$$

Consider a series RLC circuit with  $L = 0.1$  H and  $C = 100$   $\mu$ F. Calculate the  $R$  value making the RLC circuit critically damped. Use Matlab to calculate and visualise the evolution of  $Q(t)$ .

Double the  $R$  value in comparison with the critically damped circuit. Calculate and visualise  $Q(t)$  with the same  $L$  and  $C$  giving an overdamped RLC circuit.

Halve the  $R$  value in comparison with the critically damped circuit. Calculate and visualise  $Q(t)$  with the same  $L$  and  $C$  giving an underdamped RLC circuit.

#### 6.1.3.5 Transient behaviour due to an initial condition

The calculations in the previous paragraphs considered the behaviour of a pendulum and a series RLC circuit due to an initial condition (the dirac impulse actually implies an initial condition). In case of the pendulum, the initial condition includes the initial position and the initial speed. In case of the series RLC circuit, the initial condition includes the voltage across the capacitor and the current flowing through the inductor. Using differential equations, it is also possible to calculate the behaviour in case an external force is applied to the pendulum or in case an external voltage is applied to the series RLC circuit.

#### 6.1.4 The behaviour of a series RLC circuit

Consider the series RLC circuit of Fig. 6.4 containing linear time-invariant components  $R$ ,  $L$  and  $C$ . At  $t = 0$ , the switch  $S$  closes which implies the current  $i(t) = 0$  for all  $t < 0$ . By closing the switch  $S$ , an external voltage  $u(t) = \sqrt{2} U \cos(\omega t)$  is applied. Using and solving a differential equation, it is possible to calculate the current  $i(t)$  for all  $t \geq 0$ .

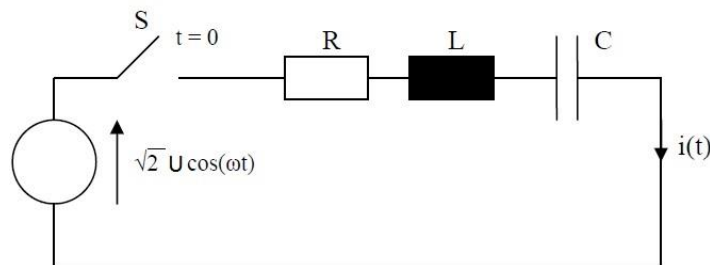


Figure 6.4. Series RLC circuit

The behaviour of the series RLC circuit is given by the integro-differential equation

$$\sqrt{2} U \cos(\omega t) \mu(t) = R i(t) + L \frac{di(t)}{dt} + \frac{1}{C} \int_{0^-}^t i(\tau) d\tau + v_C(0^-). \quad (6.22)$$

#### 6.1.4.1 The use of a Laplace transform

By taking the Laplace transform of the integro-differential equation, one obtains that

$$\sqrt{2} U \frac{s}{s^2 + \omega^2} = R I(s) + sL I(s) - L i(0^-) + \frac{1}{sC} I(s) + \frac{v_C(0^-)}{s}. \quad (6.23)$$

Since the switch is open for all  $t < 0$ ,  $i(0^-) = 0$ . Suppose  $v_C(0^-) = \frac{U}{2}$  where  $U$  equals 10 volts. After some calculations,

$$I(s) = \frac{\sqrt{2} U}{L} \frac{s^2}{(s^2 + \omega^2) \left( s^2 + s \frac{R}{L} + \frac{1}{LC} \right)} - \frac{v_C(0^-)}{L} \frac{1}{\left( s^2 + s \frac{R}{L} + \frac{1}{LC} \right)} \quad (6.24)$$

In case  $R = 50\Omega$ ,  $L = 1mH$ ,  $C = 1\mu F$  with  $\omega = 2\pi f$  (and  $f = 2 kHz$ ), one obtains that  $I(s) = \frac{(0.101s - 0.1356\omega)}{(s^2 + \omega^2)} - \frac{(0.101s - 5788)}{\left( s^2 + s \frac{R}{L} + \frac{1}{LC} \right)}$ .

By defining

$$\bar{\omega}^2 = \frac{1}{LC} - \frac{R^2}{4L^2} = 19365^2, \quad (6.25)$$

a current

$$I(s) = \frac{(0.101s - 0.1356\omega)}{(s^2 + \omega^2)} - \frac{0.101(s - 57307)}{\left( \left( s + \frac{R}{2L} \right)^2 + 19365^2 \right)}. \quad (6.26)$$

is obtained with  $\frac{R}{2L} = 25000$ .

Some basic calculations reveal that

$$I(s) = \frac{0.101s}{s^2 + \omega^2} - \frac{0.1356\omega}{s^2 + \omega^2} - \frac{0.101(s + 25000)}{(s + 25000)^2 + 19365^2} + \frac{0.4293 \cdot 19365}{(s + 25000)^2 + 19365^2} \quad (6.27)$$

and taking the inverse Laplace transform the current  $i(t)$  is the sum of a steady state term  $i_1(t)$  and a transient term  $i_2(t)$ . More precisely,

$$i(t) = i_1(t) + i_2(t) \quad (6.28)$$

with

$$i_1(t) = 0.101 \cos(\omega t) \mu(t) - 0.1356 \sin(\omega t) \mu(t) \quad (6.29)$$

( $\omega = 12566 \text{ rad/s}$ ) and

$$i_2(t) = -0.101 e^{-25000t} \cos(19365 t) \mu(t) + 0.4293 e^{-25000t} \sin(19365 t) \mu(t). \quad (6.30)$$

#### 6.1.4.2 Transient and steady state behaviour

The transient term  $i_2(t)$  originates from the initial condition  $v_C(0^-)$ . Due to the heat dissipation in the resistor,  $i_2(t)$  is damped and goes to zero. Notice the pulsation

$$\bar{\omega} = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = 19365 \text{ rad/s} \quad (6.31)$$

does not equal the resonance pulsation  $\frac{1}{\sqrt{LC}} = 31623 \text{ rad/s}$  nor the applied pulsation  $\omega = 2\pi \cdot 2000 = 12566 \text{ rad/s}$ .

The current  $i(t)$  also contains a steady state term  $i_1(t)$  having a pulsation  $\omega = 12566 \text{ rad/s}$ . Notice this steady state term can also be calculated using complex numbers. The use of complex numbers to study the behaviour in an electrical circuit is only valid in case of sinusoidal steady state voltages and currents. Using complex numbers, no transient behaviour can be studied. In case non sinusoidal voltages are applied, Laplace transforms are needed i.e. the electrical circuit can not be studied using complex numbers.

#### 6.1.5 The use of complex numbers

Sinusoidal signals (for instance voltages or currents)  $\sqrt{2} A \sin(\omega t + \varphi)$  can be represented by a complex number  $A e^{j\varphi}$ . Notice this complex number reveals the RMS value  $A$  and the phase  $\varphi$ .

##### 6.1.5.1 Behaviour of resistors, inductors and capacitors

When a voltage  $E_R(t) = \sqrt{2}U \sin(\omega t)$  is applied to a resistor  $R$ , the relationship between  $E_R(t)$  and the current  $I(t) = \sqrt{2} I \sin(\omega t)$  flowing through the resistor can be written as

$$E_R(t) = \sqrt{2} U \sin(\omega t) = R \sqrt{2} I \sin(\omega t). \quad (6.32)$$

By representing these sinusoidal signals in the complex domain, one obtains that

$$U = R I. \quad (6.33)$$

This means the resistor  $R$  can be represented by the real valued complex number  $R$ .

When a voltage  $E_L(t) = \sqrt{2}U \sin(\omega t)$  is applied to an inductor  $L$ , the relationship between  $E_L(t)$  and the current  $I(t)$  flowing through the inductor can be written as  $E_L(t) = L \frac{dI(t)}{dt}$ .

The current

$$I(t) = \frac{\sqrt{2} U}{\omega L} \sin\left(\omega t - \frac{\pi}{2}\right) \quad (6.34)$$

gives, in the complex domain, rise to the relationship

$$\bar{I} = \frac{U}{\omega L} e^{-j\frac{\pi}{2}} = -j \frac{U}{\omega L} = \frac{U}{j\omega L}. \quad (6.35)$$

This means the impedance of the inductor  $L$  equals  $j\omega L$ .

When a voltage  $E_C(t) = \sqrt{2}U \sin(\omega t)$  is applied to a capacitor  $C$ , the relationship between  $E_C(t)$  and the current  $I(t)$  flowing through the capacitor can be written as

$$I(t) = C \frac{d E_C(t)}{dt} \quad (6.36)$$

giving a current

$$I(t) = U\omega C \sin\left(\omega t + \frac{\pi}{2}\right). \quad (6.37)$$

Using the complex representation, the relationship

$$\bar{I} = U\omega C e^{+j\frac{\pi}{2}} = j\omega C U. \quad (6.38)$$

This means the impedance of the capacitor  $C$  equals  $\frac{-j}{\omega C} = \frac{1}{j\omega C}$ .

#### 6.1.5.2 Series electrical systems

When considering a resistor  $R$ , it has an impedance  $Z(j\omega) = R$  in the complex domain. An inductor  $L$  has an impedance  $Z(j\omega) = j\omega L$  and a capacitor has an impedance  $Z(j\omega) = 1/j\omega C$ . Each time, the relationship  $E(j\omega) = Z(j\omega) I(j\omega)$  is valid.

When considering  $n$  series elements having impedances  $Z_i(j\omega)$  with  $i \in \{1, \dots, n\}$ , the total impedance equals  $Z(j\omega) = Z_1(j\omega) + \dots + Z_n(j\omega)$  and  $E(j\omega) = Z(j\omega) I(j\omega)$ . Here,  $E(j\omega)$  is the total voltage and  $I(j\omega)$  is the current flowing in the series circuit.

#### 6.1.5.3 Behaviour in the complex domain of a series RLC circuit

Figure 6.5 visualises a series RLC circuit fed by a sinusoidal voltage  $U(t) = \sqrt{2} U \cos(\omega t)$  which can be represented by the complex number  $U(j\omega) = jU$ . The total impedance of the series RLC circuit equals

$$Z(j\omega) = R + j\omega L - \frac{j}{\omega C}. \quad (6.39)$$

In general, the current

$$I(j\omega) = \frac{U(j\omega)}{Z(j\omega)} = \frac{jU}{R + j\omega L - \frac{j}{\omega C}} \quad (6.40)$$

allows to calculate the current with the assumptions that the pulsation  $\omega = 2\pi 2000 = 12566 \text{ rad/s}$ ,  $R = 50\Omega$ ,  $L = 1\text{mH}$ ,  $C = 1\mu\text{F}$  and  $U$  equals 10 volts. More precisely,

$$I(j\omega) = \frac{j10}{50 + j12.57 - j79.58} = -0.0959 + j0.0715 \quad (6.41)$$

which corresponds with the steady state current

$$i(t) = -\sqrt{2} 0.0959 \sin(\omega t) + \sqrt{2} 0.0715 \cos(\omega t) \quad (6.42)$$

or  $i(t) = i_1(t) = -0.1356 \sin(\omega t) + 0.101 \cos(\omega t) \quad (6.43)$

which corresponds with the results obtained using a Laplace transform.

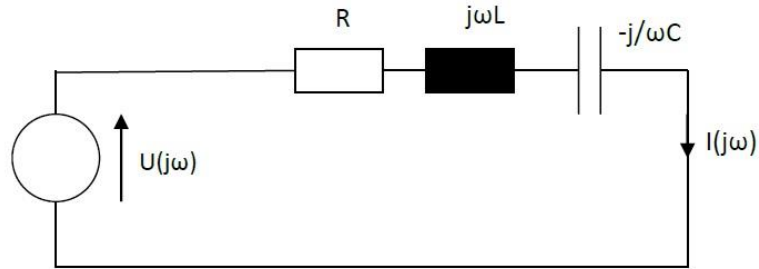


Figure 6.5. Steady state behaviour of series RLC circuit

#### 6.1.6 References

1. R.C. Buck, E.F. Buck, Introduction to Differential Equations, Houghton Mifflin Company, London, 1976.
2. E. Gillon, Wisselstroomtheorie, Standaard Boekhandel, Antwerpen, 1954.
3. H.P. Hsu, Theory and problems of signals and systems, Schaum's Outline Series, Mc Graw-Hill, London, 1995.
4. H.K. Khalil, Nonlinear Systems, Prentice Hall, Upper Saddle River, New Jersey, 2002.

## 6.2 Matlab and Simulink exercises

### 6.2.1 Matlab exercise 1: pendulum

Consider a pendulum where a bob is swinging on a rod having a length  $l = 1 \text{ m}$ . The bob has a mass  $m = 2 \text{ kg}$ . Assume at  $t = t_0$ , the bob has a position  $\theta_0 = \frac{\pi}{4}$ . Using MATLAB, calculate the evolution of the angle  $\theta(t)$  in case:

- 1) there is no friction i.e. the coefficient of friction  $k = 0$ .
- 2) the coefficient of friction  $k = 1$  and the coefficient of friction  $k = 5$ .

What happens with the oscillation if the length of the rod is increasing. Calculate the evolution of the angle  $\theta(t)$  in case the rod has a length  $l = 2 \text{ m}$ .

Using SIMULINK, use the description of the pendulum in the Laplace domain to calculate the evolution of the angle  $\theta(t)$  in case:

- 1) there is no friction i.e. the coefficient of friction  $k = 0$ .
- 2) the coefficient of friction  $k = 1$  and the coefficient of friction  $k = 5$ .

What happens with the oscillation if the length of the rod is increasing. Calculate the evolution of the angle  $\theta(t)$  in case the rod has a length  $l = 2 \text{ m}$ .

### 6.2.2 Simulink exercise: pendulum

Startup simulink and realise a simple model containing a source “pulse generator” (library: sources), a block “transfer function” (library: continuous) and finally a “scope” (library: sinks).

By setting the period of the “pulse generator” larger than the simulation interval and setting the pulse width sufficiently small, a single dirac impulse is approximated. Take care the “transfer function” realizes

$$\frac{lm s + kl}{ml s^2 + kl s + mg} \quad (6.44)$$

Let the scope simulate the output during 10 seconds and by choosing  $l = 1$ ,  $m = 2$ ,  $k = 1$  and  $\theta_0 = \frac{\pi}{4}$  (in the MATLAB environment), obtain the damped oscillation of the bob.

### 6.2.3 Matlab exercise 2 : series RLC circuit

Consider a series RLC circuit with zero initial conditions. The externally applied voltage is a dirac impulse  $E(t) = \delta(t)$  and calculate the electrical charge  $Q(t)$  at the electrodes of the capacitor.

In case  $L = 0.1 \text{ H}$  and  $C = 100 \mu\text{F}$ , calculate the  $R$  value to obtain a critically damped circuit. Use MATLAB to calculate and visualise the evolution of  $Q(t)$ .

Double the  $R$  value in comparison with the critically damped circuit. Calculate and visualise  $Q(t)$  with the same  $L$  and  $C$  giving an overdamped RLC circuit.

Halve the  $R$  value in comparison with the critically damped circuit. Calculate and visualise  $Q(t)$  with the same  $L$  and  $C$  giving an underdamped RLC circuit. mped oscillation of the bob.

#### 6.2.4 Simulink exercises: series RLC circuit

##### Simulation 1

Consider a series RLC circuit. There is a switch  $S$  between the voltage source  $u(t)$  and the resistor  $R$ . At  $t = 0$ , the switch  $S$  closes implying  $i(0^-) = x_1(0^-) = 0$ .

In case  $R = 50 \Omega$ ,  $L = 1 \text{ mH}$ ,  $C = 1 \mu\text{F}$ ,  $f = 2000 \text{ Hz}$  and  $v_C(0^-) = 5 \text{ V}$ , simulate the series RLC circuit using MATLAB/SIMULINK. Here, the input voltage  $u(t) = \sqrt{2} 10 \cos(2\pi f t)$ . Visualise  $y(t) = x_1(t) = i(t)$ . Is there a transient behaviour and a steady state behaviour?

##### Simulation 2:

Consider a series RLC circuit. There is a switch  $S$  between the voltage source  $u(t)$  and the resistor  $R$ . At  $t = 0$ , the switch  $S$  closes implying  $i(0^-) = x_1(0^-) = 0$ .

Using SIMULINK, simulate the behaviour of the same series RLC circuit in case  $f = 100 \text{ kHz}$  with  $u(t) = \sqrt{2} 10 \cos(2\pi f t)$ . Visualise  $y(t) = x_1(t) = i(t)$ . Is there a transient behaviour and a steady state behaviour?

##### Simulation 3:

Consider a series RLC circuit. There is a switch  $S$  between the voltage source  $u(t)$  and the resistor  $R$ . At  $t = 0$ , the switch  $S$  closes implying  $i(0^-) = x_1(0^-) = 0$ .

How is it possible to adapt the state space model in SIMULINK in order to simulate  $v_C(t)$  instead of  $i(t)$ ?

##### Simulation 4:

Consider a series RLC circuit. There is a switch  $S$  between the voltage source  $u(t)$  and the resistor  $R$ . At  $t = 0$ , the switch  $S$  closes implying  $i(0^-) = x_1(0^-) = 0$ .

Simulate the behaviour of the RLC circuit in case  $f = 10 \text{ kHz}$  and  $u(t) = \sqrt{2} 10 \cos(2\pi f t)$ . Visualise  $v_C(t)$  with  $v_C(0^-) = -2 \text{ V}$ . Is there a transient behaviour and a steady state behaviour?

##### Simulation 5

Consider a series RLC circuit. There is a switch  $S$  between the voltage source  $u(t)$  and the resistor  $R$ . At  $t = 0$ , the switch  $S$  closes. The voltage  $u(t) = \mu(t)$  (no voltage when  $t < 0$  and a constant voltage when  $t > 0$ ). Moreover,  $v_C(0^-) = -2 \text{ V}$ .

- 1) Simulate  $v_C(t)$  in case  $R = 5 \Omega$ ,  $L = 1 \text{ mH}$  and  $C = 1 \mu\text{F}$ . Is the RLC circuit overdamped, critically damped or underdamped?
- 2) Simulate  $v_C(t)$  in case  $R = 50 \Omega$ ,  $L = 1 \text{ mH}$  and  $C = 1 \mu\text{F}$ . Is the RLC circuit overdamped, critically damped or underdamped?



- 3) Consider the case where  $L = 1 \text{ mH}$  and  $C = 1 \text{ }\mu\text{F}$ . Calculate  $R$  such that the RLC circuit is critically damped. Simulate  $v_C(t)$ .
- 4) Simulate  $v_C(t)$  in case  $R = 630 \text{ }\Omega$ ,  $L = 1 \text{ mH}$  and  $C = 1 \text{ }\mu\text{F}$ . Is the RLC circuit overdamped, critically damped or underdamped?

## 6.3 State space analysis

### 6.3.1 Example 1: The behaviour of a pendulum

#### 6.3.1.1 A second order differential equation

A spherical bob having a mass  $m$  is attached at the end of a rod with a length  $l$ . Assume the rod is rigid and the rod has zero mass. The pendulum is free to swing in the vertical plane and the pivot is fixed.

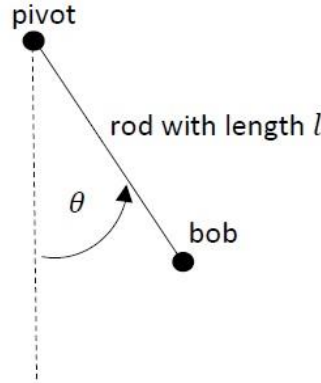


Figure 6.6. Pendulum

The gravitation exerts a vertical force  $mg$  on the bob ( $g$  is the acceleration due to gravity). There is a frictional force which is assumed to be proportional with the speed of the bob. The coefficient of friction equals  $k$ . Based on Newton's second law of motion in the tangential direction (perpendicular with the rod), the behaviour of the pendulum can be described by a nonlinear second order differential equation

$$ml \ddot{\theta}(t) = -mg \sin\theta(t) - kl \dot{\theta}(t). \quad (6.45)$$

Here,  $\ddot{\theta}(t)$  denotes the second derivative with respect to time of  $\theta(t)$  and  $\dot{\theta}(t)$  denotes the first derivative with respect to time of  $\theta(t)$ . By linearizing the differential equation, a linear differential equation

$$ml \ddot{\theta}(t) = -mg \theta(t) - kl \dot{\theta}(t) \quad (6.46)$$

is obtained.

#### 6.3.1.2 The state variable model

By defining two states, a second order differential equation can be reformulated using two first order differential equations. In case of the pendulum, a first state is defined as  $x_1 = \theta$  and a second state is defined as  $x_2 = \dot{\theta}$ . This implies the behaviour of the pendulum is described by the state equations

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{g}{l} \sin x_1 - \frac{k}{m} x_2 \end{aligned} \quad (6.47)$$

When applying an external torque  $T$  to the pendulum, the differential equation becomes

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{g}{l} \sin x_1 - \frac{k}{m} x_2 + \frac{1}{m l^2} T\end{aligned}\quad (6.48)$$

where  $T$  is an input to the system. When considering the linearized second order differential equation, the state variable model equals (without an external torque  $T$ )

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\frac{g}{l} x_1 - \frac{k}{m} x_2\end{aligned}\quad (6.49)$$

### 6.3.2 Example 2: The behaviour of a satellite

#### 6.3.2.1 The time domain and the Laplace domain

Consider the satellite visualised in Figure 6.7. The satellite is spherical and has four thrusters. The angle  $\theta(t)$  is the yaw angle of the satellite. Using the thrusters, the yaw angle can be controlled since active thrusters apply a torque  $\tau(t)$ . The torque of the two active thrusters shown in Figure 6.7 tends to reduce  $\theta(t)$ . The two other thrusters tend to increase  $\theta(t)$ .

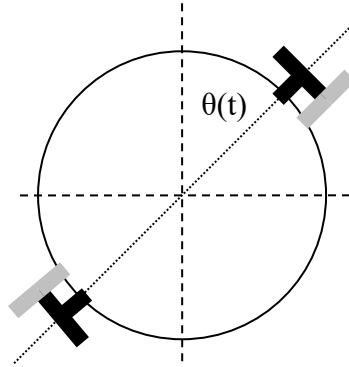


Figure 6.7. Satellite with thrusters

Since there is essentially no friction in the environment of the satellite, and assuming the satellite is rigid, the behaviour of the satellite is given by the differential equation

$$J \frac{d^2 \theta(t)}{dt^2} = \tau(t). \quad (6.50)$$

Here,  $J$  is the satellite's moment of inertia in the yaw axis. Taking the Laplace transform yields

$$J s^2 \theta(s) = \tau(s). \quad (6.51)$$

When considering the satellite, the driving torque  $\tau(t)$  is the input of the system and the yaw angle  $\theta(s)$  is the output of the system. The transfer function of a system is the ratio between the Laplace

transform of the output and the Laplace transform of the input. In the present example, the behaviour of the satellite is described by the transfer function

$$\frac{\theta(s)}{\tau(s)} = \frac{1}{Js^2}. \quad (6.52)$$

### 6.3.2.2 The state variable model

The behaviour of the satellite of Figure 6.7 is described by the second order differential equation (6.50) by defining the states  $x_1(t) = \theta(t)$

and

$$x_2(t) = \frac{dx_1(t)}{dt} = \dot{x}_1(t) = \dot{\theta}(t) = \frac{d\theta(t)}{dt}, \quad (6.53)$$

one obtains that

$$\dot{x}_2(t) = \frac{d^2\theta(t)}{dt^2} = \frac{1}{J}\tau(t). \quad (6.54)$$

Therefore, the behaviour of the satellite can be described by two first order differential equations instead of using one single second order differential equation. This state space description (with state variables  $x_1$  and  $x_2$ ) in vector-matrix form equals

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1/J \end{bmatrix} \tau. \quad (6.55)$$

## 6.3.3 Example 3: servo motor

### 6.3.3.1 The time domain and the Laplace domain

A DC servo motor can be used as a positioning system for an antenna. Figure 6.8 shows a separately excited DC motor. We assume that the excitation current  $i_F$  is constant implying a constant magnetic field. The behaviour of the motor is armature controlled by a voltage  $e(t)$ . The armature has a resistance  $R_A$  and an inductance  $L_A$ . We assume that the inductance  $L_A$  can be ignored which is the case for many servomotors.

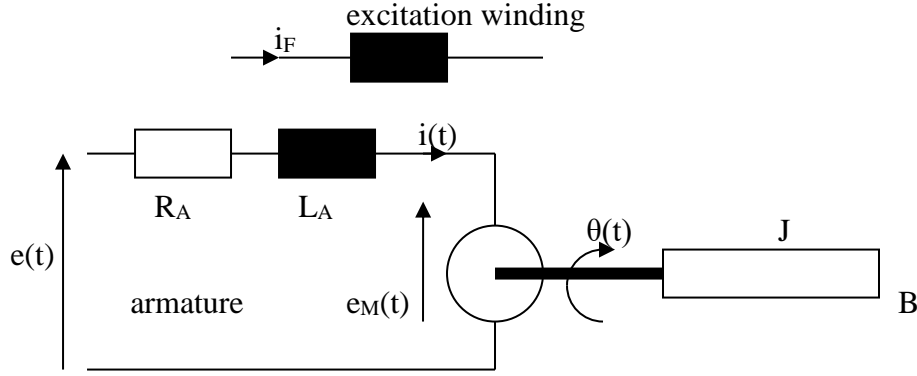


Figure 6.8. DC servo motor

The motor back emf  $e_M(t)$  is given by

$$e_M(t) = K_B \omega(t) = K_B \frac{d\theta(t)}{dt}. \quad (6.56)$$

Here,  $\theta(t)$  is the motor shaft position,  $\omega(t)$  is the shaft angular velocity and  $K_B$  is a motor dependent constant. The total moment of inertia connected to the motor shaft equals  $J$  and  $B$  is the total viscous friction.

The mechanical behaviour of the DC servo motor and the load is described by the differential equation

$$J \frac{d^2 \theta(t)}{dt^2} + B \frac{d\theta(t)}{dt} = \tau(t). \quad (6.57)$$

The developed torque for this motor is given by  $\tau(t) = K_T i(t)$

where  $i(t)$  is the armature current and  $K_T$  is a motor dependent constant. The electrical behaviour of the armature is given by the equation ( $L_A$  is ignored)  $e(t) = e_M(t) + R_A i(t)$ .

By combining these equations, one obtains that

$$i(t) = \frac{e(t) - e_M(t)}{R_A} = \frac{e(t)}{R_A} - \frac{K_B}{R_A} \frac{d\theta(t)}{dt} \quad (6.58)$$

and that

$$\tau(t) = K_T i(t) = \frac{K_T}{R_A} e(t) - \frac{K_T K_B}{R_A} \frac{d\theta(t)}{dt}. \quad (6.59)$$

The desired model of the DC servo motor is given by

$$J \frac{d^2 \theta(t)}{dt^2} + \frac{BR_A + K_T K_B}{R_A} \frac{d\theta(t)}{dt} = \frac{K_T}{R_A} e(t). \quad (6.60)$$

Here, we have a second order differential equation. If the armature inductance  $L_A$  were not negligible, this model would be a third order differential equation. If we take the Laplace transform and solve for the transfer function  $H(s)$ , the result equals

$$H(s) = \frac{\theta(s)}{E(s)} = \frac{K_T/R_A}{Js^2 + \frac{BR_A + K_T K_B}{R_A}s} = \frac{K_T/JR_A}{s \left( s + \frac{BR_A + K_T K_B}{JR_A} \right)}. \quad (6.61)$$

Since the servo motor is armature controlled, the applied voltage  $e(t)$  ( $E(s)$  in the Laplace domain) is the input and the rotor position  $\theta(t)$  ( $\theta(s)$  in the Laplace domain) is the output.

#### 6.3.3.2 The state variable model

By defining the states  $x_1(t) = \theta(t)$

and  $x_2(t) = \dot{\theta}(t) = \frac{d\theta(t)}{dt}$ ,

one obtains the relationship

$$\dot{x}_2(t) = \frac{d^2 \theta(t)}{dt^2} = - \frac{BR_A + K_T K_B}{JR_A} x_2(t) + \frac{K_T}{JR_A} e(t). \quad (6.62)$$

The state equation can be written as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{BR_A + K_T K_B}{JR_A} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ K_T/JR_A \end{bmatrix} e. \quad (6.63)$$

#### 6.3.4 State space representation

Having considered three examples, the state space representation will be generalised. Consider a linear time invariant system with an input  $u(t)$  and an output  $y(t)$ . The behaviour of the system can be described by a differential equation of the form

$$\frac{d^N y(t)}{dt^N} + a_1 \frac{d^{N-1} y(t)}{dt^{N-1}} + \dots + a_N y(t) = b u(t). \quad (6.64)$$

One possible set of initial conditions is  $y(0)$ ,  $y^{(1)}(0)$ ,  $y^{(2)}(0)$ ,  $\dots$ ,  $y^{(N-1)}(0)$ , where

$$y^{(k)}(t) = \frac{d^k y(t)}{dt^k}. \quad (6.65)$$

By defining N state variables  $x_1(t), x_2(t), \dots, x_N(t)$  as  $x_1(t) = y(t), x_2(t) = y^{(1)}(t), \dots, x_N(t) = y^{(N-1)}(t)$ , one obtains that

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= x_3(t) \\ &\vdots \\ \dot{x}_{N-1}(t) &= x_N(t) \\ \dot{x}_N(t) &= -a_N x_1(t) - a_{N-1} x_2(t) - \dots - a_1 x_N(t) + b u(t). \end{aligned} \quad (6.66)$$

The output can be written as  $y(t) = x_1(t)$

In matrix form, the behaviour of the system can be described as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_{N-1}(t) \\ \dot{x}_N(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \bullet & \bullet & \bullet & 0 \\ 0 & 0 & 1 & \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet & \bullet & \bullet & 1 \\ -a_N & -a_{N-1} & -a_{N-2} & \bullet & \bullet & \bullet & -a_1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{N-1}(t) \\ x_N(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ b \end{bmatrix} u(t) \quad (6.67)$$

and

$$y(t) = \begin{bmatrix} 1 & 0 & \bullet & \bullet & \bullet & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_{N-1}(t) \\ x_N(t) \end{bmatrix}. \quad (6.68)$$

Notice that, using the state space representation, a N-th order differential equation is replaced by N first order differential equations. By defining

$$X(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \bullet \\ \bullet \\ \bullet \\ x_{N-1}(t) \\ x_N(t) \end{bmatrix} \quad (6.69)$$

and

$$\dot{X}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \bullet \\ \bullet \\ \bullet \\ \dot{x}_{N-1}(t) \\ \dot{x}_N(t) \end{bmatrix}, \quad (6.70)$$

the behaviour of the system is described by the equations

$$\dot{X}(t) = A X(t) + B u(t) \quad (6.71)$$

and  $y(t) = C X(t)$ .

Here,

$$A = \begin{bmatrix} 0 & 1 & 0 & \bullet & \bullet & \bullet & 0 \\ 0 & 0 & 1 & \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet & \bullet & \bullet & 1 \\ -a_N & -a_{N-1} & -a_{N-2} & \bullet & \bullet & \bullet & -a_1 \end{bmatrix} \quad (6.72)$$



$$B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (6.73)$$

$$C = [1 \quad 0 \quad \bullet \quad \bullet \quad \bullet \quad 0 \quad 0]. \quad (6.74)$$

#### 6.3.4.1 Satellite and DC servo motor

When considering the satellite of Paragraph 2, the order  $N = 2$ , the input  $u(t) = \tau(t)$  and the output  $y(t) = \theta(t)$ . Verify that  $a_1 = 0$ ,  $a_2 = 0$  and that  $b = 1/J$ .

When considering the DC servo motor system of Paragraph 3, the order  $N = 2$ , the input  $u(t) = e(t)$  and the output  $y(t) = \theta(t)$ . Verify that  $a_1 = 0$ ,  $a_2 = (BR_A + K_T K_B)/JR_A$  and that  $b = K_T/JR_A$ .

### 6.3.5 Solutions of state space equations

#### 6.3.5.1 Calculating the transfer function

In general, a  $N$  dimensional state space representation with a single input  $u(t)$  and a single output  $y(t)$  can be written as  $\dot{X}(t) = A X(t) + B u(t)$  and  $y(t) = C X(t) + D u(t)$ .

In the general case,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \bullet & \bullet & \bullet & a_{1N} \\ a_{21} & a_{22} & a_{23} & \bullet & \bullet & \bullet & a_{2N} \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ a_{(N-1)1} & a_{(N-1)2} & a_{(N-1)3} & \bullet & \bullet & \bullet & a_{(N-1)N} \\ a_{N1} & a_{N2} & a_{N3} & \bullet & \bullet & \bullet & a_{NN} \end{bmatrix} \quad (6.75)$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \bullet \\ \bullet \\ \bullet \\ b_{N-1} \\ b_N \end{bmatrix} \quad (6.76)$$

$$C = [c_1 \quad c_2 \quad \bullet \quad \bullet \quad \bullet \quad c_{N-1} \quad c_N] \quad (6.77)$$

$$D = [d]. \quad (6.78)$$

By taking the Laplace transform of the differential equation, one obtains the relationship

$$s X(s) - X(0) = A X(s) + B u(s) \quad (6.79)$$

$$y(s) = C X(s) + D u(s). \quad (6.80)$$

These equations imply that

$$(sI - A)X(s) = X(0) + B u(s), \quad X(s) = (sI - A)^{-1} X(0) + (sI - A)^{-1} B u(s) \quad (6.81)$$

and that

$$y(s) = C (sI - A)^{-1} X(0) + [C (sI - A)^{-1} B + D] u(s). \quad (6.82)$$

By taking the inverse Laplace transform of  $y(s)$ , one obtains  $y(t)$ . Notice that the first term corresponds to the zero-input response due to  $X(0)$  and that the second term is determined by the input  $u(t)$ .

The transfer function equals

$$H(s) = \frac{y(s)}{u(s)} = C (sI - A)^{-1} B + D. \quad (6.83)$$

Notice the transfer function  $H(s)$  is the ratio between the Laplace transforms of the output and the input in case the initial condition  $X(0) = 0$ .

### 6.3.5.2 Solution in the time domain

Define the state transition matrix

$$\Phi(t) = e^{At} = I + At + \frac{1}{2!} A^2 t^2 + \frac{1}{3!} A^3 t^3 + \dots \quad (6.84)$$

which is the inverse Laplace transform of  $(sI - A)^{-1}$ . Verify that

$$\frac{d\Phi(t)}{dt} = A\Phi(t) = \Phi(t)A. \quad (6.85)$$

Consider

$$X(t) = \Phi(t)X(0) + \int_0^t \Phi(t-\tau)Bu(\tau)d\tau \quad (6.86)$$

and verify, by taking the derivative, that this  $X(t)$ -expression satisfies the differential equation  $\dot{X}(t) = AX(t) + Bu(t)$ .

Moreover, this  $X(t)$ -expression can also be obtained by taking the inverse Laplace transform of  $X(s) = (sI - A)^{-1}X(0) + (sI - A)^{-1}Bu(s)$ ,

These expressions imply that

$$y(t) = C\Phi(t)X(0) + \int_0^t C\Phi(t-\tau)Bu(\tau)d\tau + Du(t). \quad (6.87)$$

$$y(t) = Ce^{At}X(0) + \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Du(t). \quad (6.88)$$

Moreover, this  $y(t)$ -expression can also be obtained by taking the inverse Laplace transform of  $y(s) = C(sI - A)^{-1}X(0) + [C(sI - A)^{-1}B + D]u(s)$ .

### 6.3.6 Exercise

#### 6.3.6.1 Theoretical calculation

Consider the series RLC circuit of Figure 6.9. Write down the state space equations in case the voltage  $u(t)$  is the input signal and the current  $i(t)$  is the output  $y(t)$ .

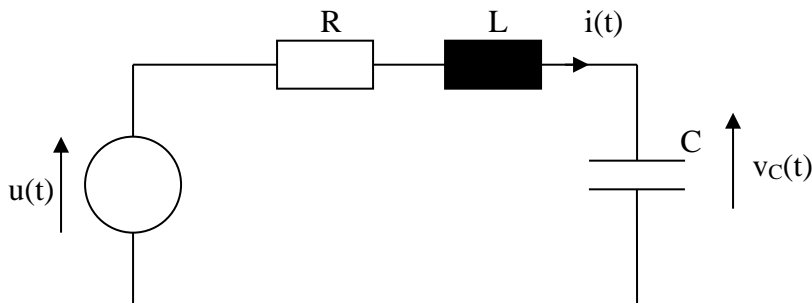


Figure 6.9. Series RLC circuit

Using Kirchoff's law, the behaviour of the series RLC circuit is described by the differential equation

$$Ri(t) + L \frac{di(t)}{dt} + \frac{1}{C} \int_0^t i(\tau) d\tau + v_C(0^-) = u(t). \quad (6.89)$$

By defining the first state  $x_1(t) = i(t)$  and  $x_2(t) = v_C(t)$ , one obtains that

$$\begin{aligned} \dot{x}_1(t) &= \frac{di(t)}{dt} = -\frac{R}{L}x_1(t) - \frac{1}{L}x_2(t) + \frac{1}{L}u(t) \\ \dot{x}_2(t) &= \frac{1}{C}x_1(t) \\ y(t) &= x_1(t). \end{aligned} \quad (6.90)$$

This state space description in vector-matrix form equals

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} &= \begin{bmatrix} -\frac{R}{L} & -\frac{1}{L} \\ \frac{1}{C} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} u. \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{aligned} \quad (6.91)$$

### 6.3.6.2 Simulation with simulink 1

Consider the series RLC circuit of Figure 6.9. There is a switch S between the voltage source  $u(t)$  and the resistor R. At  $t = 0$ , the switch S closes implying  $i(0^-) = x_1(0^-) = 0$ .

Simulation 1:

In case  $R = 50 \Omega$ ,  $L = 1 \text{ mH}$ ,  $C = 1 \mu\text{F}$ ,  $f = 2000 \text{ Hz}$  and  $v_C(0^-) = 5 \text{ V}$ , simulate the series RLC circuit using MATLAB/SIMULINK. Here, the input voltage  $u(t) = \sqrt{2} 10 \cos(2\pi f t)$ . Visualise  $y(t) = x_1(t) = i(t)$ . Is there a transient behaviour and a steady state behaviour?

Simulation 2:

Using SIMULINK, simulate the behaviour of the same series RLC circuit in case  $f = 100 \text{ kHz}$  with  $u(t) = \sqrt{2} 10 \cos(2\pi f t)$ . Visualise  $y(t) = x_1(t) = i(t)$ . Is there a transient behaviour and a steady state behaviour?

Simulation 3:

How is it possible to adapt the state space model in SIMULINK in order to simulate  $v_C(t)$  instead of  $i(t)$ ?

Simulation 4:

Simulate the behaviour of the RLC circuit in case  $f = 10 \text{ kHz}$  and  $u(t) = \sqrt{2} 10 \cos(2\pi f t)$ . Visualise  $v_C(t)$  with  $v_C(0^-) = -2 \text{ V}$ . Is there a transient behaviour and a steady state behaviour?

#### 6.3.6.3 Simulation with simulink 2

Consider the series RLC circuit of Figure 6.9. The voltage  $u(t) = \mu(t)$  and  $v_C(0^-) = -2 \text{ V}$ .

- 1) Simulate  $v_C(t)$  in case  $R = 5 \text{ } \Omega$ ,  $L = 1 \text{ mH}$  and  $C = 1 \text{ } \mu\text{F}$ . Is the RLC circuit overdamped, critically damped or underdamped?
- 2) Simulate  $v_C(t)$  in case  $R = 50 \text{ } \Omega$ ,  $L = 1 \text{ mH}$  and  $C = 1 \text{ } \mu\text{F}$ . Is the RLC circuit overdamped, critically damped or underdamped?
- 3) Consider the case where  $L = 1 \text{ mH}$  and  $C = 1 \text{ } \mu\text{F}$ . Calculate  $R$  such that the RLC circuit is critically damped. Simulate  $v_C(t)$ .
- 4) Simulate  $v_C(t)$  in case  $R = 630 \text{ } \Omega$ ,  $L = 1 \text{ mH}$  and  $C = 1 \text{ } \mu\text{F}$ . Is the RLC circuit overdamped, critically damped or underdamped?

#### 6.3.7 References

1. H.P. Hsu, Theory and problems of signals and systems, Schaum's Outline Series, Mc Graw-Hill, London, 1995.
2. H.K. Khalil, Nonlinear Systems, Prentice Hall, Upper Saddle River, New Jersey, 2002.
3. C.L. Phillips and H.T. Nagle, Digital Control System Analysis and Design, Prentice Hall, Englewood Cliffs, 1984.
4. D.M. Wiberg, Theory and Problems of State Space and Linear Systems, Schaum's Outline Series, Mc Graw-Hill Book Company, New York, 1971.

## 6.4 System theory: DSP: Analog and digital filters

### 6.4.1 Introduction

Filters are used to remove unwanted frequency components of a signal or to enhance wanted frequency components. A distinction can be made between analog filters and digital filters. In both cases, a distinction can be made between low-pass filters, band-pass filters, high-pass filters and band-reject filters.

When considering analog filters, a distinction can be made between passive filters and active filters. Passive filter implementations are based on combinations of resistors, capacitors and inductors. Active filters not only use resistors, capacitors and inductors but also active (amplifying) components are used. For instance bipolar transistors or opamps are commonly used as active components and they need an additional DC power source.

When considering digital filters, a distinction can be made between FIR filters and IIR filters. In both cases, the input signal is sampled and an analog-to-digital converter converts the signal to a combination of 0 and 1 signals. Using a processor or an FPGA, the actual digital filter is realised. The output of this actual digital filter is converted into an analog signal using a digital-to-analog converter.

When designing an analog filter, a lot of design tools are available since analog filters are already used and designed during decades. When designing a digital filter, a digital filter design methodology can be used. However, it is also possible to reformulate the digital filter specifications to a related analog filter and designing this analog filter using one of the numerous existing design methodologies. This analog filter can be transformed to the desired digital filter.

### 6.4.2 Analog low-pass filters

#### 6.4.2.1 Filter specifications

Figure 6.10 shows the typical magnitude specifications for an analog low-pass filter and notice the pulsation  $\Omega = 2\pi f$  (expressed in radians per second). The filter has a passband ranging from 0 to  $\Omega_p$  and a stopband for pulsations larger than  $\Omega_s$ . The transition band ranges from  $\Omega_p$  to  $\Omega_s$ . The analog filter in Fig. 6.10 has a transfer function  $H_a(j\Omega)$  and in the passband

$$1 - \delta_p \leq |H_a(j\Omega)| \leq 1 + \delta_p. \quad (6.92)$$

The smaller  $\delta_p$ , the smaller the ripple in the filter amplification in the passband which is in general a good property. In case of a larger  $\delta_p$ , different frequency components face different amplifications. In case the input signal is an audio signal, one can hear this commonly undesired phenomenon. Also in case the input signal is a measurement signal, different components of the original signal commonly face undesired differences in the amplifications. The smaller  $\delta_p$ , the larger the required order of the filter i.e. the more complex and the more performing the filter.

In the stopband, where  $\Omega_S \leq \Omega \leq +\infty$ ,  $|H_a(j\Omega)| \leq \delta_S$ . The smaller  $\delta_S$ , the more performing the filter implying a larger order is needed. In Figure 6.10, the magnitude  $|H_a(j\Omega)|$  of the transfer function contains a ripple in the passband and in the stopband. The ripple in the stopband is less important since for small  $\delta_S$  values, the stopband frequencies are filtered out.

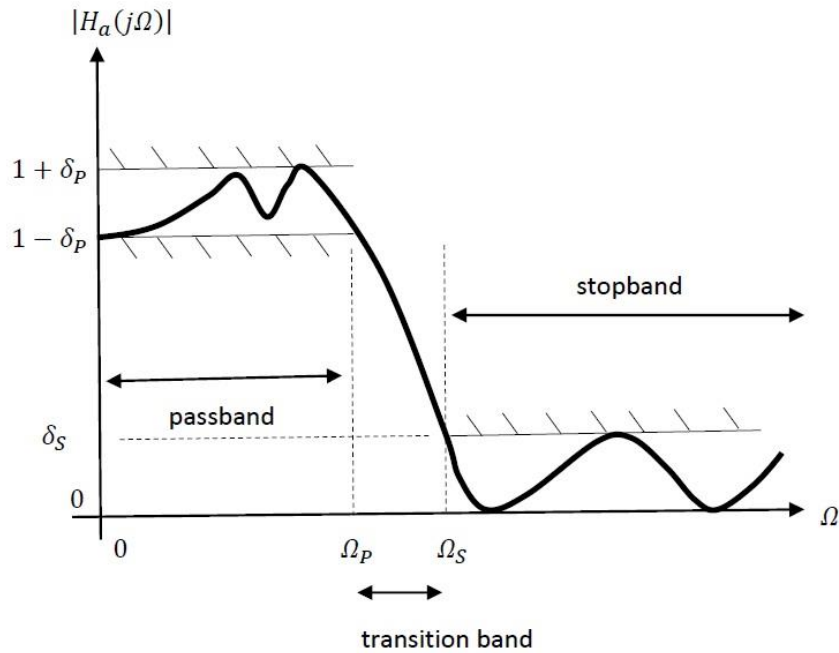


Figure 6.10. Magnitude specifications for an analog low-pass filter (based on: Mitra)

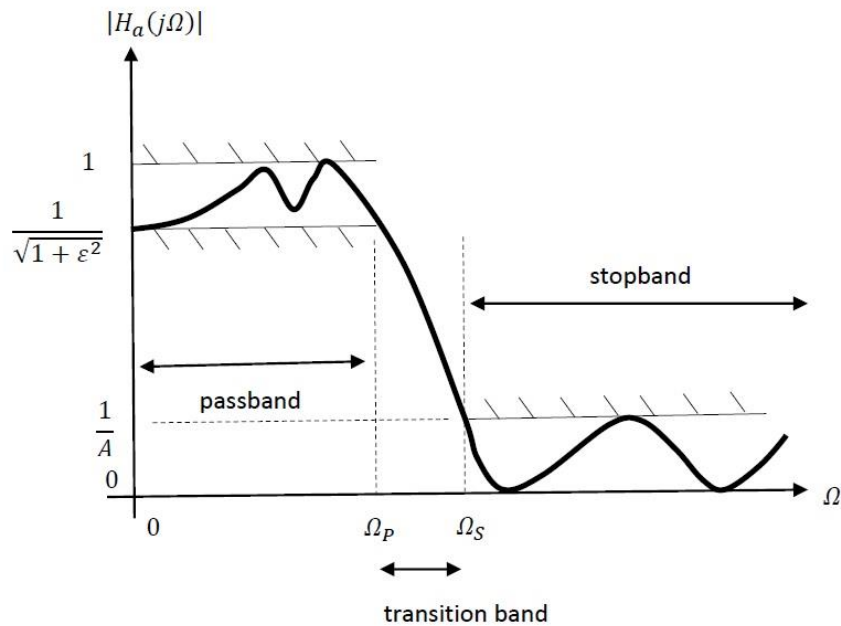


Figure 6.11. Magnitude specifications for an analog low-pass filter (based on: Mitra)

Instead of the magnitude specifications of Fig. 6.10, the specifications of Figure 6.11 can be used. The definitions of the passband, transition band and stopband have not changed. The specifications of the magnitudes have changed. In the passband,

$$\frac{1}{\sqrt{1+\varepsilon^2}} \leq |H_a(j\Omega)| \leq 1. \quad (6.92)$$

In the stopband, where  $\Omega_S \leq \Omega \leq +\infty$ ,

$$|H_a(j\Omega)| \leq \frac{1}{A} \quad (6.93)$$

As the order of the filter increases, it is possible to realise a filter

- having a smaller ripple in the passband (a smaller  $\delta_p$  in Figure 6.10, a smaller  $\varepsilon$  in Figure 6.11),
- having a smaller transition band i.e.  $\Omega_p$  and  $\Omega_S$  are closer to each other,
- with more attenuation in the stopband (a smaller  $\delta_S$  in Figure 6.10, a larger  $A$  in Figure 6.11).

When considering filter specifications as visualised in Figure 6.10 and Figure 6.11, the ranges of the allowed amplifications  $|H_a(j\Omega)|$  are considered. Alternatively, the amplifications can be expressed in decibels (dB) by using

$$20 \log(|H_a(j\Omega)|). \quad (6.94)$$

Consider the example where  $\varepsilon = 1$  which implies the amplification ranges between 0.707 and 1 in the passband. This means the amplification ranges between 0 dB and  $-3$  dB. In case  $A = 100$  in Figure 6.11, an attenuation of at least 40 dB is obtained in the stopband.

#### 6.4.2.2 Filter realisations

There are numerous ways to realise an analog filter satisfying specifications as visualised in Fig. 6.10 and Fig. 6.11. Fig. 6.12 shows an active filter structure appropriate to realise a second order behaviour. By connecting several second order filter structures in cascade, a higher order filter can be obtained. For instance a tenth order filter can be obtained by connecting five such filter structures in cascade. In case for instance an eleventh order filter is needed, an additional first order RC network can be added.

Actually, the transfer function is split up in a cascade connection of several biquads (second order transfer functions). Each biquad is individually realised as an analog electronic filter circuit. In case of an odd order, an additional first order filter circuit is needed.



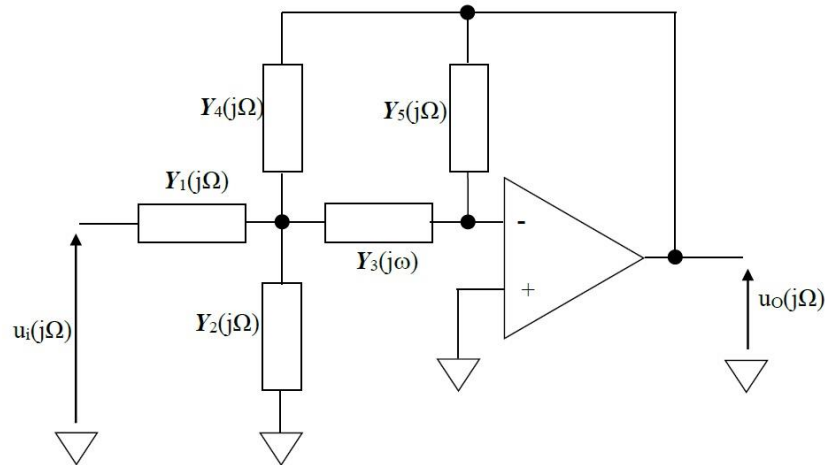


Figure 6.12. Practical realisation of a second order analog active filter

When designing an analog filter, the values for the resistors, inductors and capacitors are calculated. In reality, the values must be chosen from the E12 series (or another E series). This implies deviations between the desired values and the real values. Moreover, these resistor, inductor and capacitor values have tolerances (e.g. 10% tolerance components are used).

The real analog filter behaviour will depend on the deviations on the values of the resistors, inductors and capacitors which are used to build the filter. Depending on the filter structure, the impact of these deviations on the transfer function  $H_a(j\Omega)$  is different.

#### 6.4.2.3 Butterworth filters

Fig. 6.10 and Fig. 6.11 show ripples in the passband. Not all filter realizations have such a ripple behaviour. For instance, a Butterworth filter has a maximally flat behaviour as shown in Figure 6.13. The maximally flat behaviour is obtained in the passband and in the stopband. Especially the maximally flat behaviour in the passband is important since it implies different frequency components, close to each other, have approximately the same amplification.

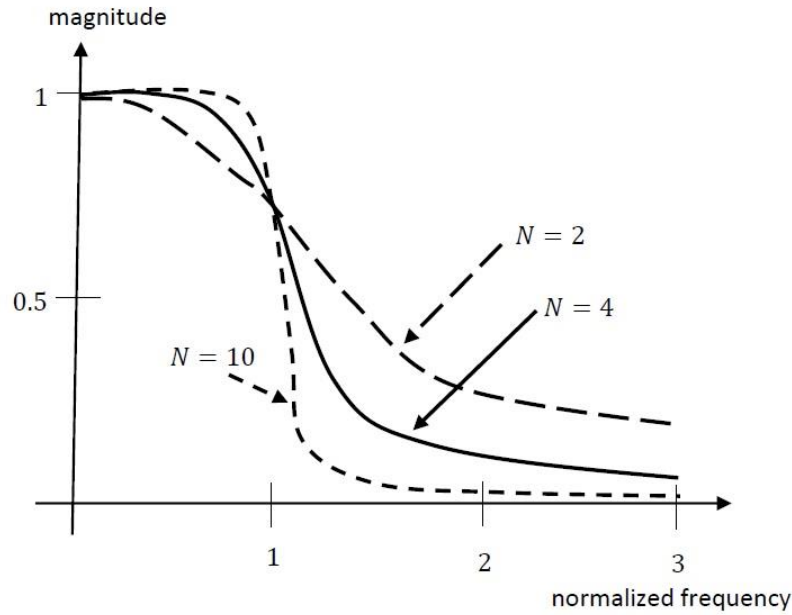


Figure 6.13. Typical frequency behaviour of a lowpass Butterworth filter (based on: Mitra)

In general, the transfer function of an analog filter equals

$$H_a(s) = \frac{B(s)}{A(s)}. \quad (6.95)$$

In case the numerator  $B(s) = 1$ , the denominator  $A(s)$  determines the behaviour of the Butterworth filter.

### Numerical examples 1

Consider a number of the simple Butterworth filters

$$H_a(s) = \frac{1}{s^2 + \sqrt{2}s + 1}.$$

$$H_a(s) = \frac{1}{(s^2 + s + 1)(s + 1)} = \frac{1}{s^3 + 2s^2 + 2s + 1}. \quad (6.96)$$

Using the MATLAB instruction *roots* (e.g. *roots([1 sqrt(2) 1])*), the locations of the poles in the s-plane are calculated. Visualise the locations of the poles of the given transfer functions in the s-plane. Notice these poles are located on a half circle in the open left half plane. Notice these poles are equally spread over this half circle.

Since the poles are located in the open left half plane, no stability problems arise. Since the locations of the poles are equally spread over a half circle, the desired maximally flat behaviour in the passband is obtained.

## Numerical examples 2

Consider a filter having a transfer function  $H_a(s) = \frac{1}{s^3 + 2s^2 + 2s + 1}$

A three dimensional view of this transfer function can be obtained using the instructions

```
x = -2:0.01:2;
y=x;
for n1 = 1:length(x)
    for n2= 1: length(y)
        s(n1,n2)=x(n1) + i*y(n2);
    end
end
H=log(abs(1./(s.^3 + 2*s.^2 + 2*s + 1)));
meshc(x,y,H);
```

## The transfer function

The magnitude-squared response of an analog Butterworth lowpass filter (having an order  $N$ ) is given by

$$|H_a(j\Omega)|^2 = \frac{1}{1 + (\Omega/\Omega_C)^{2N}}. \quad (6.97)$$

The maximally flat behaviour is obtained since the  $2N - 1$  derivatives of  $|H_a(j\Omega)|^2$  equal zero at  $\Omega = 0$ . There are actually two parameters which characterize the behaviour of the Butterworth filter i.e.  $\Omega_C$  and  $N$ . When using the specifications of Figure 6.11, based on the choice of  $\Omega_P$ ,  $\Omega_S$ ,  $\varepsilon$  and  $A$  it is possible to determine  $\Omega_C$  and  $N$ . More precisely

$$\begin{aligned} |H_a(j\Omega_P)|^2 &= \frac{1}{1 + (\Omega_P/\Omega_C)^{2N}} = \frac{1}{1 + \varepsilon^2} \\ |H_a(j\Omega_S)|^2 &= \frac{1}{1 + (\Omega_S/\Omega_C)^{2N}} = \frac{1}{A^2} \end{aligned} \quad (6.98)$$

which implies the required order equals

$$N = \frac{1}{2} \frac{\log_{10}[(A^2 - 1)/\varepsilon^2]}{\log_{10}(\Omega_S/\Omega_P)}. \quad (6.99)$$

MATLAB can be used to determine the order of a Butterworth filter using the instruction *buttord*. More precisely, try the instructions:

```
help buttord
[N,Wn] = buttord(Wp,Ws,Rp,Rs,'s')
```

The ripple  $R_P$  in the passband is expressed in dB and the stopband attenuation  $R_S$  is also expressed in dB. The pulsations  $W_P$  and  $W_S$  are expressed in radians per second (they are equal to  $\Omega_P$  and  $\Omega_S$ ). Once  $N$  and  $W_n$  are known, the transfer function of the Butterworth filter can be calculated by the instruction:

Help butter

`[B,A] = butter(N, Wn, 's')`

To get some feeling, try the instructions

`[B,A] = butter(2,1,'s')`

`[B,A] = butter(3,1,'s')`

`[B,A] = butter(2,2,'s')`

`[B,A] = butter(2,100,'s')`

In case of `[B,A] = butter(2,100,'s')`, one obtains  $B = [0 \ 0 \ 100]$  and  $A = [1 \ 14.14 \ 100]$  giving a transfer function

$$H_a(s) = \frac{100}{s^2 + 14.14 s + 100}. \quad (6.100)$$

### Example

Consider a Butterworth filter having a passband ripple  $R_P = 1 \text{ dB}$  and a stopband attenuation  $R_S = 40 \text{ dB}$ . Suppose  $W_P = \Omega_P = 2 * \pi * 1000 \text{ rad/s}$  and  $W_S = \Omega_S = 2 * \pi * 2000 \text{ rad/s}$ . The filter can be obtained using the instructions:

`[N,Wn] = buttord(2*pi*1000, 2*pi*2000, 1, 40, 's')`

This gives an  $N = 8$  and  $W_n = 7.066 e + 003 = 7,066$  implying  $W_n$  is located between  $2 * \pi * 1000 \text{ rad/s}$  and  $2 * \pi * 2000 \text{ rad/s}$ . The transfer function is obtained by

`[B,A] = butter(N, Wn, 's')`

The locations of the poles (there are no zeros) can be visualised using the instruction

`pzmap(B,A)`

Notice the poles are located on a circle having radius  $W_n$  and they are equally spread over a half circle in the open left half plane. The frequency behaviour can be visualized on the imaginary axis of the s-plane. The lower frequencies are closer to the poles which implies a high amplification. The higher frequencies are removed from the poles which implies an attenuation. This amplification as a function of the pulsation is obtained by the instruction

`freqs(B,A)`

Using the instruction `[N,Wn] = buttord(Wp,Ws,Rp,Rs,'s')`, it is possible to show the order  $N$  increases when

- $W_p = \Omega_p$  and  $W_s = \Omega_s$  are closer to each other,
- the ripple  $R_p$  decreases,
- the attenuation  $R_s$  increases.
- 

#### 6.4.2.4 Exercise

Consider a first order analog low-pass filter containing a resistor  $R$  and a capacitor  $C$  as shown in Fig. 6.14. Determine the relationships between the input voltage  $u_i(t)$ , the output voltage  $u_o(t)$  and the current  $i(t)$ . Determine these relationships in the time-domain and in the Laplace domain giving the transfer function

$$H(s) = \frac{U_o(s)}{U_i(s)} = \frac{1}{1+sRC}. \quad (6.101)$$

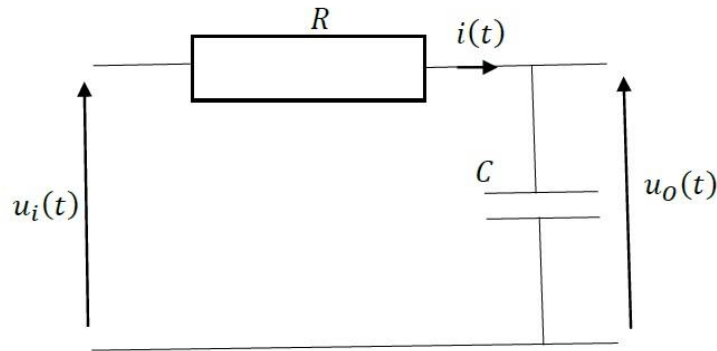


Figure 6.14. First order low-pass filter

This transfer function has no zeros but there is one single pole at  $s = -1/RC$ . Visualize the  $s$ -plane and verify

- the stability of the filter (BIBO stability: a bounded input implies a bounded output),
- the filter has a low-pass behaviour.

Using the inverse Laplace transform, calculate the impulse response of the filter and visualise this impulse response.

Consider a transfer function

$$H(s) = \frac{1}{s+a} \quad (6.102)$$

and make a difference between an  $a > 0$  and an  $a < 0$ . Discuss the stability properties and discuss the impulse response.

#### 6.4.2.5 Chebyshev filter of the first type

Fig. 6.15 shows a typical Chebyshev filter of the first type (for several orders  $N$ ). Notice the equiripple behaviour in the passband and the maximally flat behaviour in the stopband.

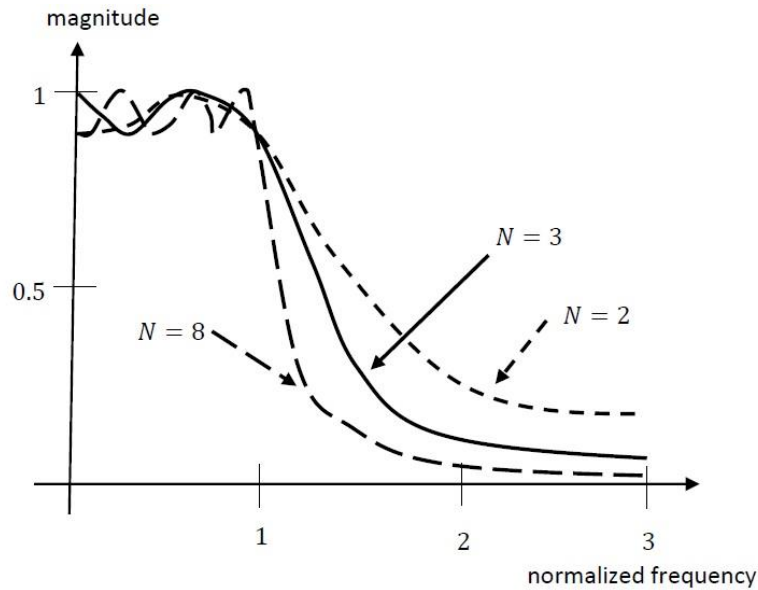


Figure 6.15. Chebyshev low-pass filter of the first type (based on: Mitra)

Consider such a Chebyshev filter satisfying the conditions  $R_p = 1 \text{ dB}$ ,  $R_s = 40 \text{ dB}$ ,  $\Omega_p = 2 * \pi * 1000 \text{ rad/s}$  and  $\Omega_s = 2 * \pi * 2000 \text{ rad/s}$ . Using MATLAB instructions, the filter can be designed giving

help cheblord

```
[N,Wn] = cheblord(2*pi*1000, 2*pi*2000, 1, 40, 's')
```

```
[B,A] = cheby1(N, 1, Wn, 's')
```

```
freqs(B,A)
```

```
pzmap(B,A)
```

The Chebyshev filter has a transfer function

$$H_a(s) = \frac{B(s)}{A(s)} \quad (6.103)$$

and the coefficients in  $B(s)$  are given by  $B$ . The coefficients in  $A(s)$  are given by  $A$ .

Using the same specifications when designing the Chebyshev filter of the first type, an order  $N = 5$  is obtained instead of the order  $N = 8$  in case of the already studied Butterworth filter.

When realizing such a Chebyshev filter of the first type, it is clear

- the transfer function contains no zeros,
- all poles are located in the open left half plane giving appropriate stability properties,
- the poles are located on an ellipse (not on a circle) (Due to this property, the amplification in the passband is not constant. There is a ripple in the passband.),

- the amplification decreases as the pulsation increases (Visualize the s-plane with the locations of the poles. Visualize the amplification  $|H_a(j\Omega)|$  on the imaginary axis.).
- 

#### 6.4.2.6 Chebyshev filter of the second type

A Chebyshev filter of the first type has an equiripple behaviour in the passband and a maximally flat behaviour in the stopband. A Chebyshev filter of the second type (also called an inverse Chebyshev filter) has an equiripple behaviour in the stopband and a maximally flat behaviour in the passband. Fig. 6.16 visualizes the frequency behaviour of such a Chebyshev filter of the second type.

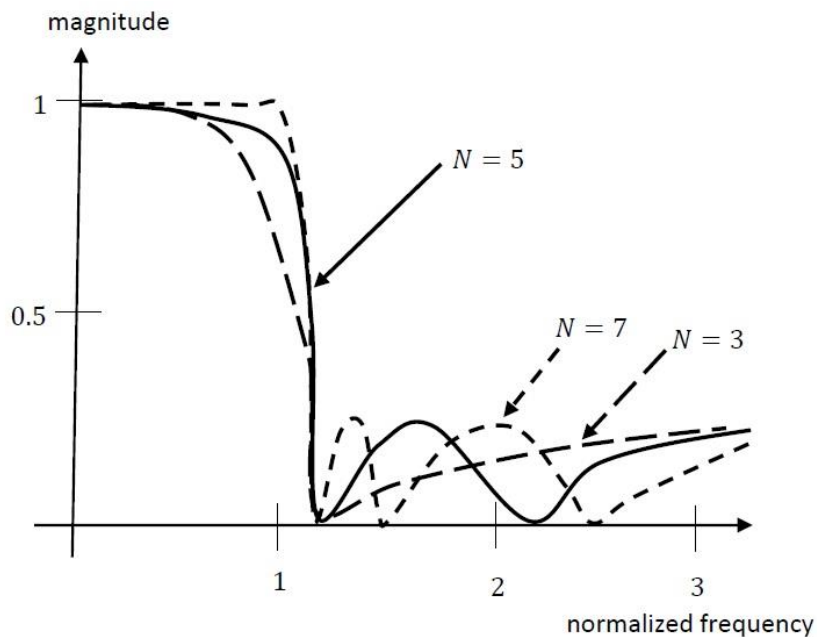


Figure 6.16. Chebyshev low-pass filter of the second type (based on: Mitra)

Using MATLAB, it is possible to design a Chebyshev filter of the second type satisfying the conditions  $R_p = 1 \text{ dB}$ ,  $R_s = 120 \text{ dB}$ ,  $\Omega_p = 2 * \pi * 1000 \text{ rad/s}$  and  $\Omega_s = 2 * \pi * 2000 \text{ rad/s}$ . More precisely,

```
help cheb2ord
```

```
[N,Wn] = cheb2ord(2*pi*1000, 2*pi*2000, 1, 120, 's')
```

```
[B,A] = cheby2(N, 120, Wn, 's')
```

```
freqs(B,A)
```

```
pzmap(B,A)
```

The Chebyshev filter has a transfer function (6.103) and the coefficients in  $B(s)$  are given by  $B$ . The coefficients in  $A(s)$  are given by  $A$ .

When realizing such a Chebyshev filter of the second type, it is clear

- all poles are located in the open left half plane giving appropriate stability properties,
- the poles are located on a circle and they are equally spread over this half circle (Due to this property, a maximally flat behaviour is obtained in the passband.),
- the transfer function contains zeros which are located at the imaginary axis of the s-plane (These zeros help to realize the stopband and they are responsible for the equiripple behaviour in this stopband).
- 

### 6.4.3 Digital low-pass filters

When considering an analog filter, the input signal and the output signal both depend on the continuous time  $t$ . The transfer function  $H_a(s)$  is described in the Laplace domain. When considering a digital filter, the input signal  $u_i[n]$  and the output signal  $u_o[n]$  both depend on the discrete time  $n$ . The transfer function is described in the Z-domain giving

$$H(z) = \frac{B(z)}{A(z)} = \frac{U_o(z)}{U_i(z)} \quad (6.104)$$

A digital filter can be designed as visualized in Figure 6.17. Notice the input signal  $u_i[n]$  and the output signal  $u_o[n]$ . The relationship

$$u_o[n] = \beta_0 u_i[n] + \beta_1 u_i[n-1] + \beta_2 u_i[n-2] + \beta_3 u_i[n-3] + \beta_4 u_i[n-4] \quad (6.105)$$

can be transformed to the Z-domain giving

$$U_o(z) = (\beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \beta_3 z^{-3} + \beta_4 z^{-4}) U_i(z). \quad (6.106)$$

This implies the transfer function equals

$$H(z) = \frac{U_o(z)}{U_i(z)} = \frac{\beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \beta_3 z^{-3} + \beta_4 z^{-4}}{1} = \frac{\beta_0 z^4 + \beta_1 z^3 + \beta_2 z^2 + \beta_3 z + \beta_4}{z^4}. \quad (6.107)$$

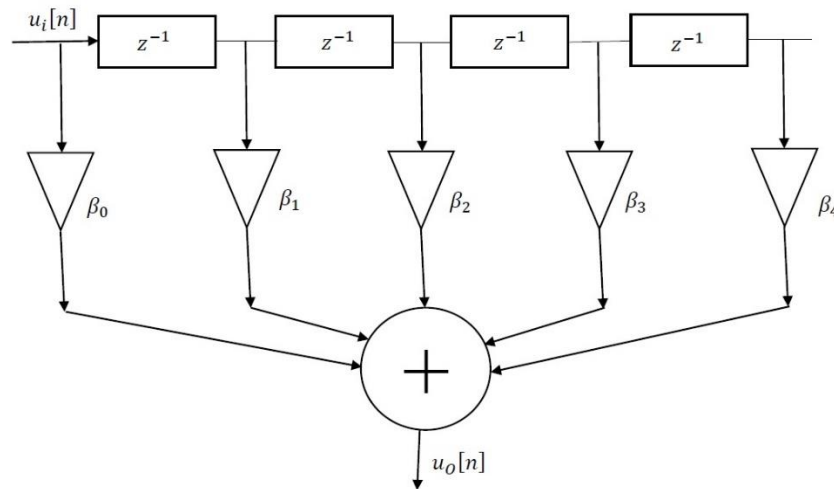


Figure 6.17. Implementation of a digital filter



#### 6.4.3.1 The s-domain and the Z-domain

When designing an analog filter, a lot of design tools are available since analog filters are already used and designed during decades. When designing a digital filter, a digital filter design methodology can be used. However, it is also possible to reformulate the digital filter specifications to a related analog filter and designing this analog filter using one of the numerous existing design methodologies. This analog filter can be transformed to the desired digital filter.

This approach needs a relationship between the s-domain and the Z-domain. More precisely,

$$z = e^{sT_S} \quad (6.108)$$

implying

$$z = \frac{e^{s\frac{T_S}{2}}}{e^{-s\frac{T_S}{2}}} \cong \frac{1+s\frac{T_S}{2}}{1-s\frac{T_S}{2}} = \frac{2+sT_S}{2-sT_S}. \quad (6.109)$$

Verify that

$$s \cong \frac{2}{T_S} \frac{z-1}{z+1} \quad (6.110)$$

which is called a bilinear transformation. Fig. 6.18 shows the relationship between the s-domain and the Z-domain and this bilinear transformation defines a bijection between both planes. The imaginary axis in the s-domain corresponds with the unit circle in the Z-domain. The open left half plane of the s-domain corresponds with the inner part of the unit circle. The open right half plane of the s-domain corresponds with the outer part of the unit circle.

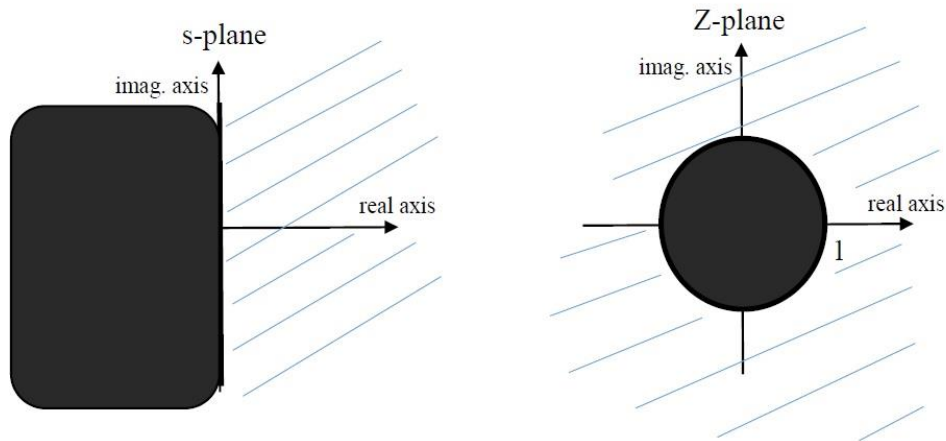


Figure 6.18. Relationship between the s-domain and the Z-domain

Although the bilinear transformation can be used as a bijection between the s-domain and the Z-domain (6.109) and (6.110)

this does not mean that for all  $s$

$$e^{sT_S} \cong \frac{1+s\frac{T_S}{2}}{1-s\frac{T_S}{2}}, \quad (6.111)$$

although this relationship is valid in case  $s = j\Omega$  and  $\Omega T_S$  is sufficiently small ( $\Omega$  must also be sufficiently small in order to respect the theorem of Shannon).

To conclude, when designing a digital filter it is possible to reformulate the digital filter specifications to a related analog filter and designing this analog filter giving the transfer function (6.103)

and calculating  $H(z)$  as

$$H(z) = \frac{B(s)}{A(s)}_{s=\frac{z-1}{T_S} \frac{z+1}{z-1}} \quad (6.112)$$

#### 6.4.3.2 Exercise

##### Analog filter

Suppose a low-pass Butterworth filter is needed having an  $\Omega_P = 2 * \pi * 1000 \text{ rad/s}$  and an  $\Omega_S = 2 * \pi * 2000 \text{ rad/s}$ . In the passband the allowed ripple equals  $R_P = 1 \text{ dB}$  and the stopband needs an attenuation  $R_S = 40 \text{ dB}$ . First an analog low-pass Butterworth filter is designed.

help buttord

```
[N,Wn] = buttord(2*pi*1000, 2*pi*2000, 1, 40, 's')
```

```
[B,A] = butter(N, Wn, 's')
```

```
[H,W]=freqs(B,A); loglog(W/(2*pi),abs(H),'g')
```

```
pzmap(B,A)
```

##### Digital filter

A digital low-pass filter is required. In the present exercise, we consider a sampling frequency  $F_S = 8000 \text{ Hz}$ .

When designing a digital filter, the relative frequencies and the relative pulsations are important. The sampling frequency  $F_S$  (expressed in Hz) corresponds with a relative frequency 2 ( $F_S/2$  corresponds with 1). The sampling frequency  $F_S$  corresponds with a relative pulsation  $2\pi \text{ rad/sample}$  ( $F_S/2$  corresponds with  $\pi \text{ rad/sample}$ ).

The digital filter ( $R_P = 1 \text{ dB}$ ,  $R_S = 40 \text{ dB}$ ) can be designed using the MATLAB instructions

```
Fs=8000
```

```
Wp=2*1000/8000
```

```
Ws=2*2000/8000
```

```
[Nd,Wd] = buttord(Wp, Ws, 1, 40)
```

```
[Bz,Az] = butter(Nd, Wd)
```

```
[Hz,Wz]=freqz(Bz,Az); loglog(Wz, abs(Hz),'g')
```

```
pzmap(Bz,Az)
```

Notice  $Wz$  is a relative pulsation whereas  $Wp$  and  $Ws$  are relative frequencies. When using instruction `loglog(Wz, abs(Hz),'g')`, the frequency behaviour is obtained using the relative pulsation on the horizontal axis. To obtain the frequency behaviour using the real frequency (expressed in  $Hz$ ) on the horizontal axis, the instruction

```
loglog((Wz*(Fs/2))/pi, abs(Hz),'g'),
```

is needed. Using the instruction

```
pzmap(Bz,Az)
```

the location of the poles and the zeros in the  $Z$ -plane are obtained. Notice these poles are all located inside the unit circle (implying no stability problems occur). The digital filter contains  $Nd = 6$  zeros at  $z = -1$ . The location of these zeros can be verified using the instruction `roots(Bz)`. Based on the locations of the poles and the zeros using instruction `pzmap(Bz,Az)`, is it possible to recognize the low-pass behaviour of the filter?

#### 6.4.4 Practical use of the bilinear transformation

Here, we will restrict ourselves to the design of digital IIR filters since these IIR filters require significantly less hardware than a FIR filter. The approach has already been mentioned, more precisely:

- formulate the specifications of the digital filter,
- using the bilinear transformation, reformulate the digital filter specifications adapted to the related analog filter,
- design the analog filter,
- transform the designed analog filter to the digital filter using the bilinear transformation.

##### 6.4.4.1 The bilinear transformation

Based on the bilinear transformation

$$s = \frac{2}{T_S} \frac{z-1}{z+1} = \frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}} \quad (6.113)$$

and since the imaginary axis  $s = j\Omega$  corresponds with the unit circle  $z = e^{j\omega}$ , one obtains that

$$j\Omega = \frac{2}{T_S} \frac{1-e^{-j\omega}}{1+e^{-j\omega}} = \frac{2}{T_S} \frac{e^{+j\frac{\omega}{2}} - e^{-j\frac{\omega}{2}}}{e^{+j\frac{\omega}{2}} + e^{-j\frac{\omega}{2}}} = \frac{2}{T_S} \tanh\left(j\frac{\omega}{2}\right) = j\frac{2}{T_S} \operatorname{tg}\left(\frac{\omega}{2}\right). \quad (6.114)$$

Since the relative pulsation  $\omega$  varies between  $-\pi \text{ rad/sample}$  and  $+\pi \text{ rad/sample}$ , the real pulsation  $\Omega$  varies between  $-\infty \text{ rad/s}$  and  $+\infty \text{ rad/s}$ . Verify and visualize this relationship using the expression

$$\Omega = \frac{2}{T_S} \operatorname{tg}\left(\frac{\omega}{2}\right). \quad (6.115)$$

Notice the transformation from  $\Omega$  to  $\omega$  and the conversion from  $\omega$  to  $\Omega$  is non linear. This implies that using the bilinear transformation is not appropriate to design the phase behaviour of a filter. Fortunately, it is realistic to design the magnitude behaviour using this bilinear transformation.

To become familiar with the bilinear transformation, the bilinear transformation will be applied on a simple low-pass filter and on a simple high-pass filter.

#### 6.4.4.2 A simple low-pass filter

Consider the analog low pass filter having a transfer function

$$H_{LP}(s) = \frac{\Omega_C}{s + \Omega_C} \quad (6.116)$$

having a  $-3 \text{ dB}$  cut-off pulsation at  $\Omega_C$ . Verify this low-pass filter is a Butterworth filter having a maximally flat behaviour. The corresponding digital filter equals

$$G_{LP}(z) = H_{LP}(s) \Big|_{s=\frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}}} = \frac{\Omega_C}{\frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}} + \Omega_C}. \quad (6.117)$$

By defining

$$\alpha = \frac{1 - \frac{\Omega_C T_S}{2}}{1 + \frac{\Omega_C T_S}{2}} \quad (6.118)$$

one also obtains that

$$\frac{1-\alpha}{2} = \frac{\frac{\Omega_C T_S}{2}}{1 + \frac{\Omega_C T_S}{2}}. \quad (6.119)$$

This allows to rewrite  $G_{LP}(z)$  as

$$G_{LP}(z) = \left(\frac{1-\alpha}{2}\right) \frac{1+z^{-1}}{1-\alpha z^{-1}}. \quad (6.120)$$

Visualize the locations of the pole and the zero in the Z-domain. Notice a zero at  $z = -1$  and a pole at  $z = \alpha$  imply  $G_{LP}(z)$  is a low-pass filter. Consider the case where  $0 < \alpha \leq 1$  and consider the case where  $\alpha < 0$ .

Due to the theorem of Shannon,

$$\Omega_C < 2\pi \frac{1}{2} \frac{1}{T_S} \quad (6.121)$$

or equivalently

$$\Omega_C \frac{T_S}{2} < \frac{\pi}{2} \quad (6.122)$$

which implies that  $-1 < \alpha < +1$ . In all cases, a low-pass filter is obtained.

#### 6.4.4.3 A simple high-pass filter

Consider the analog high pass filter having a transfer function

$$H_{HP}(s) = \frac{s}{s + \Omega_C} \quad (6.123)$$

The corresponding digital filter equals

$$G_{HP}(z) = H_{HP}(s) \Big|_{s=\frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}}} = \frac{\frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}}}{\frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}} + \Omega_C} \quad (6.124)$$

By defining

$$\alpha = \frac{1 - \frac{\Omega_C T_S}{2}}{1 + \frac{\Omega_C T_S}{2}} \quad (6.125)$$

one also obtains that

$$\frac{1+\alpha}{2} = \frac{1}{1 + \frac{\Omega_C T_S}{2}}. \quad (6.126)$$

This allows to rewrite  $G_{HP}(z)$  as

$$G_{LP}(z) = \left( \frac{1+\alpha}{2} \right) \frac{1-z^{-1}}{1-\alpha z^{-1}}. \quad (6.127)$$

Visualize the locations of the pole and the zero in the Z-domain. Notice a zero at  $z = +1$  and a pole at  $z = \alpha$  imply  $G_{HP}(z)$  is a high-pass filter. Consider the case where  $0 < \alpha \leq 1$  and consider the case where  $\alpha < 0$ .

### 6.4.5 Design of an IIR digital Butterworth filter

#### 6.4.5.1 Global approach

The approach starts by determining the specifications of the digital filter ( $\omega_p, \omega_s, R_p, R_s$ ). These digital specifications are transformed to the specifications of the related analog filter ( $\Omega_p, \Omega_s, \varepsilon, A$ ). This allows to design the order of the analog Butterworth filter. Such a Butterworth filter has two parameters i.e. the order  $N$  and  $\Omega_C$ . Based on the transfer function of the analog filter, the corresponding digital filter is obtained using the bilinear transformation.

### 6.4.5.2 The bilinear transformation

The bilinear transformation described by

$$S = \frac{2}{T_S} \frac{1-z^{-1}}{1+z^{-1}} \quad (6.128)$$

or equivalently

$$Z = \frac{\frac{2}{T_S} + S}{\frac{2}{T_S} - S} \quad (6.129)$$

can be used. Alternatively, a simplified bilinear transformation described by

$$S = \frac{1-z^{-1}}{1+z^{-1}} \quad (6.130)$$

or equivalently

$$Z = \frac{1+s}{1-s} \quad (6.131)$$

can be used. Of course either the original either the simplified transformation is used, combining the use of both transformations gives incorrect results.

### 6.4.5.3 Exercise

A digital IIR Butterworth low-pass filter with a transfer function  $G(z)$  is required. The specifications are:

- $\omega_p = 0.25 \pi \text{ rad/sample}$
- passband ripple  $R_p \leq 0.5 \text{ dB}$
- $\omega_s = 0.55 \pi \text{ rad/sample}$
- stopband attenuation  $R_s \geq 15 \text{ dB}$

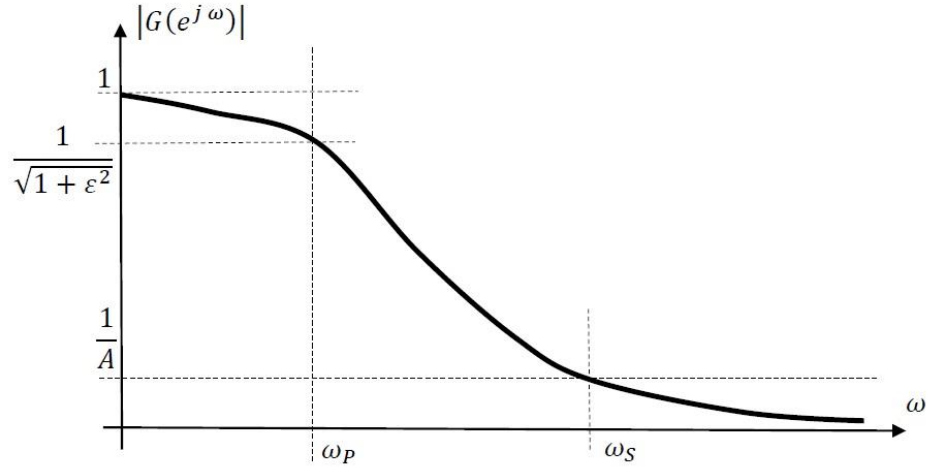


Figure 6.19. Frequency behaviour of the Butterworth low-pass filter

Solution: part 1

When considering the passband, it is important the DC gain  $|G(e^{j \cdot 0})| = 1$  and

$$20 \log_{10}|G(e^{j\omega_P})| = 20 \log_{10}|G(e^{j 0.25 \pi})| \geq -0.5 \text{ dB} . \quad (6.132)$$

When considering the stopband, it is important

$$20 \log_{10}|G(e^{j\omega_S})| = 20 \log_{10}|G(e^{j 0.55 \pi})| \leq -15 \text{ dB} . \quad (6.133)$$

These digital filter specifications must be translated to the related analog filter using the expression (based on the simplified bilinear transformation)

$$\Omega = tg\left(\frac{\omega}{2}\right) . \quad (6.134)$$

Giving  $\Omega_P = tg\left(\frac{\omega_P}{2}\right) = 0.4142$  and  $\Omega_S = tg\left(\frac{\omega_S}{2}\right) = 1.1708$ . The ratio  $\frac{\Omega_S}{\Omega_P} = \frac{1.1708}{0.4142} = 2.8267$

will be used. Since the ripple in the passband  $R_P \leq 0.5 \text{ dB}$ ,  $\epsilon$  can be obtained using expression

$$20 \log_{10}\left(\frac{1}{\sqrt{1+\epsilon^2}}\right) = -0.5 \quad (6.135)$$

giving  $\epsilon^2 = 0.1220$ . Since the stopband attenuation  $R_S \geq 15 \text{ dB}$ ,  $A$  can be obtained using expression

$$20 \log_{10}\left(\frac{1}{A}\right) = -15 \quad (6.136)$$

giving  $A = 5.6234$ .

This information will be used to determine the order of the Butterworth filter.

The order of the Butterworth filter: part 1

Consider a third order analog Butterworth filter having a transfer function

$$H_a(s) = \frac{1}{s^3 + 2s^2 + 2s + 1} = \frac{1}{(s^2 + s + 1)(s + 1)} \quad (6.137)$$

Verify  $H_a(s)$  has three poles ( $z = -1$ ,  $z = -0.5 + j 0.866$  and  $z = -0.5 - j 0.866$ ) in the open left half plane implying BIBO stability of the filter. The behaviour of this analog Butterworth filter can be written as

$$H_a(j\Omega) = \frac{1}{1 + 2j\Omega - 2\Omega^2 - j\Omega^3} \quad (6.138)$$

giving

$$|H_a(j\Omega)|^2 = \frac{1}{(1 - 2\Omega^2)^2 + (2\Omega - \Omega^3)^2} = \frac{1}{1 + \Omega^6}. \quad (6.139)$$

In general, an  $N$ -th order Butterworth filter behaves as

$$|H_a(j\Omega)|^2 = \frac{1}{1 + \left(\frac{\Omega}{\Omega_C}\right)^{2N}}. \quad (6.140)$$

In case  $\Omega = 0$ , the filter gain equals 1. In case  $\Omega = \Omega_C$ ,  $|H_a(j\Omega_C)|^2 = 1/2$  or equivalently  $|H_a(j\Omega_C)| = 1/\sqrt{2}$ . This implies  $\Omega = \Omega_C$  is the  $-3$  dB pulsation.

The Butterworth filter has a maximally flat behaviour in the passband since the first  $2N - 1$  derivatives of  $|H_a(j\Omega_C)|^2$  are zero in case  $\Omega = 0$ . The analog Butterworth filter has poles on a half circle with radius  $\Omega_C$  in the open left half plane.

#### The order of the Butterworth filter: part 2

Here, we will determine the order of the filter satisfying the specifications in Paragraph 5.3. Based on the specifications,

$$|H_a(j\Omega_P)|^2 = \frac{1}{1 + \left(\frac{\Omega_P}{\Omega_C}\right)^{2N}} = \frac{1}{1 + \varepsilon^2} \quad (6.141)$$

and

$$|H_a(j\Omega_S)|^2 = \frac{1}{1 + \left(\frac{\Omega_S}{\Omega_C}\right)^{2N}} = \frac{1}{A^2}, \quad (6.142)$$

the order  $N$  can be determined. Based on the required  $\varepsilon$ ,  $\Omega_P$ ,  $A$  and  $\Omega_S$ , it is indeed possible to determine the required order  $N$ . Since

$$\left(\frac{\Omega_P}{\Omega_C}\right)^{2N} = \varepsilon^2, \quad \left(\frac{\Omega_S}{\Omega_C}\right)^{2N} = A^2 - 1, \quad (6.143)$$

one obtains that

$$\frac{A^2 - 1}{\varepsilon^2} = \left(\frac{\Omega_S}{\Omega_P}\right)^{2N} \quad (6.144)$$

and that



$$N = \frac{1}{2} \frac{\log_{10}\left(\frac{A^2-1}{\varepsilon^2}\right)}{\log_{10}\left(\frac{\Omega_S}{\Omega_P}\right)}. \quad (6.145)$$

This reveals that an increasing  $A$ , a decreasing  $\varepsilon$  and a decreasing  $\Omega_S/\Omega_P$  implies an increasing order  $N$  (a more demanding and complex filter is needed). In the present exercise, a  $N = 2.658$  i.e. a  $N = 3$  is needed. Based on  $N = 3$ , it is possible to determine  $\Omega_C$ .

#### Solution: part 2

Finally,  $\Omega_C$  must be determined. Using  $N = 3$  in combination with  $\varepsilon^2 = 0.1220$  and  $\Omega_P = 0.4142$

$$|H_a(j\Omega_P)|^2 = \frac{1}{1+\left(\frac{\Omega_P}{\Omega_C}\right)^6} = \frac{1}{1+\varepsilon^2}. \quad (6.146)$$

provides an  $\Omega_C = 0.5881$ . Using  $N = 3$  in combination with  $A^2 = 31.6228$  and  $\Omega_S = 1.1708$ ,

$$|H_a(j\Omega_S)|^2 = \frac{1}{1+\left(\frac{\Omega_S}{\Omega_C}\right)^6} = \frac{1}{A^2}. \quad (6.147)$$

provides an  $\Omega_C = 0.6619$ .

In case the first and smaller value of  $\Omega_C$  is used ( $\Omega_C = 0.5881$ ), the specifications of the passband are reached (also at  $\Omega = \Omega_P$ ). Due to the small  $\Omega_C$ , more attenuation is obtained in the stopband than required. In case the second and larger value of  $\Omega_C$  is used ( $\Omega_C = 0.6619$ ), the specifications of the stopband are reached (also at  $\Omega = \Omega_S$ ). Due to the larger  $\Omega_C$ , less attenuation is obtained in the passband than allowed.

#### Solution: part 3

The next step requires the calculation of  $H_a(s)$ . We already know that a normalized third order Butterworth low-pass filter has a transfer function (Using the MATLAB instruction *buttap* (and *help buttap*), the zeros and the poles can be obtained. More precisely,  $[Z,P,K] = \text{buttap}(3)$  gives  $Z = []$ ,  $P = [-0.5+j 0.8660, -0.5-j 0.8660]$  and  $K=1$ .)

$$H_{an}(s) = \frac{1}{s^3+2s^2+2s+1} = \frac{1}{(s^2+s+1)(s+1)} \quad (6.148)$$

Using  $\Omega_C = 0.5881$ , one obtains that

$$H_a(s) = H_{an}\left(\frac{s}{\Omega_C}\right) = H_{an}\left(\frac{s}{0.5881}\right) = \frac{0.2034}{(s+0.5881)(s^2+0.5881s+0.3459)}. \quad (6.149)$$

#### Solution: part 4

The transfer function  $H_a(s)$  of this analog filter is used to obtain the transfer function of the digital filter based on the simplified bilinear transformation

$$S = \frac{1-z^{-1}}{1+z^{-1}}. \quad (6.150)$$

More precisely,

$$G(z) = H_a(s)_{s=\frac{1-z^{-1}}{1+z^{-1}}} \quad (6.151)$$

which gives after some calculations

$$G(z) = \frac{0.0662 (1+z^{-1})^3}{(1-0.2594 z^{-1})(1-0.6764 z^{-1}+0.3918 z^{-2})} \cdot \quad (6.152)$$

### Verification

Using MATLAB, it is possible to verify that this transfer function  $G(z)$  satisfies the filter specifications by using the instructions:

```
A1 = [1 -0.2594]
```

```
A2 = [1 -0.6764 0.3918]
```

```
A= conv(A1,A2)
```

```
B1 = [1 1]
```

```
B = conv( B1, conv(B1,B1) )*0.0662
```

```
freqz(B,A)
```

These instructions provide the frequency behaviour of  $G(z)$ . At a normalized pulsation of  $0.25 \pi \text{ rad/sample}$ , exactly a magnitude of  $-0.5 \text{ dB}$  is obtained. At a normalized pulsation of  $0.55 \pi \text{ rad/sample}$ , an attenuation of more than  $15 \text{ dB}$  is obtained.

Using the instruction `pzmap(B,A)`

it is clear the transfer function  $G(z)$  has three zeros at  $z = -1$ , one real pole at  $z = 0.2594$  and a complex conjugate pole pair at  $z = 0.338 \pm j 0.527$ . Visualize these zeros and poles in the Z-domain and verify indeed a low-pass filter is obtained.

### 6.4.6 References

1. S.K. Mitra, Digital Signal Processing: A Computer-Based Approach, second edition, Mc Graw-Hill, London, 2002.
2. S.K. Mitra, Digital Signal Processing: A Computer-Based Approach, third edition, Mc Graw-Hill, London, 2006.

## 6.5 EMC filters

### 6.5.1 Introduction

An electronic device can emit disturbances using the conductors of the electrical power grid. For instance, when considering a frequency converter the conducted emission measurement in the frequency range from 150 *kHz* to 30 *MHz* is given in Fig. 6.20. Notice the emission level is much higher than the emission limit allowed by the EMC directives (CISPR 11).

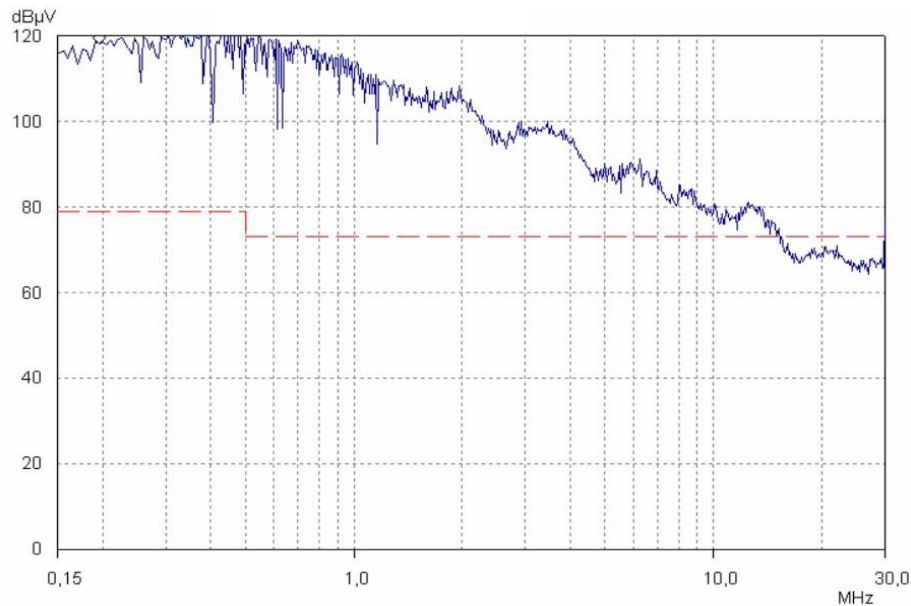


Figure 6.20. Conducted emission measurement

The radiated emission level can be reduced by mounting an EMC filter between the power grid and the frequency converter as visualised in Fig. 6.21. Such an EMC filter mainly behaves as a low pass filter i.e. the filter passes the 50 *Hz* signal but blocks the high frequent components as much as possible.

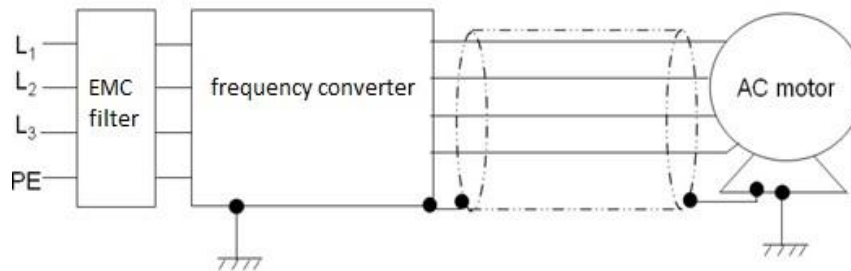


Figure 6.21. The use of an EMC filter to reduce the emission levels

Such an EMC filter is a passive filter and can be built as visualised in Figure 6.22 although a lot of other configurations are possible. It is important the EMC filter blocks high frequent common mode currents and high frequent differential mode currents.

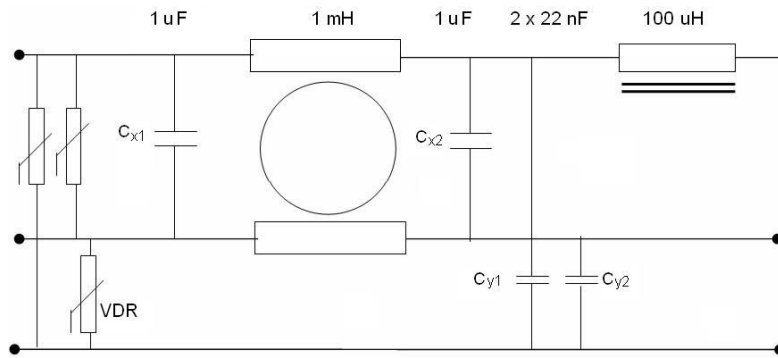


Figure 6.22. Internal circuit of a passive EMC filter (source: Deconinck et al.)

Due to such a passive filter (having a proper high frequent grounding), the conducted emission level reduces from the measurement in Fig. 6.20 to the measurement in Fig. 6.23. Notice the emission level in Fig. 6.23 does not exceed the emission limit allowed by the EMC directives (CISPR 11).

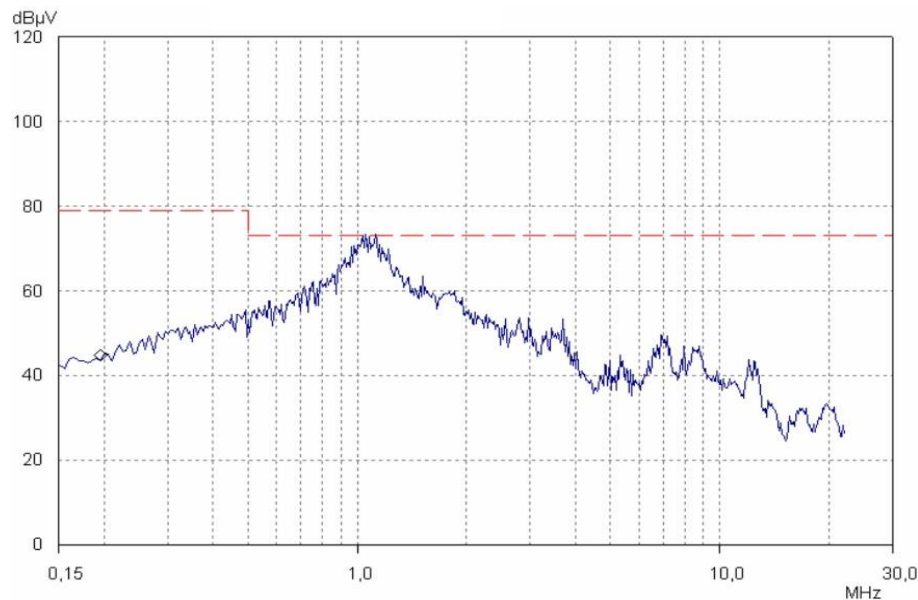


Figure 6.23. Conducted emission level when using an EMC filter

### 6.5.2 The behaviour of passive filters 1

In the present text, the behaviour of a number of passive filters will be studied in a more systematic way. The steady state behaviour will be studied in case of sinusoidal voltages and currents by calculating the transfer function. The transfer function equals the ratio of the output voltage and the input voltage. Since the steady state behaviour is studied, no Laplace transforms are used but the output voltage and the input voltage are represented by complex numbers.

### 6.5.2.1 Transfer function of a filter

Fig. 6.24 visualises a  $\pi$  filter structure containing three impedances  $\bar{Z}_1$ ,  $\bar{Z}_2$  and  $\bar{Z}_3$ . A non ideal voltage source contains an ideal voltage source  $\bar{V}_S$  and a series impedance  $\bar{Z}_S$ . At load side, the load equals the impedance  $\bar{Z}_L$ .

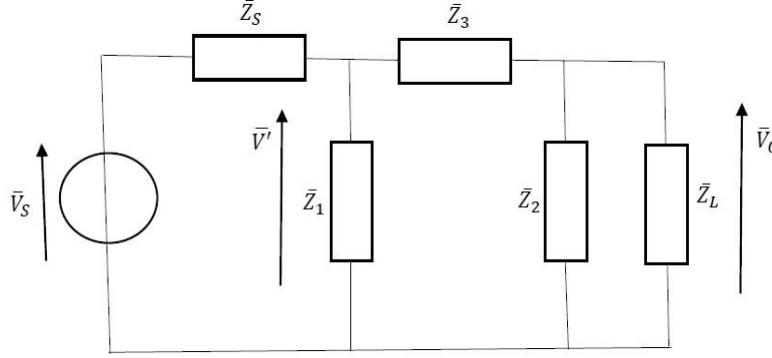


Figure 6.24. Structure of a  $\pi$  filter

In case the steady state behaviour of Fig. 6.24 will be studied, impedances  $\bar{Z}_1$ ,  $\bar{Z}_2$ ,  $\bar{Z}_3$ ,  $\bar{Z}_S$  and  $\bar{Z}_L$  are complex numbers. Since by defining

$$\bar{Z}_4 = \frac{\bar{Z}_2 \bar{Z}_L}{\bar{Z}_2 + \bar{Z}_L}, \quad (6.153)$$

the relationship

$$\frac{\bar{V}_I}{\bar{V}_S} = \frac{\bar{Z}_1 \parallel (\bar{Z}_3 + \bar{Z}_4)}{[\bar{Z}_1 \parallel (\bar{Z}_3 + \bar{Z}_4)] + \bar{Z}_S} \quad (6.154)$$

is valid. Since

$$\frac{\bar{V}_O}{\bar{V}_I} = \frac{\bar{Z}_4}{\bar{Z}_3 + \bar{Z}_4} \quad (6.155)$$

one obtains that

$$\frac{\bar{V}_O(\bar{Z}_3 + \bar{Z}_4)}{\bar{Z}_4 \bar{V}_S} = \frac{\bar{Z}_1(\bar{Z}_3 + \bar{Z}_4)}{(\bar{Z}_1 + \bar{Z}_3 + \bar{Z}_4)} \cdot \frac{1}{\frac{\bar{Z}_1(\bar{Z}_3 + \bar{Z}_4)}{(\bar{Z}_1 + \bar{Z}_3 + \bar{Z}_4)} + \bar{Z}_S}. \quad (6.156)$$

Finally, the transfer function equals

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{\bar{Z}_1 \bar{Z}_4}{\bar{Z}_1 \bar{Z}_3 + \bar{Z}_1 \bar{Z}_4 + \bar{Z}_1 \bar{Z}_S + \bar{Z}_3 \bar{Z}_S + \bar{Z}_4 \bar{Z}_S}. \quad (6.157)$$

### 6.5.2.2 Low pass filter

In a lot of applications, the impedances  $\bar{Z}_S$  and  $\bar{Z}_L$  are ohmic i.e.  $\bar{Z}_S = R_S$  and  $\bar{Z}_L = R_L$ . The  $\pi$  filter in Fig. 6.24 realises a low pass filter in case

$$\bar{Z}_1 = \frac{1}{j\omega C_1}, \bar{Z}_2 = \frac{1}{j\omega C_2}, \bar{Z}_3 = j\omega L. \quad (6.158)$$

Using these assumptions, the transfer function equals

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{R_L}{-j\omega^3 R_S R_L C_1 C_2 L - \omega^2 L(R_L C_2 + R_S C_1) + j\omega(L + R_S R_L C_2 + R_S R_L C_1) + R_L + R_S}. \quad (6.159)$$

In case the pulsation  $\omega = 0$ , the transfer function reduces to

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{R_L}{R_L + R_S} \quad (6.160)$$

since the capacitors of  $\bar{Z}_1$  and  $\bar{Z}_2$  behave as an open circuit and the inductor of  $\bar{Z}_3$  behaves as a short circuit. Indeed, a voltage division is obtained by the resistors  $R_S$  and  $R_L$ .

To simplify the calculations suppose  $C_1 = C_2 = C$  and

$$R_S = R_L = \sqrt{\frac{L}{2C}}, \quad (6.161)$$

the transfer function can be written as

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{1}{2} \cdot \frac{1}{-j\omega^3 \frac{\sqrt{LC}^3}{\sqrt{2}} - \omega^2 \sqrt{LC}^2 + \sqrt{2} j\omega \sqrt{LC} + 1} = \frac{1}{2} \frac{1}{(1 - \omega^2 \sqrt{LC}^2) + j\left(\sqrt{2} \omega \sqrt{LC} - \omega^3 \frac{\sqrt{LC}^3}{\sqrt{2}}\right)}. \quad (6.162)$$

By calculating this transfer function, the phase difference between  $\bar{V}_O$  and  $\bar{V}_S$  is known and also

$$\left| \frac{\bar{V}_O}{\bar{V}_S} \right| = \frac{1}{2} \frac{1}{\sqrt{1 + \omega^6 \frac{L^3 C^3}{8}}} \quad (6.163)$$

can be calculated. When applying a DC voltage at the input,  $|\bar{V}_O| = 0.5 |\bar{V}_S|$ . Based on this amplitude behaviour, also the cutoff pulsation of the low pass filter is obtained. More precisely, at the cutoff pulsation the filter has an additional attenuation of  $\sqrt{2}$  in comparison with the DC behaviour. This means

$$\omega^6 \frac{L^3 C^3}{8} = 1 \quad (6.164)$$

or equivalently

$$\omega = \frac{\sqrt{2}}{\sqrt{LC}}. \quad (6.165)$$

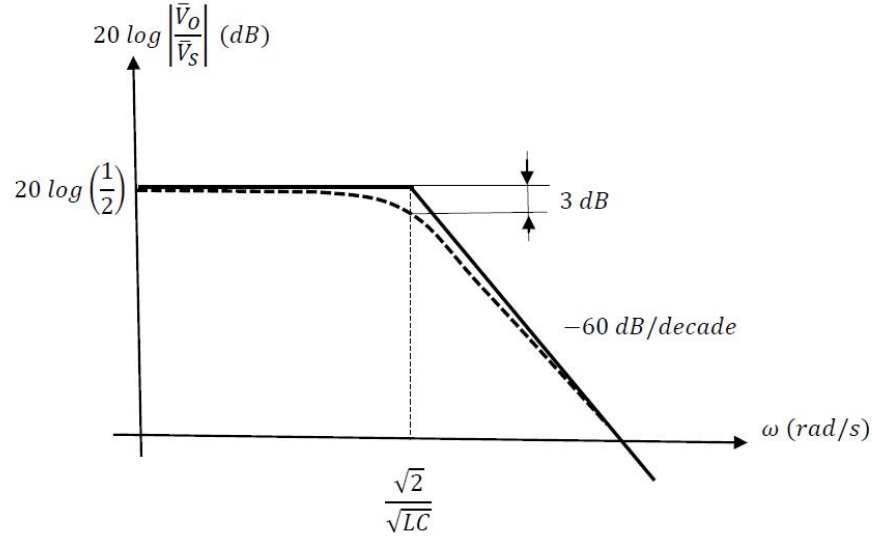


Figure 6.25. Amplification of the  $\pi$  low pass filter

The amplitude behaviour of the  $\pi$  filter is visualised in Fig. 6.25. For pulsations lower than the cutoff pulsation, the amplification is almost constant i.e. the filter passes the low frequency components. For pulsations higher than the cutoff pulsation, the amplification decreases steadily as the pulsation increases.

By expressing the amplification as

$$20 \log \left| \frac{\bar{V}_O}{\bar{V}_S} \right|, \quad (6.166)$$

a logarithmic scale expressed in  $dB$  is obtained. For pulsations higher than the cutoff pulsation, the amplification decreases with  $60 \text{ dB}$  each decade. When the pulsation becomes 10 times larger, the amplification decreases with a factor 1000.

### 6.5.2.3 High pass filter

In a lot of applications, the impedances  $\bar{Z}_S$  and  $\bar{Z}_L$  are ohmic i.e.  $\bar{Z}_S = R_S$  and  $\bar{Z}_L = R_L$ . The  $\pi$  filter in Figure 6.25 realises a high pass filter in case

$$\bar{Z}_1 = j\omega L_1, \bar{Z}_2 = j\omega L_2, \bar{Z}_3 = \frac{1}{j\omega C}. \quad (6.167)$$

Using these assumptions and in case  $L_1 = L_2 = L$ , the transfer function equals

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{-j\omega^3 L^2 C R_L}{-j\omega^3 L^2 C (R_S + R_L) - \omega^2 L (L + 2C R_S R_L) + j\omega L (R_L + R_S) + R_L R_S}. \quad (6.168)$$

In case the pulsation  $\omega = 0$ , no output voltage is obtained. Indeed, the capacitor of  $\bar{Z}_3$  behaves as an open circuit and the inductors of  $\bar{Z}_1$  and  $\bar{Z}_2$  behave as short circuits. When considering sufficiently high frequencies (a high pulsation  $\omega$ ), the transfer function reduces to

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{R_L}{R_L + R_S} \quad (6.169)$$

since a voltage division is obtained by the resistors  $R_S$  and  $R_L$ . Assuming that

$$R_S = R_L = \sqrt{\frac{L}{2C}}, \quad (6.170)$$

the transfer function can be written as

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{-j\omega^3\sqrt{LC}^3\sqrt{2}}{-j\omega^3\sqrt{LC}^32\sqrt{2}-4\omega^2\sqrt{LC}^2+2\sqrt{2}j\omega\sqrt{LC}+1} = \frac{-j\omega^3\sqrt{LC}^3\sqrt{2}}{(1-4\omega^2\sqrt{LC}^2)+j2\sqrt{2}(\omega\sqrt{LC}-\omega^3\sqrt{LC}^3)}. \quad (6.171)$$

By calculating this transfer function, the phase difference between  $\bar{V}_O$  and  $\bar{V}_S$  is known and also

$$\left|\frac{\bar{V}_O}{\bar{V}_S}\right| = \frac{\omega^3\sqrt{LC}^3\sqrt{2}}{\sqrt{1+8\omega^6L^3C^3}} \quad (6.172)$$

can be calculated. When applying a sinusoidal voltage at the input having a sufficiently high pulsation,  $|\bar{V}_O| = 0.5 |\bar{V}_S|$ . Based on this amplitude behaviour, also the cutoff pulsation of the high pass filter is obtained. More precisely, at the cutoff pulsation the filter has an additional attenuation of  $\sqrt{2}$  in comparison with the high frequent behaviour. This means

$$8\omega^6L^3C^3 = 1 \quad (6.173)$$

or equivalently

$$\omega = \frac{1}{\sqrt{2}\sqrt{LC}}. \quad (6.174)$$

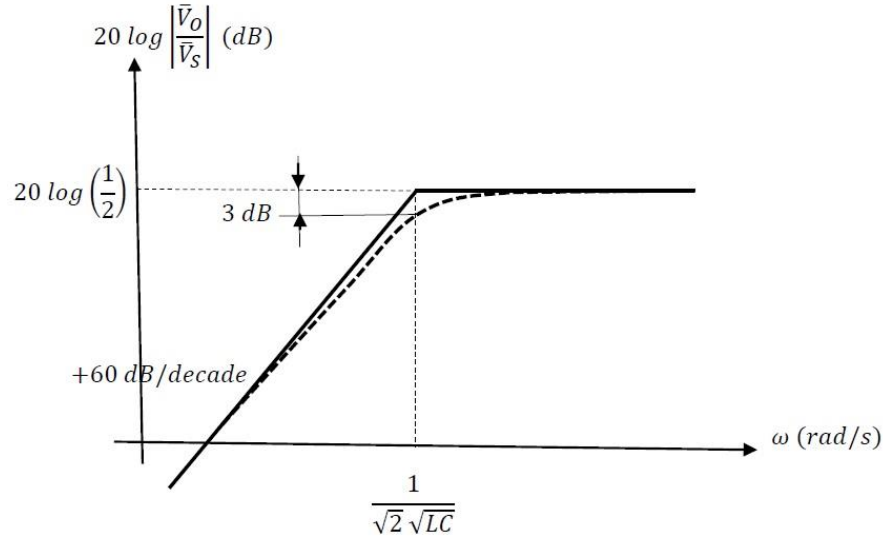


Figure 6.26. Amplification of the  $\pi$  high pass filter

The amplitude behaviour of the  $\pi$  filter is visualised in Fig. 6.26. For pulsations higher than the cutoff pulsation, the amplification is almost constant i.e. the filter passes the high frequency components. For pulsations lower than the cutoff pulsation, the amplification decreases steadily as the pulsation decreases.

By expressing the amplification as



$$20 \log \left| \frac{\bar{V}_O}{\bar{V}_S} \right|, \quad (6.175)$$

a logarithmic scale expressed in  $dB$  is obtained. For pulsations lower than the cutoff pulsation, the amplification increases with  $60 dB$  each decade as the pulsation increases. When the pulsation becomes 10 times larger, the amplification increases with a factor 1000.

#### 6.5.2.4 Exercise 1

Suppose  $L = 1 mH$  and  $C = 10 \mu F$ , calculate (6.175), as a function of the pulsation  $\omega$ . Use MATLAB to plot the amplification as a function of the pulsation  $\omega$  when considering

- the low pass  $\pi$  filter,
- the high pass  $\pi$  filter.
- 

#### 6.5.2.5 Exercise 2

Suppose the low pass filter and the high pass filter in the previous paragraph contain real world components. More precisely, suppose

- the inductor does not behave as an ideal inductor but behaves as a parallel RLC circuit (where the additional resistor accounts for the losses in the inductor and the additional capacitor models the parasitic capacitances in the inductor),
- the capacitor does not behave as an ideal capacitor but behaves as a series RLC circuit (where the additional resistor accounts for the losses in the capacitor and the additional inductance models the parasitic inductance in the capacitor).

Based on the expression

$$\frac{\bar{V}_O}{\bar{V}_S} = \frac{\bar{Z}_1 \bar{Z}_4}{\bar{Z}_1 \bar{Z}_3 + \bar{Z}_1 \bar{Z}_4 + \bar{Z}_1 \bar{Z}_S + \bar{Z}_3 \bar{Z}_S + \bar{Z}_4 \bar{Z}_S}, \quad (6.176)$$

calculate the amplification of the filter as a function of the pulsation  $\omega$  using MATLAB. Give a physical interpretation of the simulation results.

### 6.5.3 The behaviour of passive filters 2

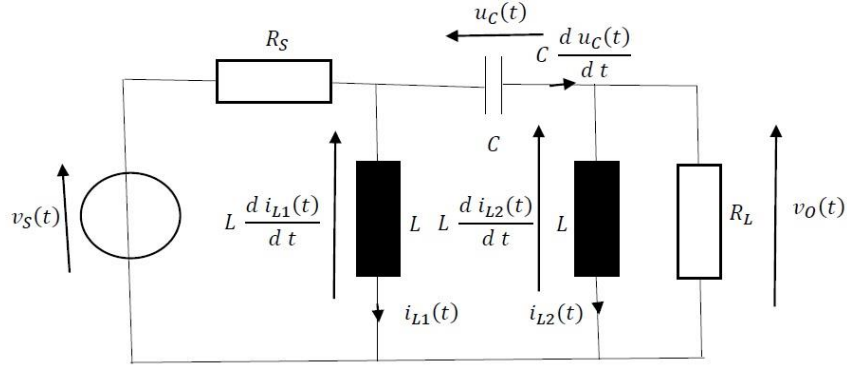


Figure 6.27. Behaviour of a  $\pi$  filter

In the previous paragraphs, the steady state behaviour of a  $\pi$  filter has been studied in case of sinusoidal voltages and currents. By studying the transfer function in the complex domain, the main properties of the filter are known. In order to investigate the behaviour of the  $\pi$  filter in case of non sinusoidal voltages and currents a differential equation is needed. Such a differential equation can also be used to calculate the transient behaviour. Using state space analysis, a higher order differential equation can be replaced by a set of first order differential equations.

#### 6.5.3.1 The behaviour of a $\pi$ filter

Notice the  $\pi$  filter in Fig. 6.27 contains three energy reservoirs (twice an inductor  $L$  and once a capacitor  $C$ ). When describing the behaviour of the filter using the state space approach, three states will be needed i.e. three first order differential equations will be needed.

The first state is the voltage  $u_C(t)$  across the capacitor  $C$ , the second state is the current  $i_{L1}(t)$  flowing in the first inductor  $L$  and the third state is the current  $i_{L2}(t)$  in the second inductor. Using the voltage law of Kirchoff, one obtains that

$$L \frac{d i_{L2}(t)}{d t} + u_C(t) = L \frac{d i_{L1}(t)}{d t} \quad (I). \quad (6.177)$$

Using the current law of Kirchoff, two relationships are obtained. More precisely,

$$i_{L1}(t) + i_{L2}(t) + \frac{L}{R_L} \frac{d i_{L2}(t)}{d t} = \frac{1}{R_S} \left( v_S(t) - L \frac{d i_{L1}(t)}{d t} \right) \quad (II) \quad (6.178)$$

and

$$C \frac{d u_C(t)}{d t} = i_{L2}(t) + \frac{v_O(t)}{R_L} \quad (III). \quad (6.179)$$

When considering the output voltage across the inductor  $L$ , one obtains that

$$v_O(t) = L \frac{d i_{L2}(t)}{d t} \quad (IV). \quad (6.180)$$

In order to use the state space approach, the final goal is writing down the behaviour of the network in Fig. 6.27 as

$$\frac{d}{dt} \begin{bmatrix} u_C(t) \\ i_{L1}(t) \\ i_{L2}(t) \end{bmatrix} = A \begin{bmatrix} u_C(t) \\ i_{L1}(t) \\ i_{L2}(t) \end{bmatrix} + B v_S(t) \quad (6.181)$$

and

$$v_O(t) = C \begin{bmatrix} u_C(t) \\ i_{L1}(t) \\ i_{L2}(t) \end{bmatrix} + D v_S(t) \quad (6.182)$$

with appropriate matrices  $A$ ,  $B$ ,  $C$  and  $D$ .

### 6.5.3.2 State Space analysis of a $\pi$ filter

By multiplying expression (II) by  $R_S$  and adding it to (I), one obtains that

$$-\left(L + \frac{R_S L}{R_L}\right) \frac{d i_{L2}(t)}{dt} = u_C(t) + R_S i_{L1}(t) + R_S i_{L2}(t) - v_S(t). \quad (6.183)$$

Taking (IV) into account, this expression also implies that

$$v_O(t) = -\frac{R_L}{R_S + R_L} (u_C(t) + R_S i_{L1}(t) + R_S i_{L2}(t) - v_S(t)). \quad (6.184)$$

Taking (III) into account, this expression also implies that

$$C \frac{d u_C(t)}{dt} = i_{L2}(t) - \frac{1}{R_S + R_L} (u_C(t) + R_S i_{L1}(t) + R_S i_{L2}(t) - v_S(t)). \quad (6.185)$$

By multiplying expression (II) by  $R_L$  and subtracting (I), one obtains that

$$-\left(L + \frac{R_L L}{R_S}\right) \frac{d i_{L1}(t)}{dt} = -u_C(t) + R_L i_{L1}(t) + R_L i_{L2}(t) - \frac{R_L}{R_S} v_S(t). \quad (6.186)$$

These four expressions imply

$$\frac{d}{dt} \begin{bmatrix} u_C(t) \\ i_{L1}(t) \\ i_{L2}(t) \end{bmatrix} = A \begin{bmatrix} u_C(t) \\ i_{L1}(t) \\ i_{L2}(t) \end{bmatrix} + B v_S(t) \quad (6.187)$$

and

$$v_O(t) = C \begin{bmatrix} u_C(t) \\ i_{L1}(t) \\ i_{L2}(t) \end{bmatrix} + D v_S(t) \quad (6.188)$$

with

$$\begin{aligned}
A &= \begin{bmatrix} -\frac{1}{(R_S + R_L) C} & -\frac{R_S}{(R_S + R_L) C} & +\frac{R_L}{(R_S + R_L) C} \\ +\frac{R_S}{(R_S + R_L) L} & -\frac{R_L R_S}{(R_S + R_L) L} & -\frac{R_L R_S}{(R_S + R_L) L} \\ -\frac{R_L}{(R_S + R_L) L} & -\frac{R_L R_S}{(R_S + R_L) L} & -\frac{R_L R_S}{(R_S + R_L) L} \end{bmatrix} \\
B &= \begin{bmatrix} \frac{1}{(R_S + R_L) C} \\ \frac{R_L}{(R_S + R_L) L} \\ \frac{R_L}{(R_S + R_L) L} \end{bmatrix} \\
C &= \begin{bmatrix} -\frac{R_L}{(R_S + R_L)} & -\frac{R_L R_S}{(R_S + R_L)} & -\frac{R_L R_S}{(R_S + R_L)} \end{bmatrix} \\
D &= \frac{R_L}{R_S + R_L}
\end{aligned} \tag{6.189}$$

#### 6.5.3.3 Exercise

When considering Fig. 6.27, suppose  $L = 1 \text{ mH}$  and  $C = 10 \text{ }\mu\text{F}$ . Assuming that

$$R_S = R_L = \sqrt{\frac{L}{2C}}, \tag{6.190}$$

calculate the matrices  $A$ ,  $B$ ,  $C$  and  $D$  of the state space model. Use SIMULINK (MATLAB) to calculate the behaviour of the high pass  $\pi$  filter in Figure 6.27.

#### 6.5.4 References

1. Deconinck G., Saey P. (ed.), *Elektrotechniek en automatisering: Capita Selecta*, Academia Press, Gent, 2011, section 8: *Parasitaire componenten in vermogenelektronische schakelingen* (Peuteman J., Knockaert J.), pp. 275 - 295.
2. Deconinck G., Saey P. (ed.), *Elektrotechniek en automatisering: Capita Selecta*, Academia Press, Gent, 2011, section 9: *EMC en elektrische aandrijvingen* (Peuteman J., Knockaert J.), pp. 297 - 325.
3. K.L. Kaiser, *Electromagnetic Compatibility Handbook*, CRC Press, London, 2005.
4. R.L. Ozenbaugh, *EMI Filter Design*, CRC Press, London, 2001.