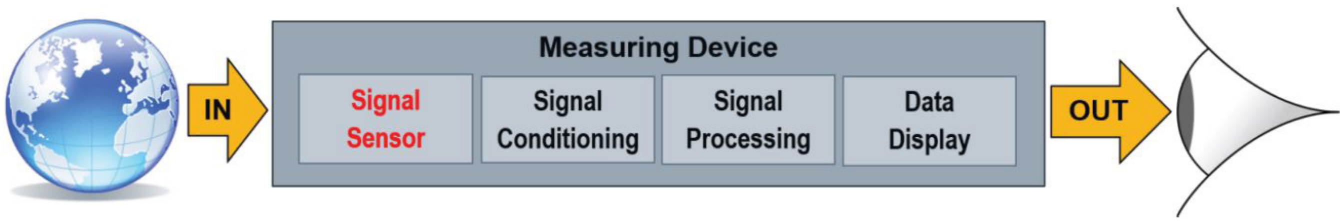
Fundamentals of Electrical Measurements 1-2



István Gyurcsek

Fundamentals of Electrical Measurements 1-2

Pécs

2019

The Fundamentals of Electrical Measurements 1-2 course material was developed under the project EFOP 3.4.3-16-2016-00005 "Innovative university in a modern city: open-minded, value-driven and inclusive approach in a 21st century higher education model".

EFOP 3.4.3-16-2016-00005 "Innovative university in a modern city: open-minded, value-driven and inclusive approach in a 21st century higher education model".

István Gyurcsek

Fundamentals of Electrical Measurements 1-2

Pécs

2019

A Fundamentals of Electrical Measurements 1-2 tananyag az EFOP-3.4.3-16-2016-00005 azonosító számú,
„Korszerű egyetem a modern városban: Értékközpontúság, nyitottság és befogadó szemlélet egy 21. századi felsőoktatási modellben” című projekt keretében valósul meg.

SZÉCHENYI 2020

PÉCSI TUDOMÁNYEGYETEM
UNIVERSITY OF PÉCS

DR. ISTVÁN GYURCSEK

FUNDAMENTALS OF ELECTRICAL MEASUREMENTS

DEPARTMENT OF ELECTRICAL NETWORKS
FACULTY OF ENGINEERING AND INFORMATION TECHNOLOGY
UNIVERSITY OF PÉCS

2019

ISBN 978-963-429-384-2

THIS COURSE MATERIAL WAS DEVELOPED UNDER THE PROJECT "MODERN UNIVERSITY IN A MODERN CITY: MODEL FOR VALUE-ORIENTED, OPENNESS AND INCLUSIVE APPROACHES IN 21. CENTURY HIGHER EDUCATION". REGISTRATION NUMBER: EFOP-3.4.3-16-2016-00005

EFOP-3.4.3-16-2016-00005 MODERN UNIVERSITY IN A MODERN CITY:
MODEL FOR VALUE-ORIENTED, OPENNESS AND INCLUSIVE
APPROACHES IN 21. CENTURY HIGHER EDUCATION

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE

Contents

Introduction and Scope of Work **5**

1. Electrical Measurements Essentials **7**

1.1 Main Terms and Definitions *7*

1.1.1 Measurement and Metrology 7

1.1.2 Measurement Procedure 8

1.1.3 Process of Measurement Modelling 10

1.1.4 Measurement Architectures 11

1.2 Methods of Measurement *13*

1.2.1 Direct and Indirect Methods 13

1.2.2 Null and Deflection Methods 14

1.2.3 Compensation Methods 15

1.2.4 Comparative Method 16

1.2.5 Bridge Circuits 17

1.2.6 Substitution Method 19

1.3 Uncertainty of Measurements *20*

1.3.1 Reasons for Uncertainty 21

1.3.2 Classification of Measuring Errors 22

1.3.3 Measurement Series 23

1.3.4 Accuracy Class 25

1.3.5 Principles in Propagation of Errors 27

1.4 Standards of Quantities *29*

1.4.1 Basic Terms and Concepts 29

1.4.2 International System of Units (SI) 30

2. Classic Electrical Measurements **34**

2.1 Indicating Measuring Instruments *34*

2.1.1 General Overview 34

2.1.2 Terms and Definitions 35

2.1.3 Moving Coil Meters 36

2.1.4 The Moving Iron Meters 42

2.1.5 The Electrodynamic Meters 44

2.2 Waveform Measuring Instruments *49*

2.2.1 Oscilloscope Basics 49

2.2.2 Features of Oscilloscopes 49

2.2.3 The Types of Oscilloscopes 50

2.2.4 Main Units and Controls 52

2.2.5 X-Y Operation Mode 56

2.3 Balanced and Unbalanced Bridge Circuits 58

2.3.1 Balance and Sensitivity 58

2.3.2 Null Type (Balanced) Bridges 61

2.3.3 Deflection Type (Unbalanced) Bridges 67

2.3.4 Alternatives to Bridge Circuits 69

2.4 Compensation Method and Comparators 70

3. Electrical Measurement Solutions 73

3.1 DC Power and Resistance Measurement 73

3.1.1 Ammeter-Voltmeter Method 73

3.1.2 Kelvin (Four-Wire) Method 74

3.2 AC Power and Impedance Measurement 77

3.2.1 Instrument Transformers 77

3.2.2 Three-Voltmeter Method 79

4. Sensor Theory 82

4.1 Positioning of Sensors 82

4.1.1 Two-Port Sensor Models 84

4.1.2 Sensor Classification 84

4.2 Modelling of Sensors 86

4.4.1 Practical Examples of Sensor Models 87

4.3 Sensor Characteristics 88

4.3.1 Static Sensor Characteristics 89

4.3.2 Dynamic Sensor Characteristics 93

4.4 Sensor Groups Overview 94

4.5 Sensor signal transmission overview 94

4.6 Sensor Application Examples 95

4.6.1 LVDT Displacement Sensor 95

4.6.2 Motion Sensor with Optocoupler 96

4.6.3 Motion, Stress, Tension and Strain Sensors 96

4.6.4 Accelerometers 96

4.6.5 Photo Sensors 97

5. Basic Sensor Networks 98

5.1 Measurement Network Architecture 98

5.1.1 Asymmetric Signal Sources 98

5.1.2 Symmetric Signal Sources 99

5.1.3 Asymmetric Signal Receivers 100

5.1.4 Symmetric Signal Receivers 101

5.1.5 Signal Source to Receiver Interconnection 102

5.2 Noise in Electrical Circuits 107

5.2.1 Noisy Signals and Noise Reduction 107

5.2.2 Classification of Electric Noise 109

5.2.3 Noise Sources 113

5.3 Groundings and Earths 114

5.3.1 Definitions 114

5.3.2 Ground Types 116

5.3.3 Grounding Symbols 118

5.3.4 The Ground Loop 119

6. Sensor Measurement Solutions 123

6.1 Temperature Sensors 123

6.1.1 Thermocouples 123

6.1.2 Resistive Temperature Measuring Devices 125

6.1.3 Thermistors 126

6.1.4 Silicon Diodes 131

6.1.5 Pyrometers or Infrared Sensors 134

6.1.6 Other Temperature Detectors 134

6.2 Radiation Sensors 136

6.2.1 Measurement of Visible Light 136

6.2.2 Infrared Measurements 151

6.2.3 Nuclear Measurements 154

6.3 Mechanical Sensors 160

6.3.1 Strain Gauges 160

6.3.2 Piezoelectric Accelerometers 164

6.3.3 Position and Displacement Sensors 165

7. Reference Sources and Recommended Materials 169

7.1 In English 169

7.2 In Hungarian 169

Introduction and Scope of Work

Accurate measurement is absolutely necessary to both scientific research and numerous other engineering work. To put into proportion the importance of measurement, and especially, electrical measurement, let's start with a citation from one of the most famous researchers and scientists in modern physics.

> *'There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery.' (Enrico Fermi)*

So no matter if the result confirms, or it is contrary to our hypothesis, we have to be skilled in measurement technologies to be able to prove that the result we get is accurate. Electrical measurement and sensor theory are the fundamental principles upon which all branches of electrical engineering are built. All branches of electrical engineering like power, electric machines, control, electronics, communications and instrumentation requires electrical measurement and it is necessary to understand the fundamentals of measurement theory. Therefore, to understand basic measurement technologies, is not only one of the most important things for students studying electrical engineering and mechanical engineering, but is always an excellent starting point for the fundamentals of any field of engineering.

Theories of measurement technology are also valuable for students specializing in other branches of physical science because the principles provide a good model for the study of measurements in general, and also because applied mathematics, physics and topology are also involved.

This textbook covers the most important commonly used electrical measurement principles and technologies. Because of the vast amount of theory and practical experience in this field, it was a huge challenge to collect and summarize even the most basic measurement methods and technologies in a short textbook. However, by focusing on the most essential parts, many interesting and important practical sections of electrical measurements are not discussed in this textbook. The topics selected and how the book has been structured is based on my extensive educational experience and previously published textbooks.

I believe that this collection of the principles in measurement, covering the most important parts of measurement technologies will support you, the students and the readers, in understanding and practicing the basics and give you the proper background to build up a deeper understanding in practical engineering.

Your feedback regarding the structure and content of this material is welcome.

I wish to express my special thanks to my colleagues for their help in the creation of this textbook - I couldn't have made such a concise selection of the topics without their valuable support.

In addition, I would like to thank the leadership of the University of Pécs, Faculty of Engineering and Information Technology for giving me the opportunity to write this book that I believe will be a useful tool for teaching both international and Hungarian students the basics of measurement technologies.

Last, but not least, I want to express appreciation to my friends and family. For the long period when I was occupied with collecting information and preparing this material, I tried their patience, though unintentionally. I would like to thank them for having so much patience - they were extremely helpful, and I really appreciate it.

István Gyurcsek
Author

1. Electrical Measurements Essentials

1.1 Main Terms and Definitions

1.1.1 Measurement and Metrology

Let us start with the definition of electrical measurements. The measurement is a cognitive process of *gathering the information* from the physical world for practical purposes. In this process a value of a *quantity is determined* (in defined time and conditions) by comparing it, with *known uncertainty,* to the *standard reference* value. The uncertainty of measurement and the way of comparison to the standard reference value both depend on the applied process and the instrument used. Fig. 1.1 illustrates two instruments for length measurement with different methods and uncertainty.

Apart from the term '*measurements*', other terms are also in use, for example '*scientific instrumentation*' (or just instrumentation) and '*metrology*'. The term of *testing* is often used as a synonym for measurements, and *metrology* is considered as the science of measurements.

Figure 1.1 Instruments for measuring length

For determining the measurement, we can distinguish different levels of the gathered info from the physical world to be examined. The highest level of that info is *knowledge* or expertise. Knowledge is the organized information framework, containing new derived info, based on the gathered elements. The *information* itself is the collection of facts and details about the examined, measured, observed and/or experienced subject or system. The next level is the *data* level. Data is the logical representation of information, usually represented by series of symbols. So-called *signal* is at the lowest level in this grouping with the meaning of physical representation of data. Signal is usually electronic code.

According to the different levels of the representation of the physical world defined above, most technical measurements are carried out at the signal level i.e. at the level of electronic codes.

Electrical measurement is increasingly important in today's world. For example, in the automotive industry, a number of sensors, i.e. ABS, ESP, GPS, airbags, rain sensors, etc. are used, as illustrated in Fig. 1.2. The connection of sensors to the control system are established by using special interfaces, like Control Area Network (CAN) bus or even RS232. In modern cars, equipped with a large number of sensors, several CAN lines are applied to reduce traffic through each line.

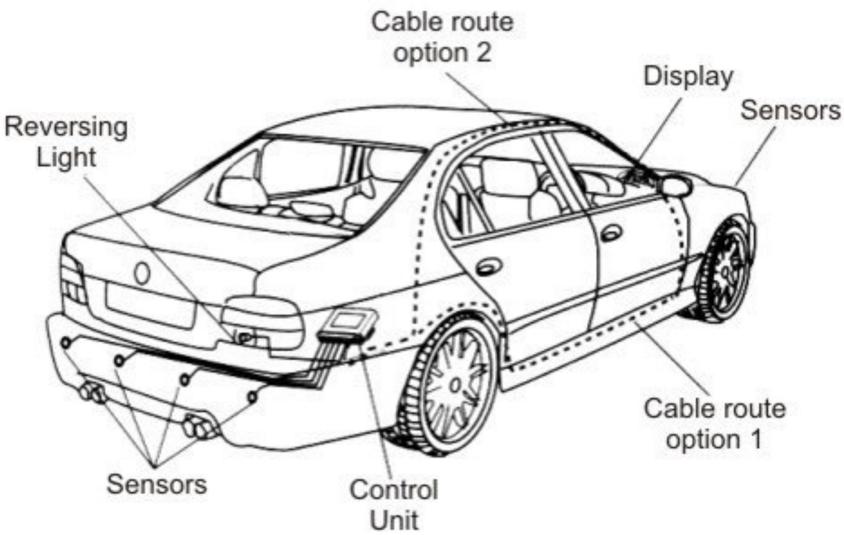


Figure 1.2 Sensors and connections in car

The importance of measurement is demonstrated by the number of Nobel Prizes (Fig. 1.3) that have been awarded recently in this field. For example, the '*Accurate measurement of resistance by means of quant Hall effect*' was awarded in 1985, '*Scanning tunnelling microscope*' was awarded in 1986, '*Cesium atomic clock*' was awarded in 1989, and '*MR imaging*' was awarded in 2003.

Figure 1.3 Nobel Prize

1.1.2 Measurement Procedure

According to the definition we discussed in the previous section, measurement is a process or procedure in which the *method of measurement*, the *unit of measurement* and the *uncertainty* are predefined, and the *quantity value* is obtained as the measurement result. Examples of measurement procedures with different methods, units, uncertainties and results are shown in Fig. 1.4. In the four illustrated methods the source voltage, circuit current, load voltage and load resistance are measured.

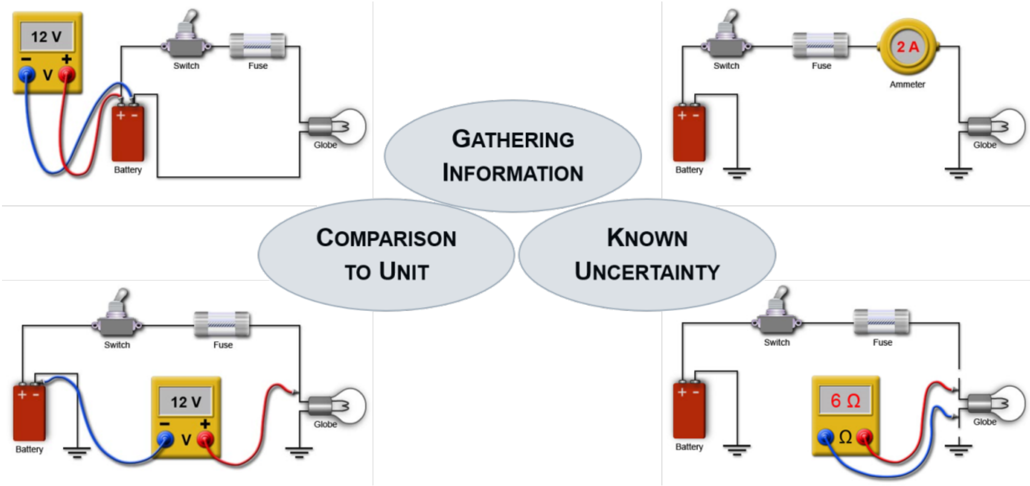


Figure 1.4 Measurement of circuit parameters (example)

The four important elements of measurement procedure are discussed below.

Method of Measurement

The method of measurement is the logical sequence of operations as defined in ISO VIM 2004, the *International Vocabulary of Metrology*. Basically, we can distinguish the analogue and digital measuring method. In case of the *analogue measuring method,* the measured value is proportional to the physical parameter and results are continuous i.e. analogue quantity. This method establishes the time continuous measurement with no information loss in the results.

In the case of the *digital measuring method*, the measured values are quantized and evaluated as discrete time samples. Because of quantized samples there is a well-defined information loss in the result.

Unit of Measurement

In general, the unit of measurement is the elementary scale interval and measurement itself is the comparison to the pre-defined and *standardized* unit. This unit must be universal and reproducible. We have two possibilities to establish a standard unit for reconstruction. On one hand it can be a *material standard,* or on the other hand, it can be a *phenomenon-based standard*. Material standards are usually physically existing *etalons* like the 1 kg platinum iridium alloy cylinder, stored (and guarded) in the Bureau International des Poids et Measures (BIMP) at Sèvres, France. Phenomenon-based standards do not need a physical object for their definition or reproduction as they are referred to the physical phenomena and laws. For example, the definition of electric current is given by Ampére's law of force as in (1.1).

$$F = \mu \cdot \frac{I_1 \cdot I_2 \cdot l}{2\pi \cdot d} \tag{1.1}$$

When $I_1 = I_2 = 1$ ampere constant electric current is flowing in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed $d = 1$ meter apart in a vacuum, it produces between these conductors a force F equal to 2 x 10^{-7} newton per $l = 1$ meter of length. In (1.1), the μ parameter is the magnetic permebility of a vacuum.

Value of Quantity

Value of quantity is a dimensionless number showing how much times higher the physical value of measurement is than the unit of measurement, as given in (1.2).

$$x = x_n \cdot \zeta \tag{1.2}$$

where x is the physical value to be measured, xn is the value of quantity and ζ is the unit of measurement. For example, if the measured electric current as a physical value is 1.25 A, the value of quantity is 1.25 and unit of measurement is A (ampere).

Known Uncertainty

A very important term to discuss is the *uncertainty of measurement*. After publishing the "*ISO Guide to the expression of uncertainty in measurement*" (ISO Guide 1993) the term uncertainty practically substituted the terms *error and accuracy*, which had been used more often in the past. What is the difference between these terms? According to the ISO Vocabulary (ISO VIM 2004), accuracy is the ability of the measuring system to provide a quantity value close to the *true value*, while the uncertainty characterizes the dispersion of the quantity values that is being attributed to the value to be measured.

We always determine the measured value *X* with an error ΔX referred to the X_T true value (or exact value). Thus, the result of measurement should be always presented as in (1.3).

$$X = X_T \pm \Delta X \tag{1.3}$$

The error of measurement can be expressed as the *absolute error* ΔX or as the *relative error h,* defined in (1.4). The relative error (usually given in %) is an absolute error related to the reference value X_R.

$$X = X_T \pm h \cdot X_R \leftarrow h = \frac{\Delta X}{X} \tag{1.4}$$

where, X_R is the reference value, ΔX and *h* are the absolute and relative errors.

It is important to know that the true value cannot be determined, because there is always some error of measurement (although this error can be infinitely small, it always exists).

1.1.3 Process of Measurement Modelling

As measurement is a process for observing physical reality it needs careful preparation and planning.

First of all, the measurement circuit as the *equivalent circuit* of the physical reality has to be planned isolating irrelevant elements and disturbing effects as far as it is possible.

The second step is planning the *measurement procedure* by defining (1) *what* we want to measure, (2) *how* it can be done, and (3) what is the required *uncertainty*.

Finally, the appropriate measurement *method* has to be chosen. The possible measuring methods, like direct comparison with reference to a standard etalon or benchmark, indirect comparison, when no etalon is required but calibration has to be applied, or differential method as comparison to a reference etalon, which will be discussed in detail later on.

1.1.4 Measurement Architectures

Traditional Measuring System (TMS)

The Traditional Measuring System, illustrated in Fig 1.5, is used when we have to measure several different parameters, usually when the unique case and measurement cannot be repeated several times. We use a number of devices and sensors placed near the tested object. Test results are followed by local evaluation and there is no need to transfer the measuring values to the distance, thus no output interface is built into the measuring system.

Figure 1.5 Traditional measuring system

Because of the unique case, the environment of the test often looks messy or a bit confused due to the numerous cables and evaluation tools, etc (as shown in Fig. 1.6).

Figure 1.6 TMS evaluated measurement

Computer Measuring System (CMS)

In CMS testing, smart devices and sensors are used instead of standalone measuring units. The measuring environment is set up for specific professional testing. The main feature of this model is the ability for remote processing and automated evaluation of test results. Smart sensors or local measuring elements are equipped with standard output interfaces, like USB or Ethernet ports, RS232 interfaces, etc. The principle of the CMS test environment is illustrated in the example in Fig. 1.7.

For convenience, most modern measuring elements, for CMS technology, are compatible with IEEE P1451 standard. These elements support 'plug and play' sensory technology and in some cases can be connected directly to the Internet.

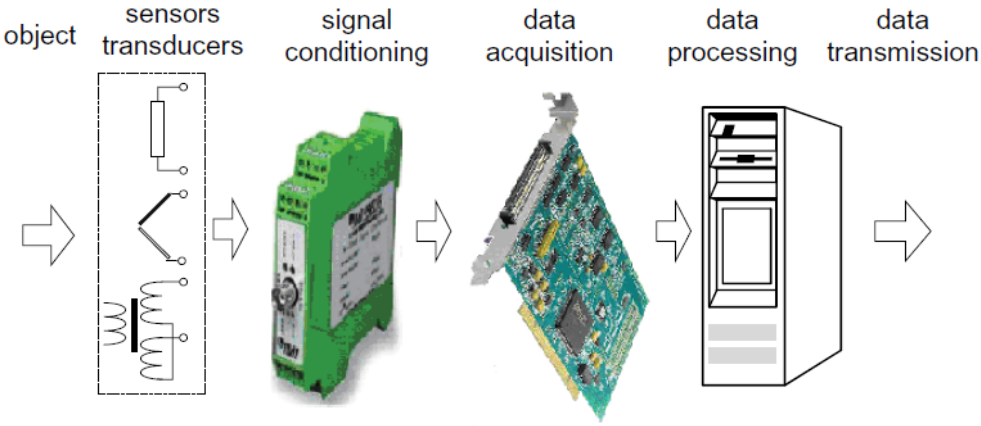


Figure 1.7 CMS testing architecture

Virtual Measuring System (VMS)

The VMS architecture provides a quick and easy solution for testing and does without the need for a large number of separate measuring devices. Test devices, such as a digital multimeter (DMM), oscilloscope, signal generator, etc., are emulated by software with a user friend interface displaying on a computer screen. Measuring terminals are simple plugins connected to the computer through USB, FireWire, etc. There is a wide range of possible evaluating applications like LabVIEW, TINA, etc. A simple VMS solution is illustrated in Fig. 1.8.

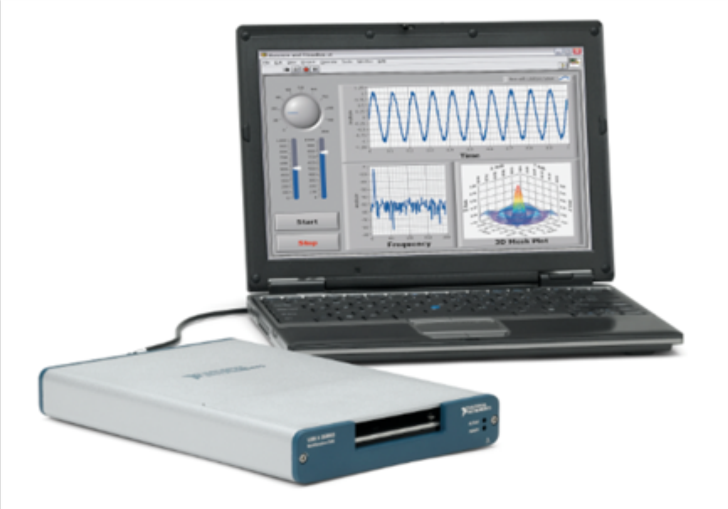


Figure 1.8 VMS architecture

1.2 Methods of Measurement

1.2.1 Direct and Indirect Methods

One of the simplest measurement methods is the direct comparison of a measured value with the standard one. An example of the *direct measurement method* is presented in Fig. 1.9a. In this method the idea of weighting is applied (this method is sometimes called the *current weight*). One coil of the electromagnet is supplied with the measured current I_X, which causes the ferromagnetic element on one arm of the balance to attract. (It is possible to also use other electric mechanisms, for example attracting of the magnet or attracting another coil – this last mechanism is very close to the definition of the ampere.)

On the second arm of the balance a similar mechanism is placed – this time the coil is supplied by the standard reference current I_S. Changing the value of the standard current we can balance the weight – the equilibrium state is when the pointer is at the zero position. We can also determine the state of equilibrium using the electrical method – for example by measuring the resistance R_X of the resistor with a moving slider causing a change in resistance (the potentiometer). This idea is presented in Fig. 1.9b.

Fig. 1.9b presents the *indirect method of measurement* of the value of the electric current. This time, the measured current causes the change of resistance R_X. In this circuit there is a lack of the standard of current, but this does not mean that this standard does not exist. It exists as the scale of resistance and could be introduced to this method by first supplying the coil by the reference standard current (in such way we introduced the dependence $R_X=f(I)$ to this method).

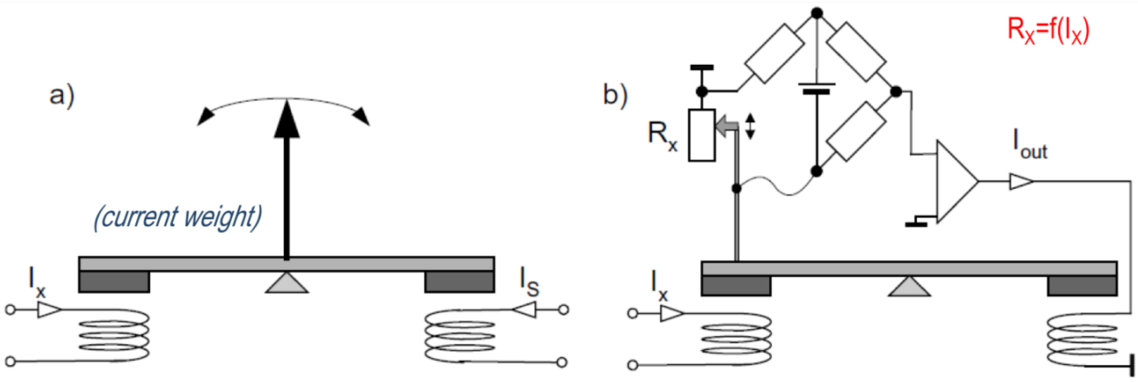


Figure 1.9 the current weight, a) direct measurement, b) indirect measurement

Figure 1.9b presents the weight that is weighted automatically. This method applies the idea of *feedback*. The resistor R_X (sensor of the position or displacement) is connected to the bridge circuit consisting of four resistors and the supply voltage. When all resistors are the same then the output signal (connected to the amplifier) is equal to zero. When one of the resistances is changed then the equilibrium is disrupted, and the signal of unbalance appears at the output of the amplifier (see the *bridge circuit* later on in this chapter).

The process of current measurement is as follows. If current $I_X = 0$ then the bridge circuit is in equilibrium state and the weight is balanced. If the current I_X changes then the electromagnet (the coil) attracts the ferromagnetic element on the arm and the resistance R_X changes. This causes the signal voltage to appear at the output of the bridge circuit. This signal after amplification supplies the second coil as the current I_{out}, which causes the movement of the second arm and balancing of the weight (similarly as it was performed manually in the example presented in Fig. 1.9a). After a short period of time (the transient state) the balance returns to the equilibrium state, which is detected as zero voltage on the output of the bridge circuit. Thus, by means of feedback we realize the automatic balancing of the weight and the output current I_{out} can be the measure of the tested current I_X.

1.2.2 Null and Deflection Methods

The classical example for the *null measurement method* is the mass measurement as shown in Fig. 1.10. In the case of a balanced (equilibrium) state of the Libra the amount of standard mass has to be equal to the unknown mass i.e. the mass to be measured. The mass (or weight) measurement is performed easily by changing the amount of standard masses on the right arm of the Libra.

Current measurement, shown in Fig. 1.9, can also be performed by the null measurement method. We balance the circuit and the balance state is indicated by the pointer or electrically by the zero-output voltage of the bridge circuit.

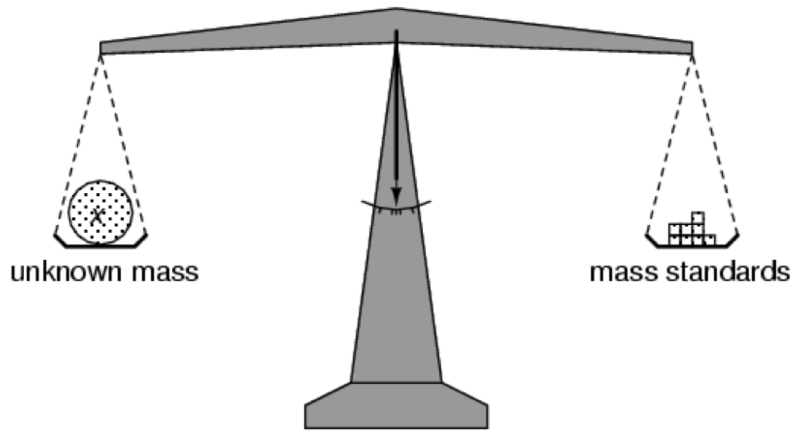


Figure 1.10 Classical example for null measurement method with Libra

The same measurement can be performed by the *differential measurement method*. In this case it is not necessary to balance the circuit – the deflection of the pointer or the output voltage of the bridge circuit can be used as the measure of the current value. It is also possible to use the *null-differential measurement method*. In this method we roughly balance the weight by the current I_0 and the deflection of the pointer (or change in the output voltage of the bridge circuit) is caused by the difference between equilibrium state and state after the change of the current $I=I_X-I_0$. Using the null-differential method we can obtain an improvement in the sensitivity of the measurement – the movement of the pointer can be realized by the smaller current I instead of the current equal to I_X.

The current measurement, shown in Fig. 1.11, applies the *indirect differential method.*

The method is indirect because the electric current is measured by a comparison to the standard mass or weight.

The deflection angle of the pointer is proportional to the measured current as given in (1.5). This function is construction dependent and is near to linear in an ideal case.

$$\alpha = f(I_X) \tag{1.5}$$

In any case, scaling and calibration are necessary in this measuring method.

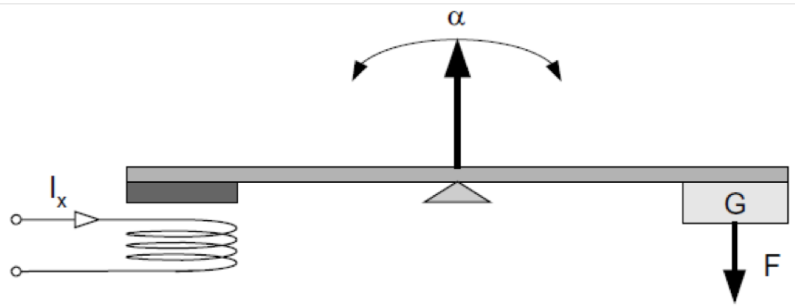


Figure 1.11 Null- differential measurement method

1.2.3 Compensation Methods

The important method of comparison of the measured value and the standard one is the *compensation measurement method* illustrated in Fig. 1.12a. In the compensation method we determine the difference between two values (measured and standard) and we can precisely determine the state when this difference is equal to zero. Thus, the compensation state (the equilibrium state) means that both values are the same and cancel each other. This results in the state of equilibrium for the signal at the output of the circuit being zero.

In the example presented in Fig. 1.12a the measured voltage U_X is compensated by the voltage drop U_S on the resistor R_S. The state of equilibrium is indicated, where the output signal is zero, by the *null indicator NI* (for this purpose a very sensitive voltmeter called a *galvanometer* can be used).

The measuring procedure consists of two steps. In the first step, the standard value of the current I_S is established (for example by comparing it with the standard value). In the second step, the resistance R_S is adjusted until the null indicator *NI* indicates a state of equilibrium. For a known value of current I_S the resistor R_S can be scaled directly in voltage units.

Besides the advantage of accurate measurements, another benefit is the non-invasive way of testing i.e. no energy consumption from the tested source. In other words, the measuring device has infinite input impedance so has no influence on the measured environment.

Figure 1.12b presents the realization of the compensation idea performed automatically by applying the feedback technique. The amplifier works as the null indicator amplifying the difference between the measured voltage U_X and the drop in voltage U_S. This difference causes the output current I_{out}, which is increased until the input voltage of the amplifier again returns to zero.

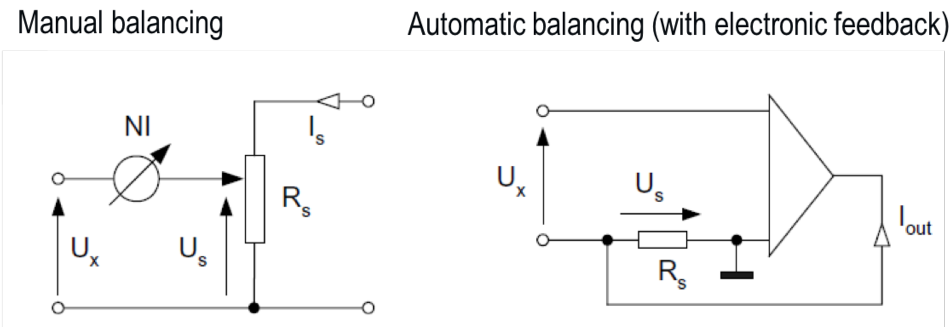


Figure 1.12 Principle of compensation in a) manual and b) automatic balancing

1.2.4 Comparative Method

In the circuit presented in Fig. 1.13 we can obtain the equilibrium by the compensation of the currents I_X and I_S. See the equation (1.6).

$$I_X - I_S = 0 \tag{1.6}$$

This state of equilibrium can be realized by the change of the voltage U_1 or U_2. The condition of the equilibrium is given in (1.7).

$$\frac{R_x}{R_S} = \frac{U_1}{U_2} \tag{1.7}$$

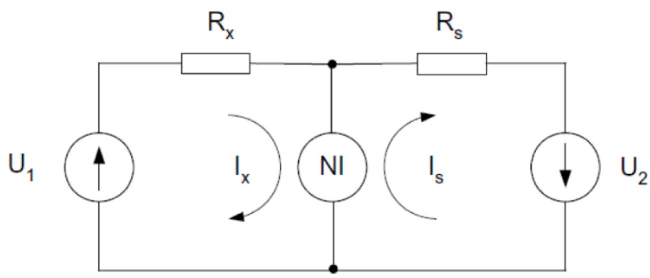


Figure 1.13 the comparator of two resistances

Thus, we can use this circuit for resistance measurement. The current compensation circuit was used as the resistance comparator.

1.2.5 Bridge Circuits

Figure 1.14 presents the *bridge circuit*, which is very often used for the measurements of the resistance (or impedance).

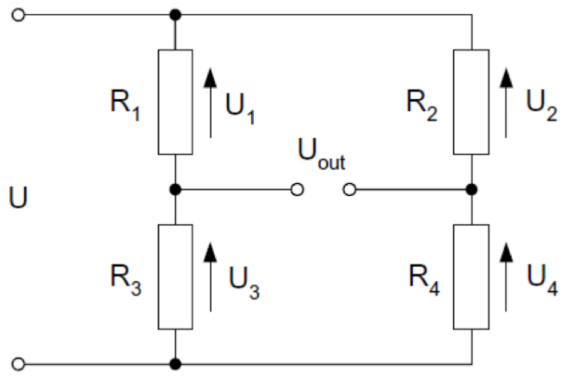


Figure 1.14 the bridge circuit

The voltage drops across the resistors $R1$, $R2$, $R3$ and $R4$ are given in (1.8), (1.9), (1.10) and (1.11).

$$U_1 = U\frac{R_1}{R_1 + R_3} \tag{1.8}$$

$$U_2 = U\frac{R_2}{R_2 + R_4} \tag{1.9}$$

$$U_3 = U\frac{R_3}{R_1 + R_3} \tag{1.10}$$

$$U_4 = U\frac{R_4}{R_2 + R_4}; \tag{1.11}$$

To obtain the equilibrium (U_{out} = 0) the conditions: $U_1 = U_2$ and $U_3 = U_4$ should be fulfilled. From the equations (1.8), (1.9), (1.10) and (1.11) we obtain following condition of the equilibrium of the bridge circuit

$$R_1 R_4 = R_2 R_3 \tag{1.12}$$

The condition shown in 1.12 is a general condition of the balance of the bridge circuit: the products of the resistance of opposite arms of the bridge circuit should be equal.

A deflection type bridge circuit for resistance measurement is shown in Fig. 1.15.

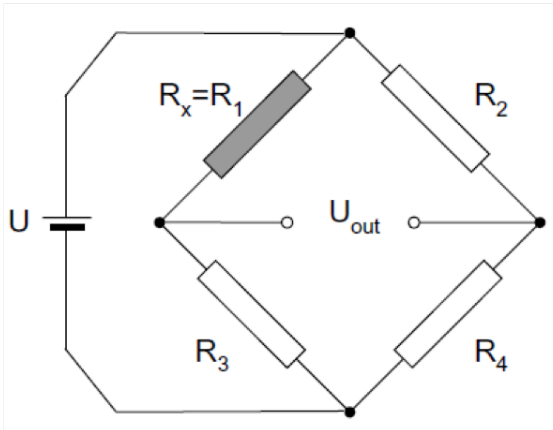


Figure 1.15 Deflection type bridge circuit

By applying the condition (1.12) we can determine the measured resistance $R_X = R_1$ from the equation (1.13).

$$R_x = R_2 \frac{R_3}{R_4} \tag{1.13}$$

In practice there are two kinds of bridge circuits: *null-type bridge circuits* (or *balanced bridge circuits*) and *deflection-type bridge circuits* (or *unbalanced bridge circuits)*. In the null-type bridge a null indicator is connected to the output and the bridge is balanced – for example by the change of one of the resistors. Most often the resistor R_2 is used for balance, while change of the ratio R_3/R_4 is used for the range selection.

The null-type bridge is not often used as a measuring device nowadays, instead, the deflection-type bridge circuit is commonly used as the conditioning circuit enabling the conversion of a change in resistance (or generally impedance) of the sensor into the voltage signal. This type of the bridge circuit is first balanced and next the output voltage (voltage of unbalance) is used as the output signal of the *U = f(R) (resistance-voltage transducer).*

Figure 1.16 presents the *transfer characteristic* $U_{out} = f(\Delta R_X/R_X)$ of such a transducer. We can see that this characteristic is *nonlinear.*

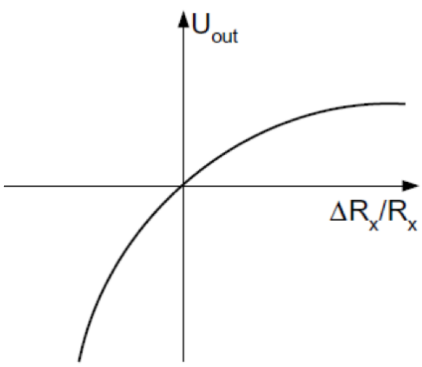


Figure 1.16 Transfer characteristic of the resistance-voltage transducer

There are various methods for the linearization of the resistance-voltage conversion. One of these methods of linearization is presented in Fig. 1.17 showing the bridge circuit with automatic balancing by means of a feedback circuit.

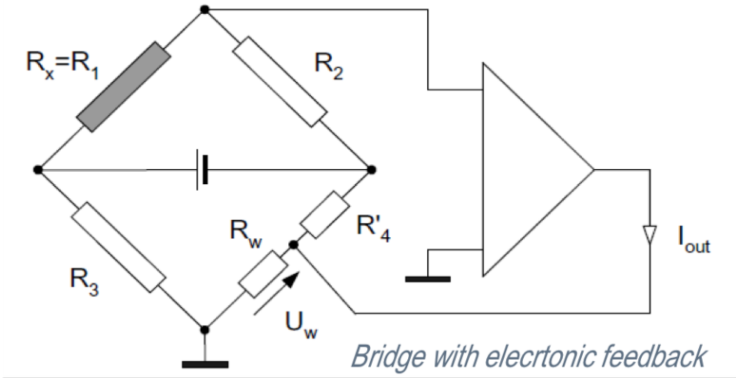


Figure 1.17 Linearization of the resistance-voltage conversion

We can see from Fig. 1.14 that the bridge circuit is balanced by changing one of the resistances and thus changing the voltage drop across it. Therefore, the change of resistance causes the same effect as change of the voltage on one of the arms. In the circuit presented in Fig. 1.17 the output signal of the bridge circuit after amplification causes the output current I_{out} to create an additional voltage drop U_W across the resistance R_W. This additional voltage can drive the balancing of the bridge circuit until it returns again to the balanced state.

The output current is the measure of the investigated resistance R_X. Because the bridge circuit is automatically balanced its output signal is very small – if the amplification is very large only a very small signal (in range of *V*) is required to generate the output current I_{out}. Therefore, only the small, linear part of the whole nonlinear transfer characteristic is used – thus the whole characteristic of the transducer $I_{out} = f(\Delta R_X/R_X)$ is linear.

1.2.6 Substitution Method

Analysing the equation (1.13) we see that the accuracy of the measurement of value R_X depends on the accuracy of all three other resistors. We can improve the accuracy of the measurement by applying the *substitution measurement method*. In that method the measuring procedure consists of two steps. In the first step the bridge circuit is balanced. Next, we substitute the measured resistor R_X by the standard resistor R_S. This time we do not balance the bridge changing the resistances of the bridge circuit, but we balance it by changing the standard resistance. If the time period between these two operations is not long (to avoid potential influence of the change of the temperature or other factors) the accuracy of measurement does not depend on the accuracy of the resistors in the bridge (it depends only on the accuracy of the standard resistor). This way, the bridge circuit is only used as the device testing that nothing changed after the resistors were substituted.

An example of the substitution method is presented in Fig. 1.18. Measuring the value of the alternating current (AC) is rather difficult, especially if this current is not sinusoidal. In contrast, we are able to measure the value of direct current (DC) very accurately. In the circuit presented in Fig. 1.18 the measurement procedure consists of two steps. In the first step we connect to the circuit the measured alternating current I_X. This current causes heating of the resistor (heater) R_T. A thermocouple (temperature sensor) is connected to this heater – in such a sensor the change of the temperature causes the change of the output voltage U_T. In the second step, we connect the standard direct current I_S to the heater. We change this current until the temperature of the heater is the same as in the first step. Because the effect of heating was the same in both cases, the values of both currents are the same. Thus, we substituted the measurement of the alternating current by the measurement of the direct current. The *root mean square* (RMS) value of the AC current is equal to the value of the DC current, and causes the same change in temperature.

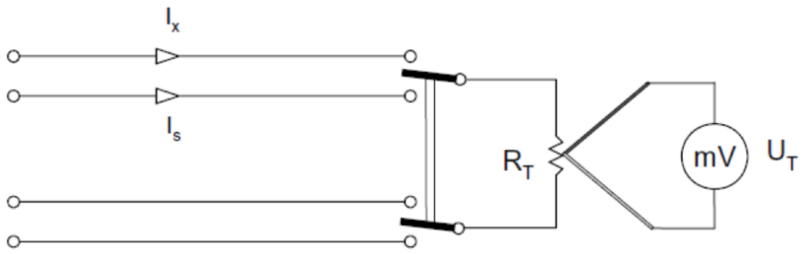


Figure 1.18 Measuring AC RMS value by means of substitution method

1.3 Uncertainty of Measurements

It is important to recognize that it is impossible to determine the true value of the measured quantity, because the measurement is always performed with a degree of uncertainty. Therefore, we can state that the measurement without the estimation of this uncertainty is worthless. Thus, an analysis of uncertainty always accompanies the measurement and it is crucially important.

We can rewrite (1.3) in the following form

$$X_T - \Delta X \leq X \leq X_T + \Delta X \qquad (1.14)$$

which can be read as follows: the result of measurement *X* is determined with the dispersion *X* around the true value X_T (bearing in mind that *X* is an absolute error of measurement).

In 1993, the International Organization of Standardization (ISO), with collaboration of many other prestigious organizations, published a "*Guide to the expression of uncertainty in measurement*". According to the concept presented in the *Guide* the dependence (1.14) should be substituted by the dependence

$$\Pr(X_0 - u \leq X \leq X_0 + u) = 1 - \alpha \qquad (1.15)$$

Thus, the result of measurement *X* is determined with the uncertainty *u* around the estimated value X_0 with the level of confidence $(1 - \alpha)$. Symbol *Pr* in the equation (1.15) denotes the probability.

We can see that the *true value* in (1.14) - which we never know - is now substituted by *the estimated value*. Similarly, the *error* is now substituted by the *uncertainty* because we also do not know the value of that error. Because we do not know the true value, we cannot determine the error using the equation (1.14).

In case of practical applications, it is recommended to limit the assumed uncertainty of measurement to a certain useful level. An increase in accuracy corresponds to higher costs as we have to use more expensive measuring devices. (The application of overly precise instruments can also be inconvenient. For example, when we use a five-digit instrument for the measurement of a non-stable source, the last digits keep fluctuating and are therefore useless.)

Fig. 1.19 illustrates the relation between cost and uncertainty. We can reduce the costs of the measuring procedure by increasing the uncertainty, but there is a risk that this will increase the total costs as a result of an incorrect decision. Taking this into consideration we can establish the optimal value of uncertainty.

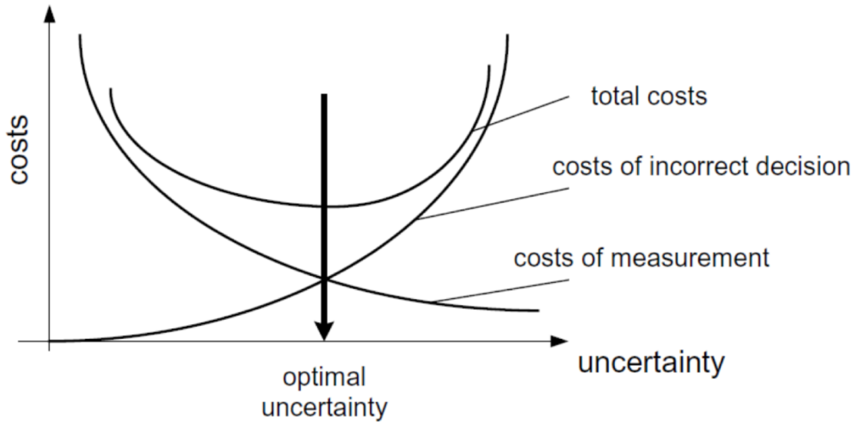


Figure 1.19 Relation between the cost and uncertainty

Note that we substituted *error* with *uncertainty* because the accurate value of error is never known in measurements. Despite this, we also use error in short for describing the upper limit of uncertainty in the following.

1.3.1 Reasons for Uncertainty

The reasons for uncertainty of measurements can be classified as presented in Fig. 1.20.

In general, *modelling errors* are the result of insufficient preparation and measurement planning. *Device errors* occur when incorrect scaling is applied or unsatisfactory accuracy of the applied measuring device and *transmission errors* include the uncertainties caused by the signal transmission channel according to its noisy conditions for example.

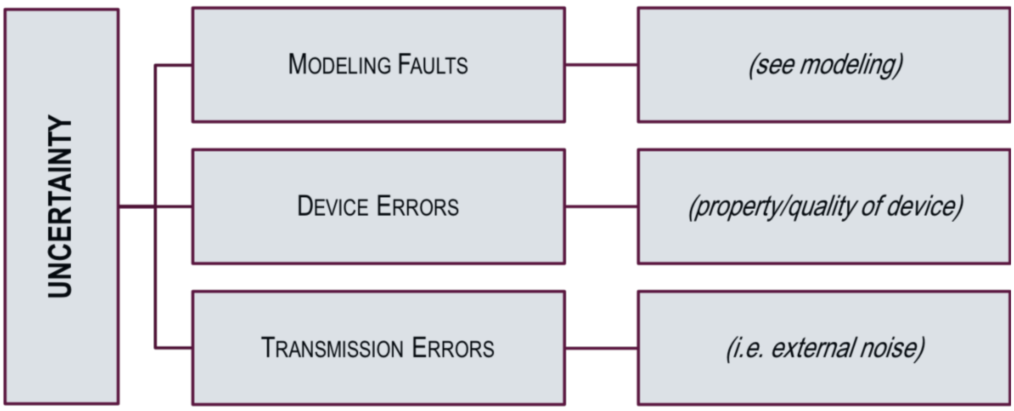


Figure 1.20 Main reasons for uncertainty

Because of their practical importance, some additional device errors and factors of uncertainty have to be mentioned in addition to the accuracy and scaling errors. Fig. 1.21 presents the results of four very common device errors showing the relationship between the true value (x) and measured value (y).

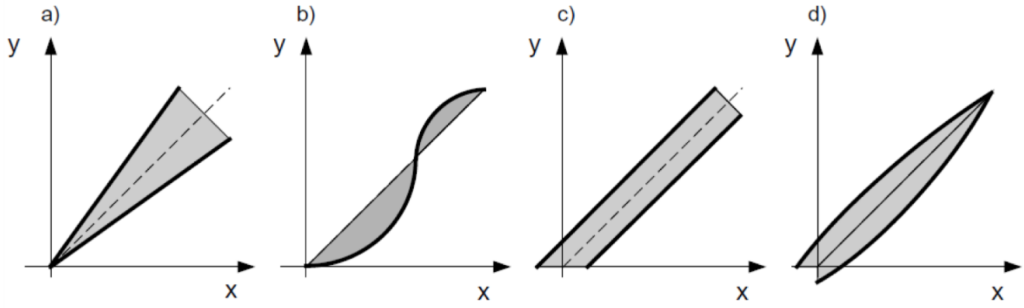


Figure 1.21 Common device error factors

Fig 1.21a demonstrates the effect of uncertainty (error) in device sensitivity. In Fig 1.21) we see the error of linearity, Fig. 1.21c is for error of resolution and Fig. 1.21d shows the hysteresis error of the measuring device.

1.3.2 Classification of Measuring Errors

Regarding time dependent behaviour, we have to distinguish between *static errors* and *dynamic errors*. While static errors are time independent, dynamic errors are caused by certain transients. The classification of measuring errors is presented in Fig. 1.22. The first main group within static errors is the *gross error*. This rough error is caused by an environmental error or human error during the measurement process. *Systematic error*, often called *absolute error* has a fixed value (K) added to (or subtracted from) the measured value thus, it can be corrected during evaluation. The relation between the X measured value, K absolute error and X_0 estimated value is given in (1.16).

$$X = X_0 + K \qquad (1.16)$$

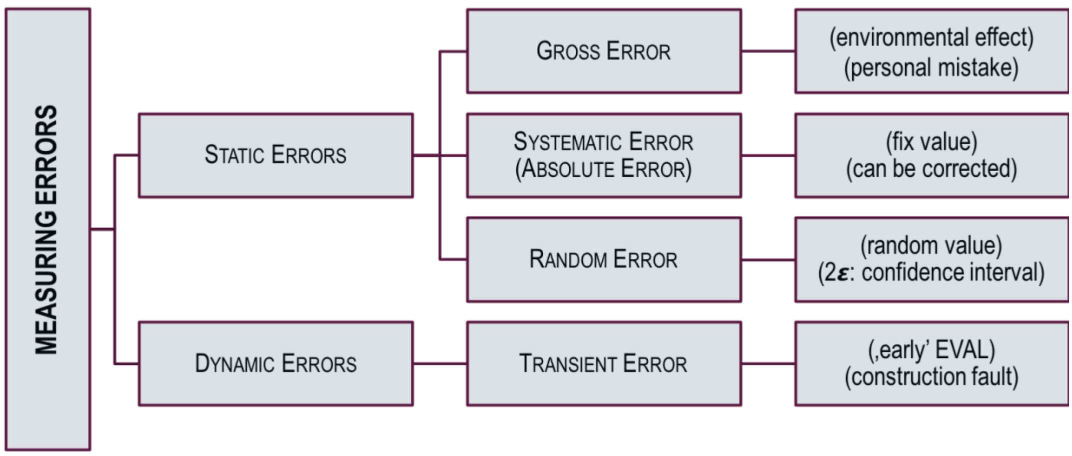


Figure 1.22 Classification of measuring errors

Random errors have no fixed values or even no 'rule' in their time dependent behaviour. We can state that the measured value is in a predefined (ε) confidence interval of the estimated value as described in (1.17).

$$X = X_0 + K \pm \varepsilon = X_{corr} \pm \varepsilon \tag{1.17}$$

Or, it can be interpreted as in (1.18). Probability of that measured value is in a confidence interval of the estimated value that is not less than the predefined confidence probability (P_ε).

$$Pr(|X_0 - X| \leq \varepsilon) \geq P_\varepsilon \tag{1.18}$$

1.3.3 Measurement Series

A measurement series is given by measuring the same parameter several times using the same device under unchanged conditions. When performing a series of measurements, we obtain a certain number of results to be statistically evaluated. An example of a measuring series is presented in Fig. 1.23.

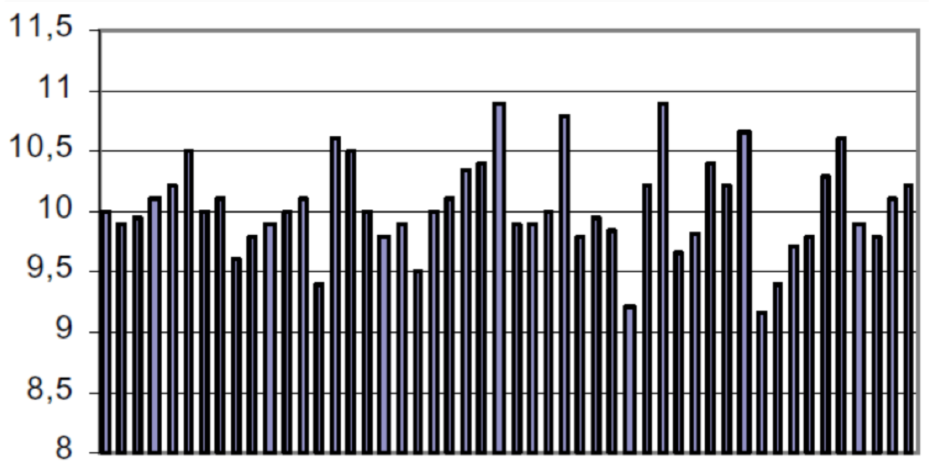


Figure 1.23 Series of measurements

A simple and useful evaluation method is drawing a histogram as shown in Fig. 1.24. In a histogram, the number of certain values are 'counted' so it is easy to see the most common value of the series. The estimated value in the example in Fig. 1.24 is around 9.9.

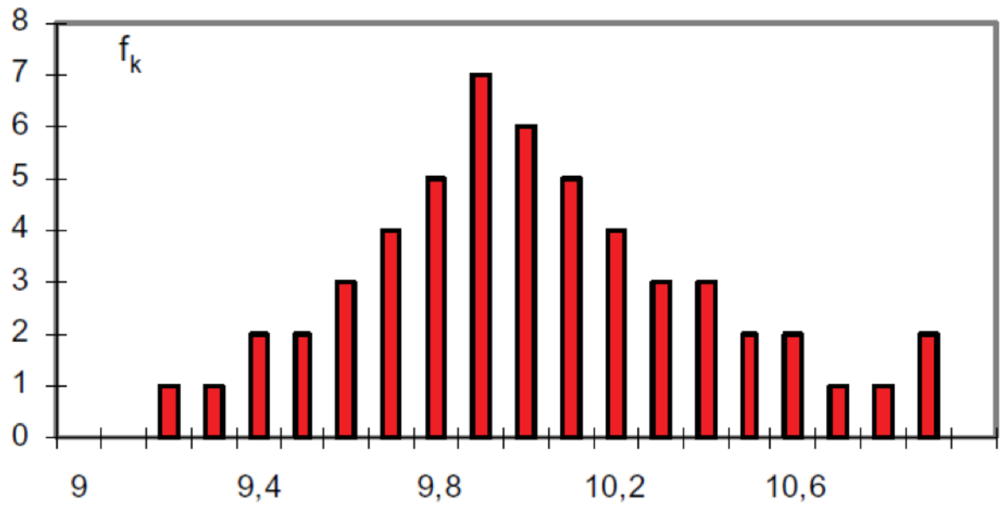


Figure 1.24 Histogram of the measuring series given in Fig. 1.24

The histogram is a useful tool in measurement evaluation, even just to check for homogeneity. As an example Fig. 1.25 demonstrates histograms of magnetic field scanning within two magnetic materials describing their homogeneity. The red 'area' is wider than the green one, which leads us to the conclusion that the measurements in the green histogram show better homogeneity in the magnetic field within the scanned material.

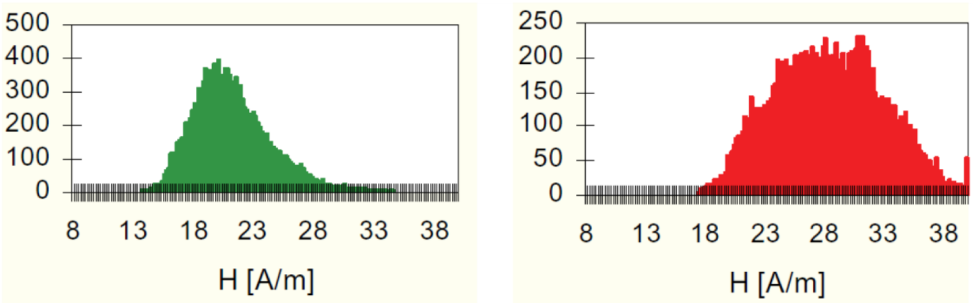


Figure 1.25 Histograms of scanned magnetic fields

Besides the histogram method, we can evaluate the measurement series by also calculating the following statistical parameters.

Mean value is equal to the expected value that is the best approximation of true value. It is calculated as in (1.19).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1.19}$$

The *range of measurement* is the difference between the maximum and minimum values in a series, defined in (1.20)

$$R = x_{\max} - x_{\min} \tag{1.20}$$

Mean absolute deviation is the deviation within the range, as shown in (1.21). So, it is the mean value of the absolute deviations from the mean value.

$$E = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}| \tag{1.21}$$

Standard deviation is the best undistorted empiric estimation with weighted calculation of the greater deviations where the degree of freedom is *(n-1)* instead of *n* as the denominator. (Calculation is weighted because of the applied square function in sum.)

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{1.22}$$

Relative deviation is the standard deviation relative to the mean value for easier comparison as defined in (1.23)

$$S_v = \frac{S}{\bar{x}}. \tag{1.23}$$

And *variance* is the square of deviation

$$V = S^2. \tag{1.24}$$

The estimated value can be calculated by the statistical parameters as given in (1.25).

$$X_0 = \bar{X} \pm \varepsilon, \;\; \varepsilon = t \cdot S, \tag{1.25}$$

where t is an integer giving the probability interval for Normal (Gaussian) distribution. For example, if t=1 then 68.3%, if t=2 then 95.5%, if t=3 then 99.7% of the values are within the interval.

Calculation Example

Evaluate the random erroneous measuring series when the voltage measurement series is given as the following. *V* = [100, 102, 102, 104, 103, 99, 103, 102, 101, 104, 101, 103, 102] V (*n* = 13)

Solution

- Mean value of the series is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{13} \cdot (99 + 100 + 2 \cdot 101 + 4 \cdot 102 + 3 \cdot 103 + 2 \cdot 104) = 102\ V$$

  Range

$$R = x_{\max} - x_{\min} = 104 - 99 = 5\ V$$

- Mean absolute deviation

$$E = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}| = \frac{1}{13} \cdot (2 + 2 + 1 + 3 + 1 + 1 + 2 + 1 + 1) = 1.077\ V$$

- Standard deviation

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{\frac{26}{12}} = 1.47\ V$$

- Relative deviation

$$S_v = \frac{S}{\bar{x}} = \frac{1.47}{102} = 1.44\ \%$$

1.3.4 Accuracy Class

We have already defined absolute error and relative error of the measured value in (1.3) and (1.4). *Absolute error* is the difference between the true (or estimated) value and measured values. It is a specified value often corresponding to the full-scale range of the measuring device. *Relative error* is the ratio of absolute error and the true (or estimated) value.

Measuring instruments are classified by their accuracy with the parameter of *accuracy class*, written on devices. For example, the voltage meter, shown in Fig. 1.26, has an accuracy class of 1.5 %.

Figure 1.25 Voltage meter with accuracy class of 1.5 %

Accuracy class (h_{op}) is the relative error compared to the full-scale of measuring range as in (1.26)

$$h_{op} = \frac{H}{X_R} \qquad (1.26)$$

where H is the absolute error and X_R is the full-scale range of the device.

According to (1.26) we can define the *absolute error upper limit* as in (1.27). Note that the upper limit is independent of the measured value.

$$H_{max} = h_{po} \cdot X_R \qquad (1.27)$$

As expected, the *Relative error upper limit* as defined in (1.28) depends on the measured value. The lower the measured value, the higher the relative error at the device with the given absolute error.

$$h_{max} = \frac{H_{max}}{X} = h_{po} \cdot \frac{X_R}{X} \qquad (1.28)$$

In order to minimize measuring errors, the upper third part of the measuring scale should be used.

A rule of thumb for checking the *error upper limits of calculated values* is the following:If addition and subtraction are used in the calculation the absolute error upper limits of measured values add up. If multiplication and division are applied in calculation the relative error upper limits of measured values add up.

Calculation Example

An electrical current of 6 A is measured with an ammeter. The full-scale measuring range is 10 A and the accuracy class of the device is 1 %. We have to determine the absolute error, relative error and range of accuracy of the measurement.

Solution

The absolute error of the measurement is

$$H = \frac{h_{op}}{100} I_R = 100 \text{ mA}$$

The relative error is

$$h = \frac{H}{I} \cdot 100 = 1.67\%$$

The range of accuracy is

$$I - H \leq I_0 \leq I + H \rightarrow 5.9 \text{ A} \leq I_0 \leq 6.1 \text{ A}$$

1.3.5 Principles in Propagation of Errors

In statistics, the propagation of errors is the effect of different variables' errors on the uncertainty of a function that are based on them. When the variables are the values of measurements, they have uncertainties due to measurement limitations (e.g. instrument precision) which propagate due to the combination of variables in the function i.e. the mathematical operations performed on test results.

Most commonly, the uncertainty of a quantity is quantified in terms of the standard deviation (*S*), the positive square root of variance (*V*). The value of a quantity and its error are then expressed as an interval $x \pm u$. If the statistical probability distribution of the variable is known or can be assumed, it is possible to derive confidence limits to describe the region within which the true value of the variable may be found.

If the uncertainties are correlated, then covariance must be considered. Correlation can arise from two different sources. First, the *measurement errors* may be correlated. Second, when the underlying values are correlated across a population, the *uncertainties in the group averages* will be correlated.

Propagation of Errors in Case of Systematic Errors

Let's have a look at the mathematical function in (1.29) for calculating *y* evaluation result based in x_i measured variables.

$$y = f(x_1, x_2, \dots, x_n) \tag{1.29}$$

In this case the propagated error is obtained from (1.30)

$$\Delta y = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \Delta x_i \tag{1.30}$$

where Δx_i is the systematic (absolute) error of x_i variable, Δy is the systematic error of y calculated result, and $\partial f / \partial x_i$ is the weight coefficient by x_i variable.

Calculation Example 1

The calculated result (*y*) is the linear combination of measured variables (x_i) as follows.

$$y = \sum_{i=1}^{n} a_i x_i$$

The relative error of *y* has to be found.

Solution

According to (1.30) the absolute error is

$$\Delta y = \sum_{i=1}^{n} a_i \Delta x_i$$

The relative error is

$$h_y = \frac{\Delta y}{y} = \sum_{i=1}^{n} a_i \frac{x_i \cdot \Delta x_i}{y \cdot x_i} = \sum_{i=1}^{n} a_i \frac{x_i}{y} h_{x_i}$$

Calculation Example 2

The result (*y*) is calculated with multiplications as given below.

$$y = \prod_{i=1}^{n} {x_i}^{\alpha_i}$$

To simplify things, let's use n = 2 in this example.

$$y = {x_1}^{\alpha_1} \cdot {x_2}^{\alpha_2}$$

The absolute error from (1.30) is

$$\Delta y = \alpha_1 {x_1}^{\alpha_1 - 1} {x_2}^{\alpha_2} \Delta x_1 + \alpha_2 {x_2}^{\alpha_2 - 1} {x_1}^{\alpha_1} \Delta x_2$$

And the relative error is

$$\frac{\Delta y}{y} = \alpha_1 \frac{\Delta x_1}{x_1} + \alpha_2 \frac{\Delta x_2}{x_2}$$

In general, we can write the following.

$$h_y = \frac{\Delta y}{y} = \sum_{i=1}^{n} \alpha_i \frac{\Delta x_i}{x_i} = \sum_{i=1}^{n} \alpha_i h_{x_i}$$

(In error propagation the component relative errors are weighted by their exponents.)

Propagation of Errors in the Case of Random Errors

The calculation of *y*, based in x_i measured variables, is the same as given in (1.29) with the 'only' difference that x_i and *y* are random variables.

Mean values and variances (squares of deviations) of the components are derived from (1.19), (1.22) and (1.24). Thus,

$$\overline{x_i} = \frac{1}{N} \sum_{k=1}^{N} x_{i_k} \tag{1.31}$$

$$S_i^2 = \frac{1}{N-1} \sum_{k=1}^{N} \left(x_{i_k} - \overline{x_i}\right)^2 \tag{1.32}$$

In the case of independent x_i variables

$$S_y^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 S_i^2 \tag{1.33}$$

The equivalent variance is the weighted sum of elements' variances and the confidence interval (random error parameter) is the variance multiplied by a constant! (see also the section on 'measurement series')

Note that in some cases the mathematical 'sum of modulus' is calculated in error propagation as the *upper limit of (relative) uncertainty*, as it is given in (1.34).

$$\delta_{(y)}^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 \left(\frac{x_i}{y}\right)^2 \delta_{x_i}^2 \tag{1.34}$$

1.4 Standards of Quantities

1.4.1 Basic Terms and Concepts

The measurement is always related to the standard unit. The *standard* is a given quantity with a stated value and measurement uncertainty, used as a reference. Using the standard, we can accurately calibrate measuring instruments. The *calibration* is the operation establishing the relationship between quantity values provided by measurement standards and the corresponding indications of the measuring system, carried out under specified conditions and including the determination of measurement uncertainty. It is defined in the ISO International Vocabulary of Metrology – Basic and General Concepts and Associated Terms published by *'Bureau International des Poids et Measures'* (BIPM) - the intergovernmental organization through which Member States act together on matters related to measurement science and measurement standards.

The first standard of *meter* unit was created in 1793 in the form of a brass bar. In 1799 this provisional bar was substituted by a more accurate platinum bar, and in 1889 the standard was substituted with an X-shaped bar made from iridium-platinum. This standard is preserved in the Bureau International Poids at Mesures BIPM in Sèvres (Paris). A copy of the "provisional" meter used from 1796-1797 is located in the wall of 36 rue de Vaugirard, Paris (Fig. 1.26).

Figure 1.26 Public copy of the first standard on a wall (1796-1797)

Standards are used as a reference with their given value and given uncertainty. Regarding the hierarchy of standards, we can distinguish *international standards, national standards, primary standards, secondary standards,* and *working standards*. Calibration, also defined by ISO VIM, is an operation under specific conditions and uncertainty establishing the relationship between standards and applied indication system i.e. measuring device. A measuring device has to be regularly *validated* by comparing it to a so called *etalon* that is calibrated by accurate standards. *Validation* is a certified comparison process.

Two types of standards are *material standards* and *standards, which signify physical phenomena*. Due to the advantages of better and easier reproduction, transferability and accuracy, standards which correspond to the physical phenomena are preferred. Replacement of material standards by standards which refer to physical phenomena are on-going. For example, the iridium-platinum meter as the standard of length has been replaced by the physical phenomena of the path of light travelled in defined time interval.

1.4.2 International System of Units (SI)

The *International System of Units* (SI), abbreviated from the French 'Système international d'unités' is the most widely used system of measurement throughout the world. It comprises of a *coherent system* of units of measurement built on several base units, a set of decimal prefixes to the unit names, and unit symbols that may be used when specifying multiples and fractions of the units. The system also specifies names for a certain number of derived units for other common physical quantities like lumen, watt, etc.

Base units are derived from those constants of nature, such as the speed of light and the triple point of water, which can be observed and measured with great accuracy, but also includes one physical object. This is the international prototype kilogram, certified in 1889, and consisting of a cylinder of platinum-iridium. *Derived units* may be defined either in terms of base units or other derived units.

The International System of Units has 7 base units, in which three are mechanical, one is electrical, one is thermal, one is photometric, and the last one is chemical. Base units are represented with their measuring units in Fig. 1.27 where 'm', 'kg' and 's' measure mechanical quantities (length, mass, time) and 'K', 'A', 'mole' measure thermal, electrical and chemical units. The last, 'cd', measures photometric units. There are two supplementary units to the base units for measuring plane angle and solid angle – 'rad' and 'sr'.

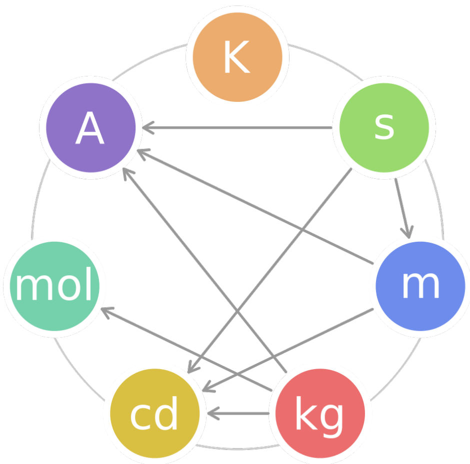


Figure 1.27 The seven base units of SI

The definitions of the seven base units are:

- *Unit of length,* symbol: *l,* unit: m, - meter is the length of the path travelled by light in a vacuum during the time interval of 1/299 792 458 of a second. (Replaced the prototype of platinum-iridium in 1889.)
- *Unit of mass,* symbol: *m*, unit: kg, - kilogram is the Hz unit of mass; it is equal to the mass of the international prototype of the kilogram. (Sévres/FR, 1889, the only prototype);
- *Unit of time,* symbol: *t*, unit: s, - second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom.;
- *Unit of electric current,* symbol: *I*, unit: A, - ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2 x 10^{-7} N per meter of length.
- *Unit of thermodynamic temperature,* symbol: *T*, unit: K, - kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.
- *Unit of amount of substance,* symbol: *M*, unit: mole, - mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12. (The elementary entities may be atoms, molecules, ions, electrons)
- *Unit of luminous intensity,* symbol: I_v, unit: cd, - candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540 10^{12} Hz and that has a radiant intensity in that direction of 1/683 watt per steradian

The definitions two supplementary units are the followings.

- *Plane angle,* symbol: $\alpha, \beta, \ldots$, unit: rad, - central angle belonging to the length of arc of the circle equal to the radius.;
- *Solid angle,* symbol: $\alpha, \beta, \ldots$, unit: sr, - central angle belonging to the area of spherical part equal to the square of radius of the sphere.

According to the rule of coherence, a derived unit is mutually related by the rules of multiplication and division so that no other numerical factor than one is needed.

There are a number of derived units in the measuring system. For example, Table 1.1 shows some of the important derived units, which are used when measuring electricity, with their symbols and measuring units.

Table 1.1 Some of derived units in electricity

Derived Unit	Symbol	Measuring Unit
Frequency	*f*	Hz, 1/s
Angular frequency	*ω*	rad/s
Force	*F*	N, kgm/s^2
Pressure	*p*	Pa, N/m^2
Energy, work	*W, E*	J, Nm
Power	*P*	W, J/s
Charge	*Q, q*	C, As
Voltage	*U, u, V, v*	V, Nm/As
Resistance	*R*	Ω, V/A
Conductance	*G*	S, A/V
Capacitance	*C*	F, C/V
Inductance	*L*	H, Vs/A
Electric field intensity	*E*	V/m
Magnetic field intensity	*H*	A/m
Magnetic flux	*Φ*	Vs
Flux density (Induction)	*B*	T, Vs/m^2

The system of units defines *metric prefixes* for convenient number handling. Metric prefix is a unit prefix, that precedes a basic unit of measure, to indicate a multiple or fraction of the unit. While all metric prefixes in common use today are decadic, historically there were a number of binary metric prefixes as well. Each prefix has a unique symbol that is attached to the beginning of the unit symbol.

Decimal multiplicative prefixes have been a feature of all forms of the metric system, with six dating back to the system's introduction in the 1790's. Metric prefixes have even been prepended by non-metric units. The SI prefixes are standardized and used to form decimal multiples and submultiples of SI units. They should be used to avoid very large or very small numeric values. The prefix attaches directly to the name of a unit, and a prefix symbol attaches directly to the symbol for a unit.

The most important advantages of the application of the system of units are the followings.

- *Coherence*, no numerical factor (other than one) is used in calculation of derived units. Although, there are some exceptions in practical life, like the consumption of electrical energy is still measured in Wh (kWh) instead of the standard unit of energy (J = Ws). The numerical factor between Wh and Ws (J) is 3600, according to the numerical factor between seconds and an hour.;

- *Simplicity*, because of the easy (short) equations for calculating derived units;
- *Comparability and universality*, the same unit system is applied throughout all the fields of disciplines.;
- *Continuity*, as the main units of previous systems are still maintained in the system of units;
- *Consistency*, because no conflicts in naming and interpretation between different disciplines.

The standardized prefixes with their symbols and multipliers are listed in Table 1.2.

Table 1.2 Metric system prefixes

Prefix	Symbol	Multiplier	
exa	E	10^{18}	1,000,000,000,000,000,000
peta	P	10^{15}	1,000,000,000,000,000
tera	T	10^{12}	1,000,000,000,000
giga	G	10^{9}	1,000,000,000
mega	M	10^{6}	1,000,000
kilo	k	10^{3}	1,000
hecto	h	10^{2}	100
deka	da	10^{1}	10
deci	d	10^{-1}	0.1
centi	c	10^{-2}	0.01
milli	m	10^{-3}	0.001
micro	μ	10^{-6}	0.000,001
nano	n	10^{-9}	0.000,000,001
pico micro micro	p μμ	10^{-12}	0.000,000,000,001
femto	f	10^{-15}	0.000,000,000,000,001
atto	a	10^{-18}	0.000,000,000,000,000,001

2. Classic Electrical Measurements

2.1 Indicating Measuring Instruments

2.1.1 General Overview

There are many commonly used measuring instruments, which were considered traditional or classic once. Electromechanical instruments, cheap manually balanced bridge instruments or even the induction type watt-hour meters are still present in our World.

There are several advantages of traditional electromechanical instruments, such as their simplicity, reliability or low price. Perhaps the biggest advantage is that the majority of such instruments can work without an external power supply. Since people's eyes are sensitive to movement, this psycho-physiological aspect of analogue instruments (with a moving needle) is also appreciated.

On the other hand, there are several drawbacks associated with electromechanical analogue instruments. The main disadvantage is that they do not provide an electrical output signal, thus there is a need for an operator to monitor activity during the measurement (at least for the reading of an indicated value). Another drawback is that such instruments generally use moving mechanical parts, which are sensitive to shocks, aging or wearing out. The relatively low cost of moving pointer instruments today is not as apparent as earlier, because of the availability of cheap digital measuring devices with a virtual pointer. For illustration, a traditional, single-phase watt-hour meter is shown in Fig. 2.1.

Figure 2.1 Electromechanical single-phase watt-hour meter

Even though there is no doubt that the future is for automatic, computer supported measuring systems, electromechanical instruments are still present in our lives. For example, attempts to replace analogue instrumentation in cars has not been successful although some dashboards are equipped with 'analogue like' measuring instruments, (or more accurately, displays of measuring instruments), in some new cars as illustrated in Fig. 2.2.

Figure 2.2 Dashboards equipped with 'analogue like' displays

Moreover, without understanding of the principles of older analogue measuring methods it can be difficult to understand more complicated digital instruments, that often use traditional principles of operation. Traditional or classic measurements are still present in our life, and progress is usually observed in more modern measuring techniques.

Electromechanical measuring instruments have several advantages over digital electronic measuring systems such as simplicity, reliability and a lower price. In addition, they do not need additional power in their operation. As mentioned earlier, because there is no electrical output there is need for close monitoring by the operator, but because they have moving parts they are sensitive to shocks. Another negative point is that their power consumption is derived from the measured system, so measurements may be influenced by this.

Finally, it has to be mentioned that measuring techniques have changed significantly recently. Due to the development of informatics, microelectronics and mechatronics a revolution in the science of measurements has occurred recently. In general, measuring devices have been substituted by more flexible and comprehensive computer based measuring systems.

2.1.2 Terms and Definitions

For the characterization of measuring instruments, the following important parameters are defined.

Measuring range is the difference between the highest (x_{max}) and the lowest (x_{min}) measured values with given, predefined uncertainty as given in (2.1).

$$R = X_{max} - X_{min}. \tag{2.1}$$

Measuring sensitivity is a differential value, defined as the difference of indication ($\Delta\alpha$) caused by the unit change in measured value (X). See equation (2.2).

$$E = \frac{\Delta\alpha}{\Delta X} \tag{2.2}$$

In practical life, the so-called *instrument constant* (*C*) is also used. This shows the measured value that causes unit (degree) indication. We can see that the instrument constant is simply the reciprocal of measuring sensitivity. It is referred usually to the upper limit of measured value belonging to the scale end as in (2.3).

$$\mathrm{C} = \frac{X_{max}}{\alpha_{max}} \tag{2.3}$$

For example, the scale of a universal *voltammeter* is shown in Fig. 2.3. If the measuring range is in between 0 and 1000 V, then the instrument constant is 10 (V/^{0}), according to (2.3). If the current measuring range is in between 0 and 3 A (for example), the instrument constant, calculated by (2.3) is 0.1 (A/^{0}). So, the instrument constant is used for transforming measured 'degree in scale' to the measured value.

Figure 2.3 Scale of a voltameter

2.1.3 Moving Coil Meters

These instruments are also called Deprez meters after Marcel Deprez, who together with Jacques-Arsène d'Arsonval, developed a device with a stationary permanent magnet and a moving coil of wire, suspended by fine spring. The structure of Deprez meter is shown in Fig. 2.4.

An iron tube between the magnet's poles defined a circular gap through which the coil rotated. This gap produced a consistent, radial magnetic field across the coil, giving a linear response throughout the instrument's range. A pointer attached to the coil indicates the coil's position. The concentrated magnetic field and delicate suspension meant that these instruments were sensitive.

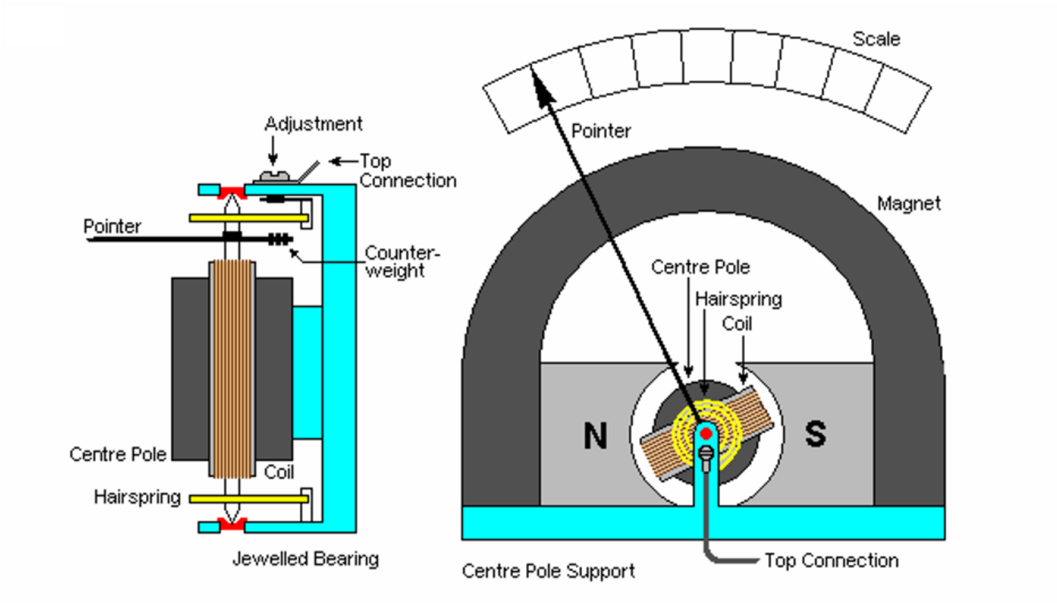


Figure 2.4 Deprez instrument

The rotating torque within the gap can be calculated according to (2.4)

$$M = Bndl \cdot I_m \tag{2.4}$$

where *M* is the rotating torque, *B* is the magnetic flux density within the gap, *n* is the number of turns, *d* is diameter and *l* is the length of coil. Finally, I_m is the electric current to be measured. The returning torque (M_S) caused by the hairspring wires is expressed in (2.5)

$$M_S = k \cdot \alpha \tag{2.5}$$

where *k* is the spring constant and α is the angle of rotation. From the equilibrium condition we can write (2.6).

$$M = M_S \rightarrow \alpha = \frac{Bndl}{k} \cdot I \tag{2.6}$$

This means that the scale of the Deprez instrument is *linear* and it measures the *simple (electrolytic) mean value* of a time varying electric current. Thus, it cannot be used directly for measuring AC, having zero simple mean. Application of moving coil meters for AC measurement needs supplementary elements and will be discussed later on in this chapter.

The measurement sensitivity of the Deprez instrument, expressed from (2.6), is the following

$$E = \frac{\alpha}{I} = \frac{Bndl}{k} \tag{2.7}$$

Regarding the dynamic behaviour, the display mechanism with the mass (of the moving pointer) and the spring constant is a second-order dynamic system that can be over-damped, critically damped or under-damped. Where the mechanism is under-damped, the pointer will move under the measurement as given in Fig. 2.5. This oscillation of the pointer is undesirable because of the useless extra movement and difficulty of getting a reading. For easy calculation the period of oscillation (T_0) and decay time i.e. the time constant (T) are given in (2.8), where b is the degree of damping and P is the damping coefficient.

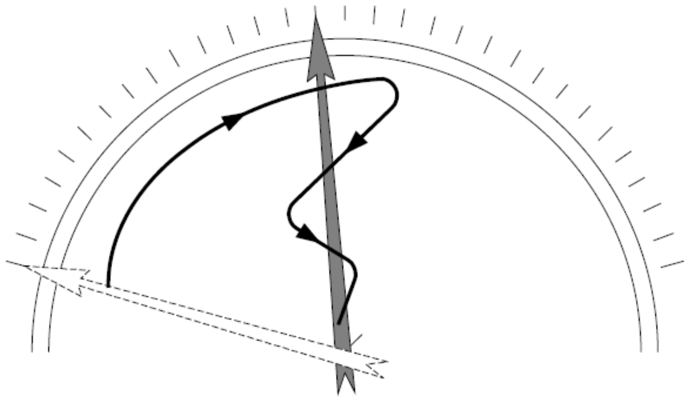


Figure 2.5 Under-damped mechanical construction

$$T_0 = 2\pi\sqrt{\frac{m}{k}}\,, \qquad T = \frac{T_0}{\sqrt{1-b^2}}\,, \qquad b = \frac{P}{2\sqrt{mk}} \tag{2.8}$$

Drawing symbol of moving coil meter is shown in Fig. 2.6.

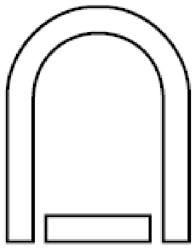


Figure 2.6 Drawing symbol of Deprez meter

2.1.3.1 Current and Voltage Measurement, Measuring Range Extension

The moving coil device can be used as a micro-ammeter without any additional elements. If the coil has r resistance, see Fig. 2.7, the measured current I_x causes $U_x = r \cdot I_x$ voltage drop across the device, which means it can also be used as millivolt meter. For example, if 1 mA current belongs to the scale end i.e. 1 mA is the current measuring range and coil resistance is 60 Ω, then the voltage measuring range is 60 mV. This device, without any additional element, is often called a *default meter*.

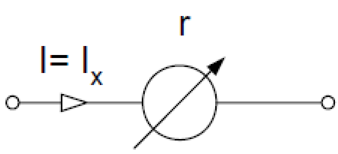


Figure 2.7 Default meter as micro-ammeter

If an additional resistor R_d (*series resistor*) is connected in series with the moving coil device, as it is shown in Fig. 2.8, then we obtain the millivolt-meter or voltmeter with an extended measuring range.

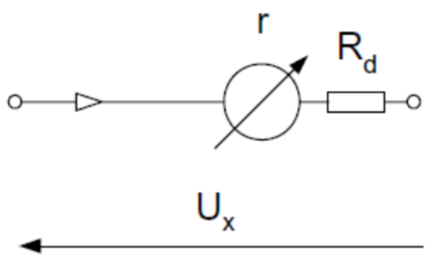


Figure 2.8 Voltage measuring range extension

The measured U_x voltage is

$$U_x = I(r + R_d) \tag{2.9}$$

And the device voltage is U_b, then

$$U_x = \frac{U_b}{r}(r + R_d) \tag{2.10}$$

Thus, we can express the ratio of the measured U_x to U_b as the function of the ratio of R_d and r.

$$\frac{U_x}{U_b} = \frac{R_d}{r} + 1 \tag{2.11}$$

Or, expressing the ratio of R_d to r with the ration of U_x and U_b

$$\frac{R_d}{r} = \frac{U_x}{U_b} - 1 \tag{2.12}$$

When the millivolt-meter in Fig. 2.8 is connected in parallel with another resistor R_b, called a *shunt resistor* we get an ammeter with an extended measuring range, as is shown in Fig. 2.9.

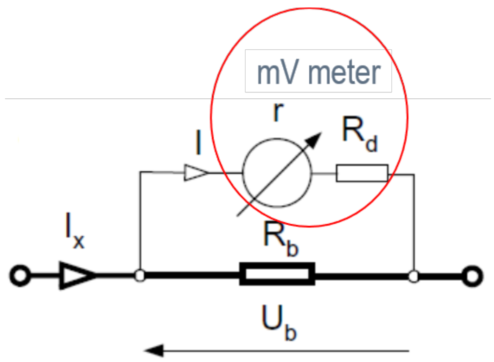


Figure 2.9 Current measuring range extension

Introducing internal device resistance R_m as given in (2.13)

$$R_m \triangleq r + R_d \tag{2.13}$$

The measured current is

$$I_x = \frac{U_b}{R_m \times R_b} = \frac{U_b(R_m + R_b)}{R_m \cdot R_b} \quad (2.14)$$

From this equation we can express the measured current as the function of the shunt resistor R_b and device parameters I and R_m, as can be seen in (2.15).

Thus, we can express the ratio of the measured I_x to I as the function of the ratio of R_d and r.

$$\frac{I_x}{I} = \frac{R_m}{R_b} + 1 \quad (2.15)$$

Or, expressing the ratio of R_m to R_b with the ration of I_x and I

$$\frac{R_m}{R_b} = \frac{I_x}{I} - 1 \quad (2.16)$$

Example – Four-Terminal Shunt

In practical measurements we often use the so-called four-terminal shunt connected in parallel to the millivolt-meter or even *Digital Volt Meter* (*DVM*) Fig. 2.10. demonstrates the connection of the four-terminal shunt and also its possible practical construction.

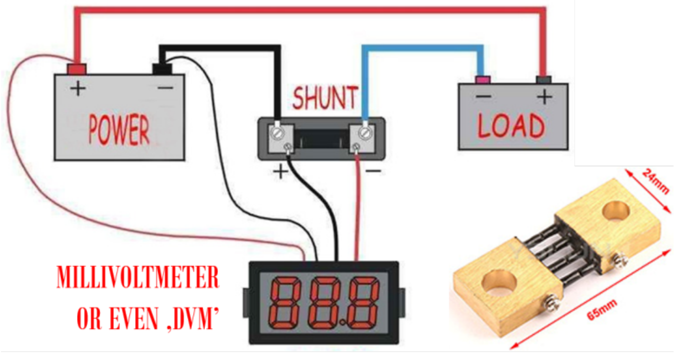


Figure 2.10 Four-terminal shunt connection

Four-terminal connections (4T connection), or the 4-wire connection method is a measuring technique that uses separate pairs of current-carrying and voltage-sensing electrodes to make more accurate measurements than the simpler and more usual two-terminal (2T) connection. Separation of current and voltage (low current!) connections, eliminates the lead and contact resistance from the measurement. This is an advantage for precise current measurement. Four-terminal sensing is also known as *Kelvin sensing*, after William Thomson, Lord Kelvin, who invented the Kelvin bridge in 1861 to measure very low resistances using four-terminal sensing. The bridge method for measurement will be discussed later on.

Universal Multi-Range Voltammeter (Example)

Finally, a possible design of a combined ammeter and voltmeter (called a voltammeter) with selectable ranges is presented in Fig. 2.11.

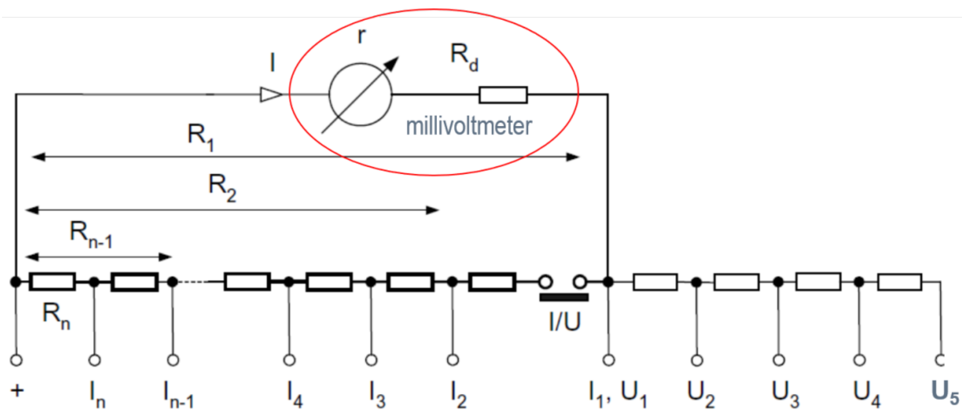


Figure 2.11 Multi-range voltammeter

To make a multi-range ammeter, the application of a *universal shunt resistor* can be useful. The universal shunt resistor is designed to obtain the same current *I* for various input currents, according to (2.17) and (2.18).

$$(I_n - I)R_n = I[(r + R_d) + (R_1 - R_n)] \tag{2.17}$$

$$(I_{n-1} - I)R_{n-1} = I[(r + R_d) + (R_1 - R_{n-1})] \tag{2.18}$$

After simple calculations we obtain the condition of universal shunt resistor in the form of (2.19).

$$\frac{I_n}{I_1} = \frac{R_1}{R_n} \tag{2.19}$$

2.1.3.2 The Moving Coil Meters for AC Measurement

We have already seen that a moving coil meter measures the *simple (electrolytic) mean value* of the electric current, according to its working principle. Thus, it can be used directly for measuring DC current. Because the simple mean of the AC signal is zero, the Deprez meter can only be used for AC measurement with an additional rectifier applied. The output of the rectifier circuit can be either the absolute mean or the peak value of the AC signal, thus we can measure either the absolute mean or the peak value with this instrument. There is no way to measure the RMS value of the AC signal when these rectifiers are applied. Because the RMS value is the most commonly used and required mean value of AC signals, these devices are scaled according to the RMS value of the AC (pure sinusoidal) signal. When measuring non-sinusoidal signals, we cannot say anything about the RMS value, but the following considerations are valid.

If a peak rectifier is applied in the device, we can calculate the peak value from the measured signal according to (2.20).

$$U_{peak} = \sqrt{2} \cdot U_{MTR} \tag{2.20}$$

where U_{MTR} is the measured value.

If an absolute mean value rectifier is applied in the device, we can calculate only the absolute mean value from the measured signal according to (2.21).

$$U_{abs} = \frac{U_{MTR}}{1.11} \tag{2.21}$$

The symbol of the rectifier equipped moving coil meter is shown in Fig. 2.12.

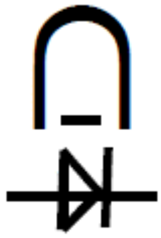


Figure 2.12 Symbol of the Deprez device for AC measurements

2.1.4 The Moving Iron Meters

The main advantage of a moving iron meter is that it can measure the RMS value of the signal. Therefore, it directly can be used for AC measurements. A possible design of a moving iron meter with stationary coil and moving iron is presented in Fig. 2.13. The deflection in the pointer is caused by the attraction between the current carrying coil and the iron.

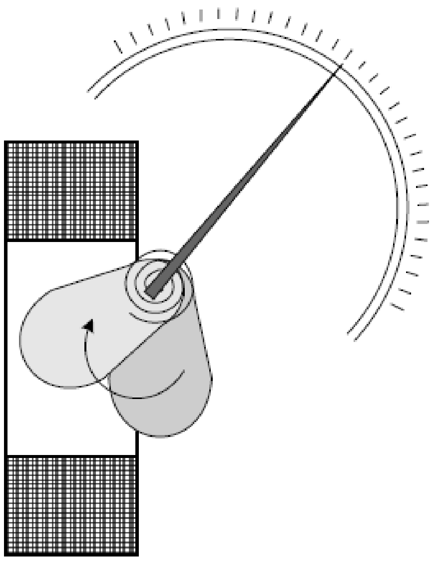


Figure 2.13 Moving iron meter with single moving iron

Another possible design of moving iron meter is shown in Fig. 2.14 where both a stationary and a moving iron is located within the stationary coil. The deflection of the pointer is caused by the repelling force between the irons.

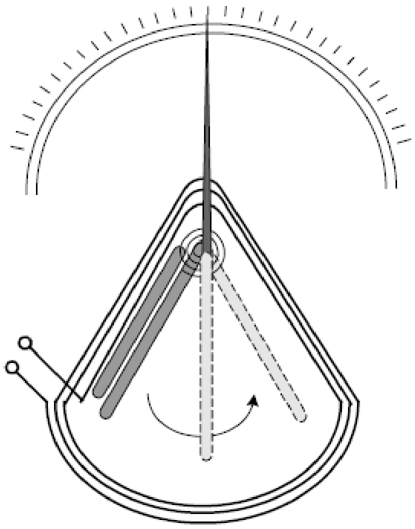


Figure 2.14 Moving iron meter with stationary and moving iron

The angular deflection depends on the measured current *I* and the change of the inductance *dL* caused by this deflection as described in (2.23).

$$\alpha = \frac{1}{2 \cdot k} \cdot \left\{ \boldsymbol{\frac{dL}{d\alpha} \cdot I^2} \right\} \qquad (2.23)$$

where α is the angular deflection and *k* is a construction dependent constant.

Although, the deflection is a nonlinear function of the measured current it is possible to design the device in such a way that the right side of the expression in brackets, **signed bold** in (2.23), is close to linear. Because the response of the device depends on the squared value of the current it is possible to obtain the RMS value. Due to errors caused by magnetic hysteresis (when *DC* current is measured) these devices are used almost exclusively for *AC* measurements.

Because of the low mass of the moving iron the mechanical construction is under-damped, and in most cases, additional mechanical damping is applied. A possible solution is shown in Fig. 2.15.

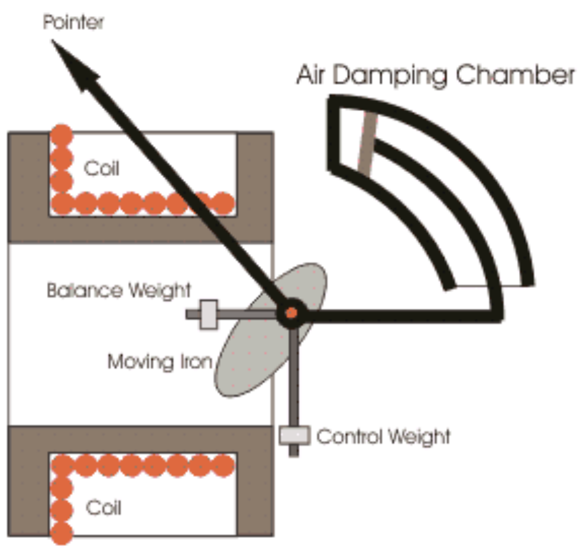


Figure 2.15 Additional damping in moving iron meter

The moving iron meter exhibits several advantages: simplicity of design – no need to supply the moving element, the range can be easily changed by selecting the number of the turns in the coil. The drawbacks of moving iron devices are their relatively large power consumption (0.1 – 1 VA) and small sensitivity (in comparison with a moving coil device). The smallest obtainable range of a moving iron milliammeter is several mA. The frequency bandwidth is also limited to less than approximately 150 Hz.

The symbol of the moving iron meter is show in Fig. 2.16.

Figure 2.16 Symbol of moving iron meter

2.1.5 The Electrodynamic Meters

The electrodynamic meters were formerly the most accurate indicating instrument but nowadays these instruments have been substituted for more accurate digital devices. The electrodynamics devices are still used as a wattmeter. The principle of operation is illustrated in Fig. 2.17.

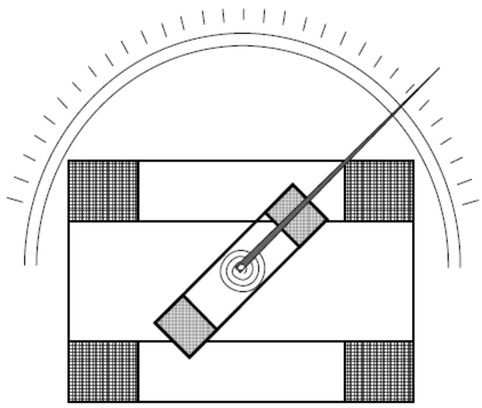


Figure 2.17 The structure of the electrodynamic meter

The electrodynamics device has two coils: a stationary coil and a moving one. The currents flowing through these coils induce a force, which causes rotation of the movable coil. The torque M resulting from the interaction between two coils depends on currents: I_1 in the stationary coil, I_2 in the movable one and the phase shift between these currents as given in (2.24).

$$M = c \cdot I_1 \cdot I_2 \cdot \cos\varphi \tag{2.24}$$

Thus, if one coil is connected to the current and the second to the voltage we can directly measure the power, given by (2.25).

$$P = U \cdot I \cdot \cos\varphi \tag{2.25}$$

An electrodynamic device can also be used for current or voltage measurement, connecting both coils in series to the same signal. This device has a linear scale for power measurement and square scale for current and voltage measurements as seen in (2.24).

The symbol of the electrodynamic device is shown in Fig. 2.18.

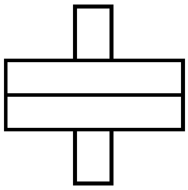


Figure 2.18 Symbol of the electrodynamic meter

Measuring Electrical Power

Fig. 2.19 presents a typical connection of electrodynamic meter as the wattmeter.

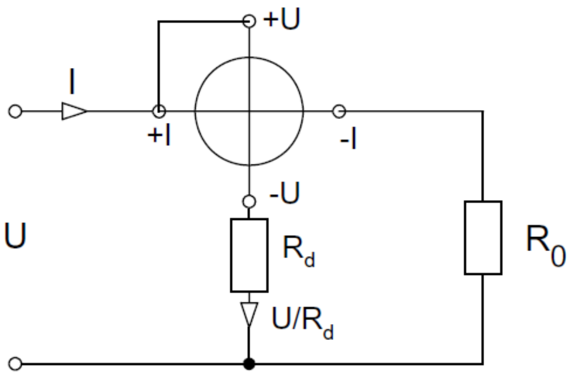


Figure 2.19 The wattmeter

The wattmeter has two pairs of terminals – the current and the voltage terminals. In the voltage circuit there is usually a series resistor R_d added. Thus, according to (2.24), the torque can be calculated as given in (2.26).

$$M = c \cdot \frac{1}{R_d} \cdot U \cdot I \cdot \cos\varphi = k \cdot P \qquad (2.26)$$

where c is the constant from (2.24) depending on the construction and *k* is the constant (containing also the R_d series resistance) showing the torque is proportional to the power *P* to be measured.

Measuring Three-Phase Active Power

The wattmeter is often used for measurements in three-phase systems. A simple example of three-phase power measurement is given in Fig. 2.20 The three-phase power meters can be used as three separate meters or one meter with three electrodynamic devices with a common axle.

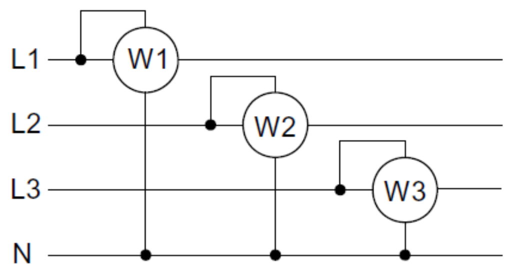


Figure 2.20 Three-phase wattmeter with neutral wire connected

The connection, shown in Fig. 2.21, is also often used because of the virtual neutral point at the common point of the voltage coils.

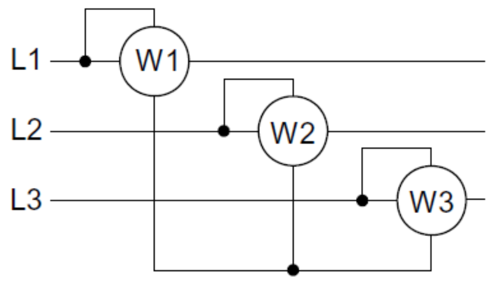


Figure 2.21 Three-phase wattmeter without neutral wire connected

The total three-phase power is the sum of the measured values as given in (2.27).

$$P = P_{W1} + P_{W2} + P_{W3} \tag{2.27}$$

Where a load is balanced, the value of three phase currents is the same, thus, it is sufficient to connect one wattmeter in single phase. The total power is given in (2.28).

$$P = 3 \cdot P_{W1} \tag{2.28}$$

Measuring Three-Phase Reactive Power

The reactive power can also be measured by the connection of the wattmeters as shown in Fig. 2.22.

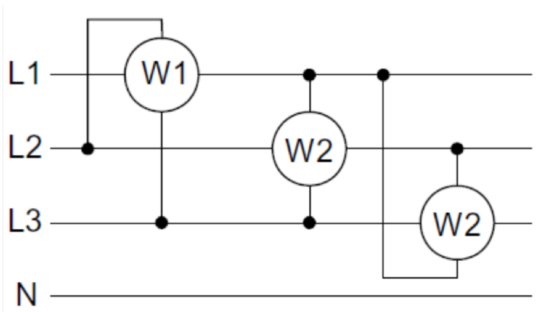


Figure 2.22 Measuring three-phase reactive power

The voltage-current phasor diagram of the phase L1 is given in Fig. 2.23.

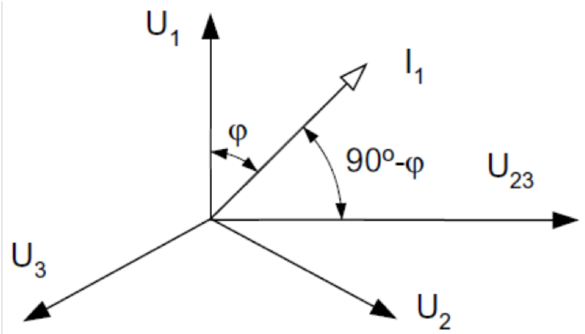


Figure 2.23 Phasor diagram of the three-phase load

Where U_{23}, line voltage is measured between U_2 and U_3 phase voltages ie. $U_{23} = U_2 - U_3$.

In the case of a symmetric voltage system, P_{W1} power, measured by W1 wattmeter, is calculated in (2.29).

$$P_{W1} = I_1 U_{23} \cos(90 - \varphi) = \sqrt{3} U_1 I_1 \sin\varphi = \sqrt{3} Q_1 \tag{2.29}$$

Thus, the total reactive power in the three-phase system is given in (2.30).

$$Q = \frac{P_{W1} + P_{W2} + P_{W3}}{\sqrt{3}} \tag{2.30}$$

Notice that (2.30) is valid even in case of an *unbalanced load.*

Aron Method for Three-Phase Power Measurement

In a three-phase system without the neutral wire – the so-called *three-wire system* makes it possible to use the measuring system with two watt meters. This method, known as the *Aron method*, is presented in Fig. 2.24.

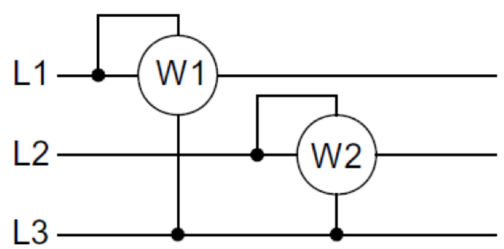


Figure 2.24 Aron method

The condition of the three-wire connection is given in (2.31)

$$i_1 + i_2 + i_3 = 0 \tag{2.31}$$

from which we can express i_3 as in (2.32).

$$i_3 = -(i_1 + i_2) \tag{2.32}$$

The instantaneous power in three-phase system is calculated in (2.33).

$$p = u_1 i_1 + u_2 i_2 + u_3 i_3 \tag{2.33}$$

Substituting (2.32) into (2.33) we obtain the followings.

$$p = (u_1 - u_3)i_1 + (u_2 - u_3)i_2 \tag{2.34}$$

Because the voltage system is balanced, we can write it as (2.35).

$$p = u_{13}i_1 + u_{23}i_2 \tag{2.35}$$

Thus, the total power is the sum of the powers indicated by the wattmeters as shown in (2.36)

$$P = \frac{1}{T}\int_0^T p\,dt = U_{13}I_1\cos(U_{13}, I_1) + U_{23}I_2\cos(U_{23}, I_2) = P_{W1} + P_{W2} \tag{2.36}$$

The phasor diagram in case of the balanced load is given in Fig. 2.25.

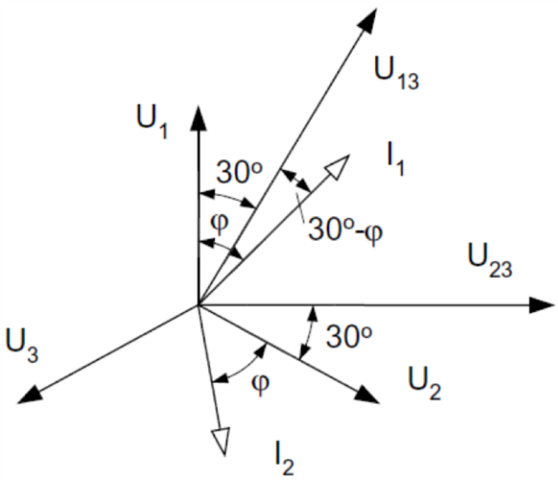


Figure 2.25 Phasor diagram in case of balanced load

According to the phasor diagram the wattmeters indicate the values given in (2.37) and (2.38).

$$P_{W1} = U_L I_P \cos(30 - \varphi) \tag{2.37}$$

$$P_{W2} = U_L I_P \cos(30 + \varphi) \tag{2.38}$$

where U_L is the line voltage and I_P is the phase current. Using the Aron circuit for a balanced system we can easily determine the reactive power Q and the phase angle as shown in (2.39) and (2.40).

$$\varphi = \tan^{-1}\left\{\sqrt{3}\frac{P_{W1} - P_{W2}}{P_{W1} + P_{W2}}\right\} \tag{2.39}$$

$$Q = \sqrt{3}(P_{W1} - P_{W2}) \tag{2.40}$$

This method has some drawbacks; first of all, it requires a three-wire connection. Each short circuit to the ground or leakage current causes an incorrect measurement. But in the power electrical systems, two wattmeters measuring systems are often used for economic reasons. The wattmeters are relatively cheap but the measuring transformers, which are necessary to isolate and reduce the currents and voltages, are often relatively expensive.

2.2 Waveform Measuring Instruments

A waveform measurement is typically used to measure and display the voltage of a signal with respect to time. The electronic test instrument that allows observation of varying signal voltages is called an oscilloscope that we will discuss in the following section.

2.2.1 Oscilloscope Basics

The oscilloscope is arguably one of the most useful general-purpose tools ever created for electronic engineers. Since its invention more than 100 years ago, new types, features and functionalities have been introduced. In the beginning all oscilloscopes were analogue. As digital technologies advanced oscilloscopes have since moved from being analogue to digital, and more features and measurement functionalities have been added, but some of the key basic considerations remain the same.

The principle of displaying the signal in the oscilloscope is illustrated in Fig. 2.26. The picture on the screen is obtained in such a way that during the horizontal movement of luminous point it is deflected vertically proportional to the value of the detected U_x signal. The horizontal movement is obtained by the sweep oscillator generating a saw-tooth signal. If the frequency of the signal is larger than several Hz due to the inertia of our eyesight it is not possible to see such a picture of the signal. Therefore, the main function of the oscilloscope is to somehow freeze the picture on the screen.

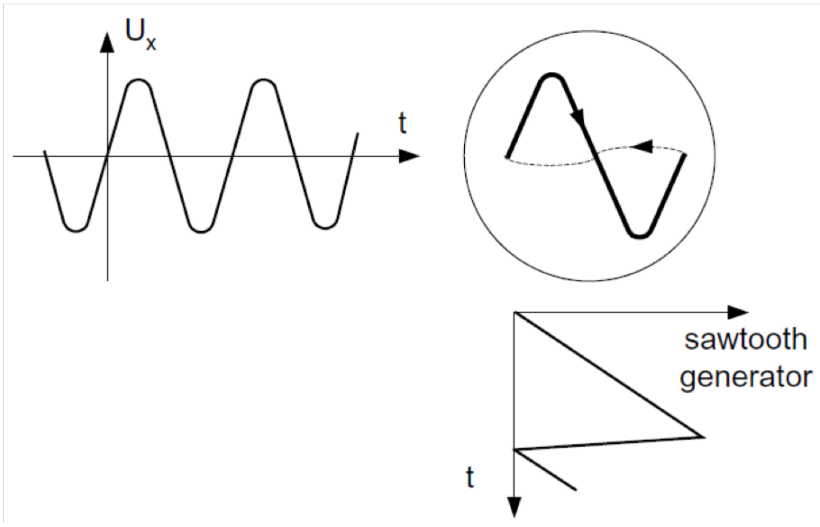


Figure 2.26 Principle of operation of the oscilloscope

If the period of oscillation of the saw-tooth signal is the same (or multiple) as the period of investigated signal the successive pictures appear to be the same. This creates the illusion that the picture is standing still. We say that the investigated signal is *synchronized* with the sweep signal.

2.2.2 Features of Oscilloscopes

Bandwidth

Oscilloscope bandwidth is the range of frequency between the lower and the upper cut-off frequency belonging to the given attenuation. Oscilloscope bandwidth is specified at the -3 dB or 3 dB down point. Therefore, the maximum bandwidth of an oscilloscope is defined as the frequency at which the sinusoidal input signal amplitude is decreased by 3dB. Typically, the bandwidth is around 1 GHz for modern oscilloscopes.

Sensitivity

This parameter is adjustable and belongs to the vertical deflection or vertical gain. Providing the amount of voltage, causes one division (DIV), ie. 1 cm vertical deflection on the screen. Typical values are between 1 mV/cm and 100 V/cm.

Time Deflection Rate

This parameter is also adjustable and belongs to the horizontal deflection or horizontal gain, giving the *t* which corresponds to one DIV (1 cm) horizontal deflection on the screen. Typical values are between 10 ns/cm and 50 s/cm.

Additional Features

There are some additional features like synchronization or trigger modes, multichannel or multibeam features, *On Screen Display* (OSD) and *Auto Set Function* (ASF) features, they will be discussed in more details later on. For illustration, Fig. 2.27 gives an example for an oscilloscope 'screen' displaying not only the wave form but the numeric value of the measured voltage at the cursor (marker) position. Settings of the oscilloscope are also displayed such as coupling mode of the measured signal, attenuation of test probe, displayed channel, etc.

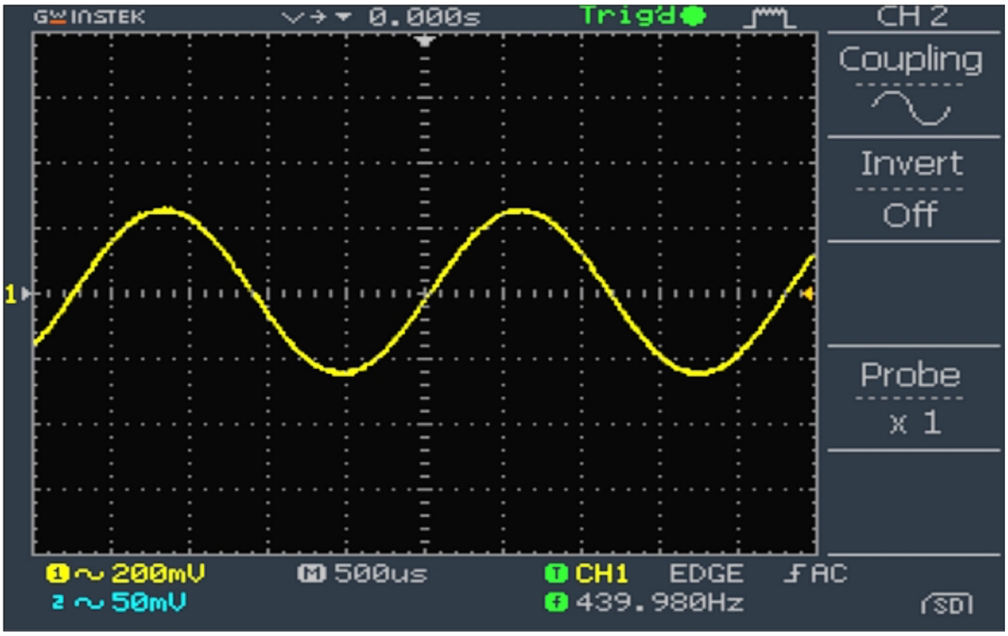


Figure 2.27 OSD screen of an oscilloscope

2.2.3 The Types of Oscilloscopes

Oscilloscopes can be categorized from different points of view, but the two main types are *analogue* and *digital*. When using the *time sampling* operation mode, we can distinguish between *real time* and *equivalent time* modes, that are high level operation methods for increasing the bandwidth of the measurement. According to the applied technologies oscilloscopes can be equipped with *Cathode Ray Tube* (*CRT*) screens or *LCD* (TFT ...OLED) displays.

Analog Oscilloscopes

The block diagram of a typical Analog Real Time (ART) oscilloscope, equipped with CRT display is presented in Fig. 2.28.

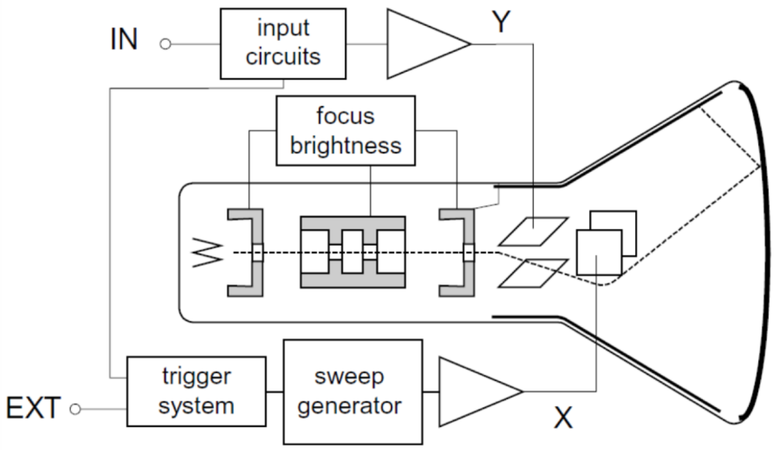


Figure 2.28 Analogue CRT oscilloscope

The electron beam from the cathode of the CRT is passes through the acceleration anode and focusing area. This beam then strikes the phosphor coating on the screen causing the light. Depending on the composition of the phosphor coating it is possible to obtain various colours of the light (green, blue or yellowish green) and time of the lighting. Usually the system of deflection of the electron beam consists of two pairs of plates: horizontal and vertical ones. The electrostatic force, depending on the voltage, causes the deflection of the electron beam and the amount of deflection is directly proportional to the voltage applied to the plates.

The 'standing picture illusion' can be obtained if the same signal part is presented on the same position on the screen again and again. The trigger system is responsible for this synchronization as the trigger signal starts (triggers) the sweep pulse for horizontal deflection at a certain trigger level in a given direction. After running-off the sweep pulse, the new trigger makes it possible to start the next sweep.

It is important to note that, based on this idea, *periodic signals* can be observed although, applying *long post-luminescence* CRT this method is also applicable for testing *non-periodic signals*. The *signal storage feature* of these CRTs is used for example in former locator technologies.

Digital Oscilloscopes

A Digital Storage Oscilloscope (often abbreviated DSO) is an oscilloscope which stores and analyses the signal digitally rather than using analogue techniques. It is now the most common type of oscilloscope in use because of the advanced trigger, storage, display and measurement features which it typically provides.

The input analogue signal is sampled and then converted into a digital record of the amplitude of the signal at each sample time. The sampling frequency should be not less than the Nyquist rate to avoid aliasing. These digital values are then converted back into an analogue signal for display on a cathode ray tube (CRT) or transformed as needed for the required output, such as liquid crystal display, chart recorder, plotter or network interface.

The principle operation of DSO can be followed in Fig. 2.29. The input signal to be measured is connected to an Analog to Digital Converter (ADC) after the (optional) signal condition.

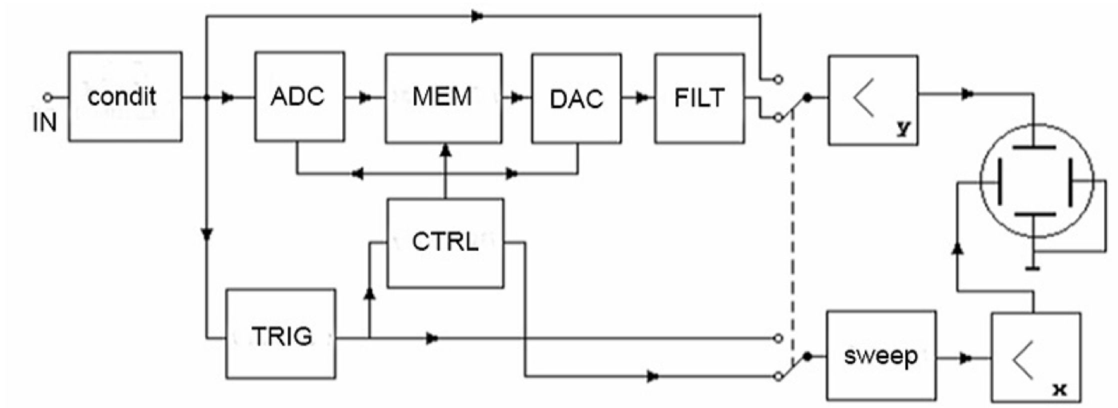


Figure 2.29 Digital storage CRT oscilloscope

The ADC circuit converts the analogue signal to 8-bit digital data allowing it to distinguish 256 different levels in the measuring signal. This digital data can easily be stored in the MEMory (MEM) unit. To enable this data to be displayed through an analogue oscilloscope requires a Digital to Analog Converter (DAC). The control (CTRL) of the 'ADC-MEM-DAC' chain of digital signal processing and storing is driven by the trigger system (TRIG) that can be started by either a prescribed level of input signal or by an external trigger signal.

2.2.4 Main Units and Controls

2.2.4.1 CRT and LCD Display Units

The Cathode Ray Tube (CRT) is a vacuum tube that contains one or more electron guns and a phosphorescent screen. It modulates, accelerates, and deflects electron beam(s) onto the screen to create the images representing electrical waveforms. The principle of operation of a CRT tube can be followed in Fig. 2.30.

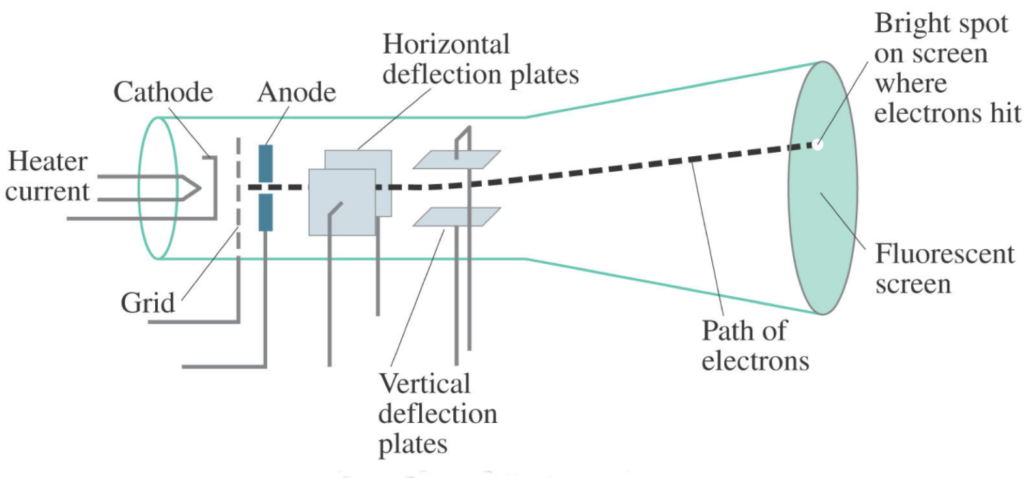


Figure 2.30 The cathode ray tube

The electron beam, emitted by the heated cathode, is controlled and focused by grid and anode electrodes. In CRTs, used in oscilloscopes, the electron beam is bent by an electrostatic deflection system. A CRT is constructed from a glass envelope which is large, deep (i.e. long from front screen to the rear end), fairly heavy, and relatively fragile. The interior of a CRT is evacuated to facilitate the free flight of electrons from the cathode to the tube's face.

Since 2000, many CRTs have been substituted by newer 'flat panel' displays such as LCD or OLED displays, especially in hand-held oscilloscopes. Liquid crystal materials consisting of rod-shaped molecules that can be ordered by an electrostatic field. The display is divided into pixels, each of which is driven by an individual electrode.

A hand-held oscilloscope is shown in Fig. 2.31 for demonstration.

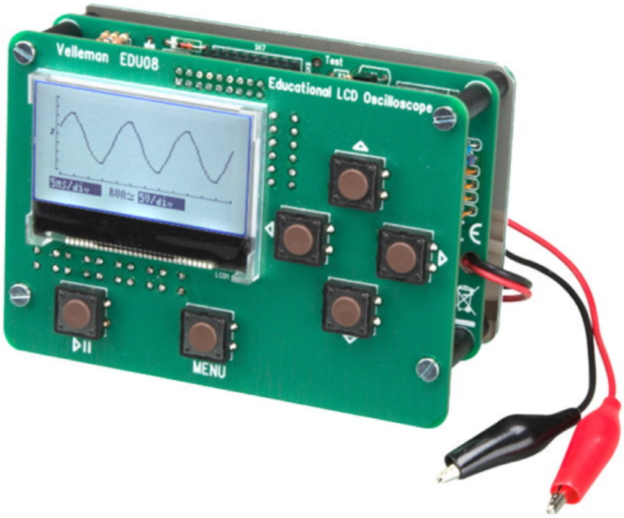


Figure 2.31 Hand-held oscilloscope equipped with LCD

2.2.4.2 Vertical Deflection System

The function of the vertical deflection system is to provide an amplified signal to drive the vertical deflection plates without introducing any appreciable distortion into the system.

The input sensitivity of many oscilloscopes is of the order of a few millivolts per division and the voltage required for deflecting the electron beam varies from approximately 100 V (peak to peak) to 500 V depending on the accelerating voltage and the construction of the CRT. Thus, the vertical amplifier is required to provide this desired gain from a millivolt input to several hundred volts (peak to peak) output. Also, the vertical amplifier should not distort the input waveform and should have a suitable response for the entire band of frequencies that are measured. The deflection plates of CRT act as plates of a capacitor and when the input signal frequency exceeds 1 MHz, the current required for charging and discharging the capacitor formed by the deflection plates increases. So, the vertical amplifier should be capable of supplying enough current to charge and discharge the deflection plate capacitor.

Because the electrical signal is slightly delayed when transmitted through an electronic circuit, with CRT oscilloscopes, the output signal voltage of the vertical amplifier is fed to the vertical plates of CRT and some of it is used for triggering the time base generator circuit, whose output is supplied to the horizontal deflection plates through the horizontal amplifier. The whole process, which includes generating and shaping a trigger pulse and starting a time-base generator and then its amplification, takes approximately 100 nanoseconds. Thus, the input signal of the vertical deflection plates of a CRT needs to be delayed by at least the same time or slightly more to allow the operator to see the leading edge of the signal waveform under study on the screen. For this purpose, a delay line circuit is introduced between the vertical amplifier and the plates of CRT.

In typical oscilloscopes, it is possible to use the dual channel mode, although the display uses only a single beam of electrons. A vertical system, equipped with signal *multiplexer* (MPX), is shown in Fig. 2.32. The so called *dual-channel electro switch* can be controlled by a free-running oscillator or this control signal can be triggered by the horizontal system.

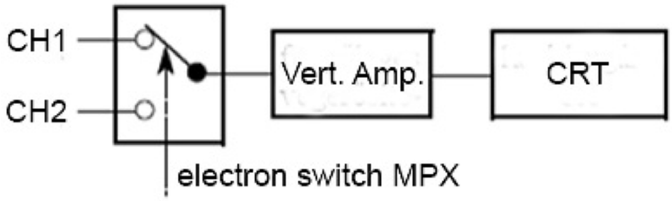


Figure 2.32 Dual channel vertical deflection system

Two signals can be displayed using the *chopper* or *alternate* mode as is illustrated in Fig. 2.33. In the alternate mode, both signals are flashed on the screen alternately – but due to the inertia of human eyesight and the persistence of lighting, it appears that both signals are shown at the same time.

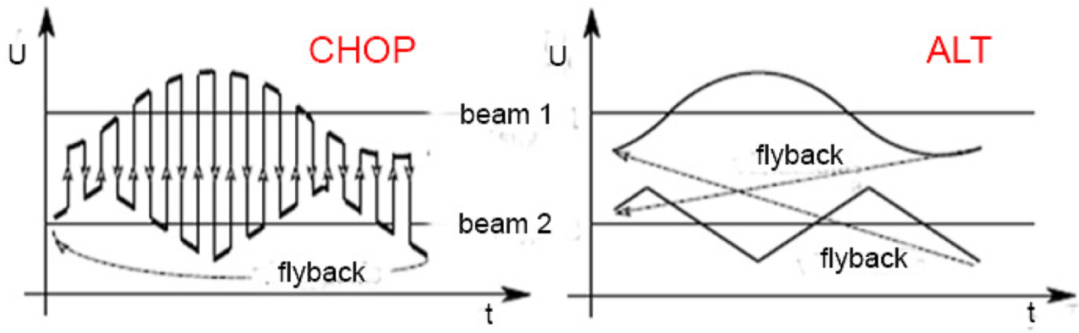


Figure 2.33 Chopper and alternate modes

The alternate mode cannot be used for low frequency signals because it can result in blinking of the picture. For low frequency signals the chopper mode should be used where both signals are chopped – a small part of each signal flashes alternately on the screen. If the chopping rate is much higher than the frequency of the signals, then the observer sees both signals as continuous lines on the screen.

2.2.4.3 Horizontal Deflection and Trigger System

The horizontal amplifier, similar to the vertical amplifier, increases the amplitude of the input signal to the level required by the horizontal deflection plates of CRT. Fig. 2.34 presents the principle of synchronization of the investigated signal. The saw-tooth voltage of the sweep generator is initiated by the pulse from the triggering system. In the simplest case, the automatic trigger mode can occur for the zero value of signal – it is important to start every time with a precisely defined point of signal, because only in such cases can we get a stationary picture on the screen. Different sources, such as internal or external signals, or even power networks (50 Hz) can be applied to horizontal deflection plates through the horizontal amplifier according to the INT, EXT or LINE position of the sweep selector switch, as shown in Fig. 2.34.

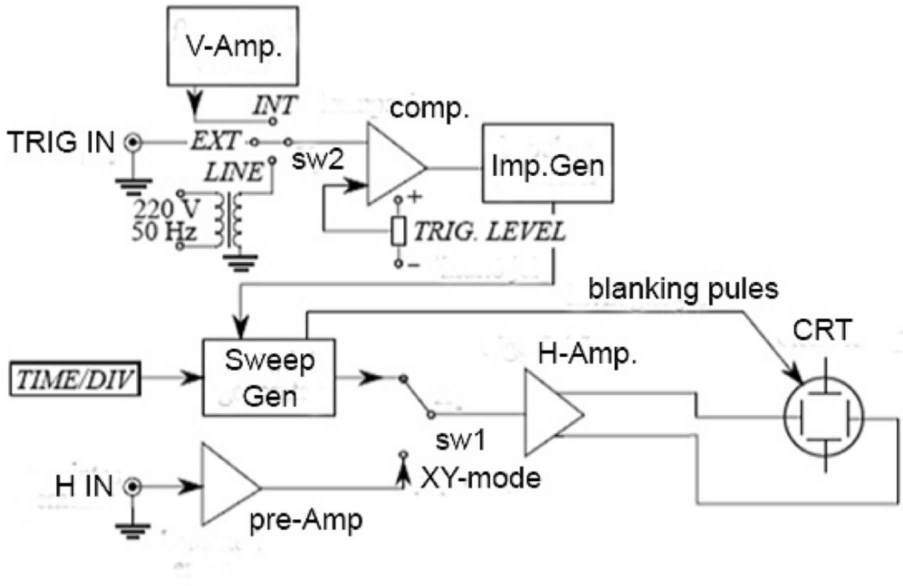


Figure 2.34 Horizontal deflection system

When the function of time is required to be displayed on the screen of the CRT, the INT position of the sweep selector switch should be used. The U_S output of the saw-tooth signal generator is shown in Fig. 2.35.

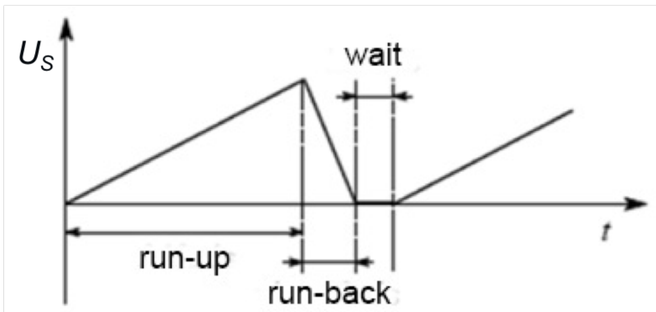


Figure 2.35 The saw-tooth signal of the sweep generator

At the trigger starting point the U_S voltage starts at zero and increases to the spot on the far left of CRT oscilloscope screen. Then, the U_S voltage reaches its maximum shown on the far right side of the screen. Thus, saw-tooth sweep voltage is applied to horizontal deflection plates of CRT, which moves the spot on the screen following a straight horizontal line from left to right during the sweep period. The slow sweep period is followed by a fast run-back period and a wait period until the next triggered cycle of saw-tooth voltage signal starts again. During the run-back and wait periods the spot on the screen is invisible because the electron beam is blanked. This blanking circuit is controlled also by the trigger system.

Based on the previous description we can conclude that:

- The spot moves from left to right over the same path again for every cycle of the saw-tooth voltage applied to the horizontal deflection plates, so a horizontal line appears on the screen of CRT oscilloscope;
- The spot moves from left to right on the screen with at a constant speed, thus, it produces a linear time base to display the function of time on the screen of the CRT oscilloscope.

Now, let's suppose a sine-wave voltage signal U_M of time period T, is applied to the vertical deflection plates and a saw-tooth voltage signal U_S of time period T is applied to the horizontal deflection plates. At zero time, the spot is far left of the vertically central position on the screen. Because of zero value of U_M and zero values of U_S. At time $T/4$, the spot is at one-fourth of the way on the screen in the horizontal direction and at maximum positive deflection above the centre line in the vertical direction because of the maximum positive value of U_M. At time $T/2$, the value of U_M is zero and U_S reaches the half of the maximum value, so the spot is on the central position of the screen. At time $3T/4$, the spot is the three-fourths the way on the screen in a horizontal direction and the maximum negative deflection in the vertical direction. Finally, at the end of time T, the spot is on the far right vertically central position of the screen and then it moves back to begin a new trace with a blank screen. In this way, a sine-wave voltage is applied to the vertical deflection system which appears on the screen. If the period of sine-wave is reduced to half, then two sine-wave cycles appear on the screen.

2.2.5 X-Y Operation Mode

Usually the oscilloscope displays the time varying signals, so the horizontal axis is the time axis. But it is also possible to connect another signal to the horizontal deflection system. This way, we can display on the screen the function $Y=f(X)$. Fig. 2.36 and Fig 2.37 present two examples of the X-Y operation.

In the first example, the signal proportional to the current (voltage drop on the resistor) is applied to the vertical plates while the signal proportional to the voltage on the diode is connected to the horizontal plates. Thus, we obtain the characteristic $I=f(U)$ of the diode on the screen of the oscilloscope.

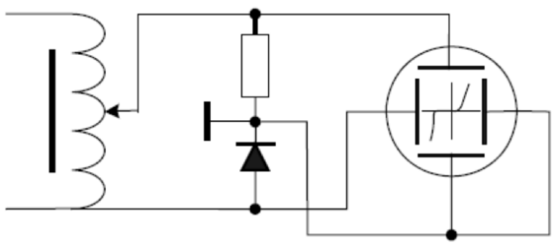


Figure 2.36 Test circuit for displaying diode characteristic

The second example in Fig. 2.37 illustrates the possibility of investigating the hysteresis loop (*B-H* loop) of magnetic soft materials. The signal, proportional to the magnetic field strength (that is proportional to the magnetizing current), is connected to the horizontal plates and to the signal, proportional to the flux density i.e. the voltage U induced in the winding after integration, according to the Faradays law given in (2.41), is connected to the vertical pair of plates.

$$U = f\left\{\frac{dB}{dt}\right\} \tag{2.41}$$

Thus, we can observe the $B = f\{H\}$ loop on the screen of the oscilloscope.

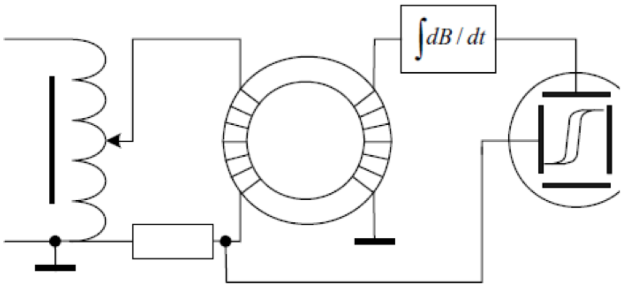


Figure 2.37 Displaying a hysteresis loop for a soft magnetic material

Finally, we mention the measuring method of the X-Y operation mode of CRT oscilloscopes with Lissajous curves as they are commonly used in period, frequency and phase measurement.

Lissajous curves are patterns generated by the pair of sinusoidal waves with axes that are perpendicular to one another. The curves, sometimes called Bowditch curves, were first described in the early 19th century by American mathematician Nathaniel Bowditch. Later that century, French mathematician Jules-Antoine Lissajous began his own study of the curves, which he produced in multiple ways. Nowadays, Lissajous curves are commonly used in CRT oscilloscope X-Y mode measurements of frequency, phase shift, etc. providing a picture of electrical signals in the form of a graph.

For example, if we have a pair of time varying signals, given in (2.42) and (2.43).

$$x(t) = A\sin(\alpha t + \delta) \tag{2.42}$$

$$y(t) = B\sin\beta t \tag{2.43}$$

These signals are connected to the horizontal and vertical inputs of the CRT oscilloscope, operating in X-Y mode. The output curves on the screen, depending on A and B amplitudes, α, β frequencies and δ phase delay, are simulated by a computer program. The result of the simulation (calculation) is shown in Fig. 2.38.

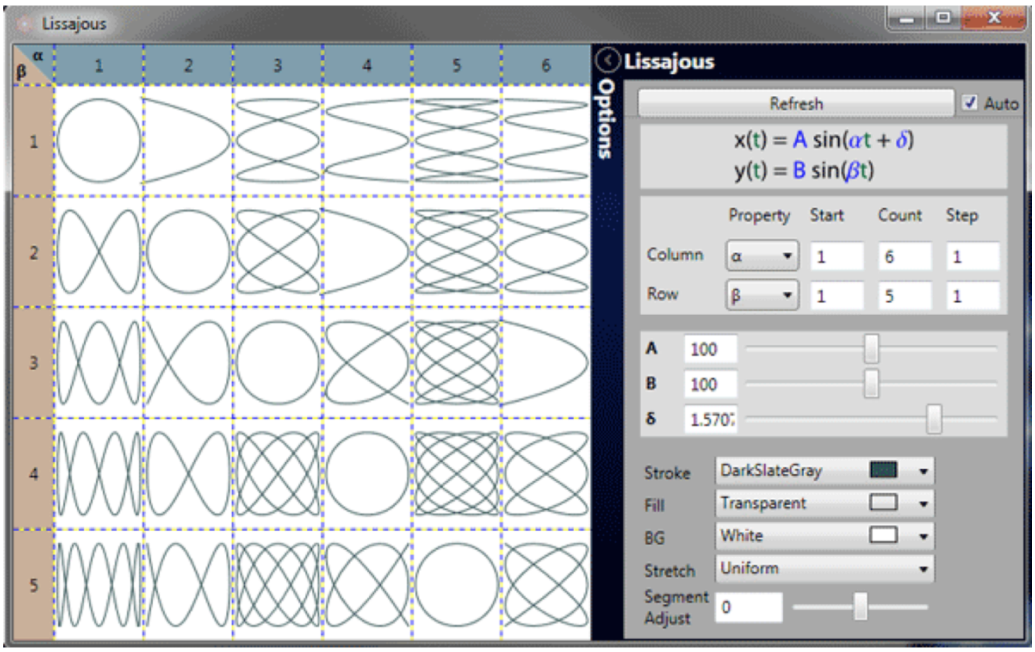


Figure 2.38 Lissajous curves in a computer application

When analysing the Lissajous curves we can describe the following.

- If both of signals have the same frequency the curve on the screen is a circle or ellipse.
- If the frequency of the *y* signal is, for example, twice the frequency of *x* signal, then the curve has (i.e.) two horizontal maximums and minimums on screen.
- If the frequency of *x* signal is, for example, twice the frequency of *y* signal, then the curve has (i.e.) two vertical maximums and minimums on screen.

Other results can also be followed easily in Fig. 2.38.

2.3 Balanced and Unbalanced Bridge Circuits

Formerly, bridge circuits were used as the most accurate devices for measuring resistance and impedance. Nowadays, bridge circuits are not as important as they used to be, because new, more effective methods of impedance measurement have been developed. Bridge circuits are commonly used as the resistance (impedance) to voltage converters and the bridge principle is utilized in digital RLC meters. The main idea of bridge circuits is that, in the case of a balanced condition of the bridge, an output voltage of zero can easily be shown.

2.3.1 Balance and Sensitivity

Two main bridge circuits: supplied by the voltage source or the current source. The voltage supplied bridge is shown in Fig. 2.39.

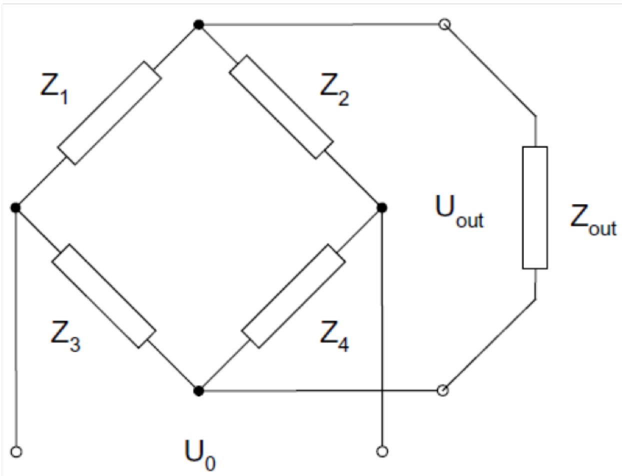


Figure 2.39 Bridge circuit supplied by a voltage source

The output voltage of the voltage supplied bridge is calculated below in (2.44).

$$U_{out} = \frac{Z_1 Z_4 - Z_2 Z_3}{(Z_1 + Z_2)(Z_3 + Z_4) + \frac{Z_1 Z_2}{Z_{out}}(Z_3 + Z_4) + \frac{Z_3 Z_4}{Z_{out}}(Z_1 + Z_2)} U_0 \qquad (2.44)$$

The current supplied bridge circuit is presented in Fig. 2.40 and its output voltage is given by (2.45).

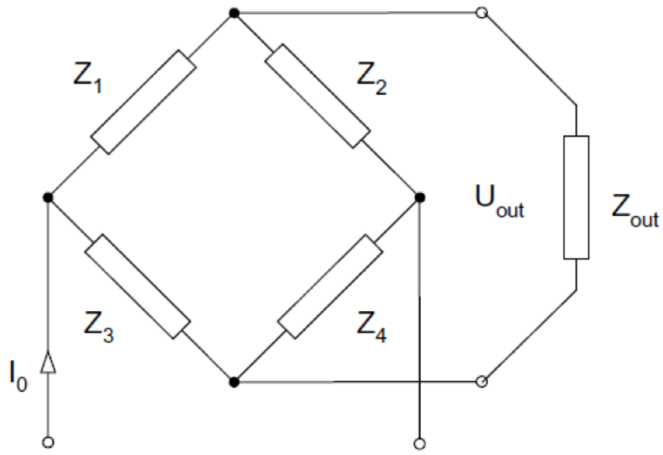


Figure 2.40 Bridge circuit supplied by current source

$$U_{out} = \frac{Z_1 Z_4 - Z_2 Z_3}{(Z_1 + Z_2 + Z_3 + Z_4) + \frac{(Z_1 + Z_3)(Z_2 + Z_4)}{Z_{out}}} I_0 \tag{2.45}$$

When the Z_{out} load impedance is infinitely high, i.e. the load is an open circuit (2.44) and (2.45) can be written as in (2.46) and (2.47).

$$U_{out} = \frac{Z_1 Z_4 - Z_2 Z_3}{(Z_1 + Z_2)(Z_3 + Z_4)} U_0 \tag{2.46}$$

$$U_{out} = \frac{Z_1 Z_4 - Z_2 Z_3}{(Z_1 + Z_2 + Z_3 + Z_4)} I_0 \tag{2.47}$$

From these equations we see that the universal condition of *balance,* i.e. when the output voltage is zero is the following.

$$Z_1 Z_4 - Z_2 Z_3 = 0 \rightarrow Z_1 Z_4 = Z_2 Z_3 \tag{2.48}$$

Because the impedances are described by complex values, (2.48) can be written in the following form

$$|Z_1| \cdot |Z_4| \cdot e^{j(\varphi_1 + \varphi_4)} = |Z_2| \cdot |Z_3| \cdot e^{j(\varphi_2 + \varphi_3)} \tag{2.49}$$

where φ_1, φ_2, φ_3, φ_4 are the phase angles of Z_1, Z_2, Z_3, Z_4 impedances.

To split up the complex equation of (2.49) we can have the magnitude and phase condition of the balanced bridge as written in (2.50) and (2.51).

$$|Z_1| \cdot |Z_4| = |Z_2| \cdot |Z_3| \tag{2.50}$$

$$\varphi_1 + \varphi_4 = \varphi_2 + \varphi_3 \tag{2.51}$$

There are three important outcomes from (2.50) and (2.51). First of all, the condition of the balance is independent from the value of the supplier source (U_0, I_0) and from the value of the Z_{out} load that is usually the indicator device. The second thing is that from the balancing point of view the linearity of the indicator device is irrelevant as the only thing we want to know is if the output voltage is zero or not. The third outcome is if the indicator and the source can be exchanged.

An important parameter of the bridge circuit, shown in Fig. 2.39, is the *measuring sensitivity* (S). This parameter describes the change in output voltage (U_{out}) with regard to the source voltage (U_0) caused by an imbalance of the bridge, i.e. the change of Z_1 impedance with 1% of its value. According to this definition the complex value of the sensitivity is given in (2.52).

$$S = \frac{U_{out}/U_0}{\Delta Z_1/Z_1} \tag{2.52}$$

Let's calculate the sensitivity as the function of bridge circuit elements in Fig. 2.39. If Z_1 impedance changes its value with ΔZ_1.

Then U_2 voltage across Z_2 will also be changed by ΔU_2. For the easy calculation, i.e. to be able to calculate with unloaded voltage dividers, let's suppose an infinite high load (indicator) impedance. That is a good supposition for voltmeters. The output voltage of the bridge is given in (2.53).

$$U_{out} = -\Delta U_2 = U_0\left(\frac{Z_2}{Z_1+Z_2} - \frac{Z_2}{Z_1+\Delta Z_1+Z_2}\right) \tag{2.53}$$

For small relative changes in Z_1 impedance the linear approximation can be applied as written in (2.54).

$$U_{out} = -\frac{\partial U_2}{\partial Z_1}\Delta Z_1 \tag{2.54}$$

Substituting U_2 with the applied (unloaded) voltage division, we have (2.55).

$$U_{out} = -\frac{\partial}{\partial Z_1}\left(U_0\frac{Z_2}{Z_1+Z_2}\right)\Delta Z_1 \tag{2.55}$$

Evaluating the differential calculus, we get the following.

$$U_{out} = U_0\frac{Z_2}{(Z_1+Z_2)^2}\Delta Z_1 \tag{2.56}$$

Dividing the numerator and denominator by Z_1^2 we have (2.57).

$$U_{out} = U_0\frac{\frac{Z_2}{Z_1}}{\left(1+\frac{Z_2}{Z_1}\right)^2}\frac{\Delta Z_1}{Z_1} \tag{2.57}$$

Introducing the *gear* of bridge (m) that is simply the ratio of Z_2/Z_1 ratio, (2.57) can be shortened as the following.

$$U_{out} = U_0 \frac{m}{(1+m)^2} \frac{\Delta Z_1}{Z_1} \tag{2.57}$$

Thus, the sensitivity of the bridge according to (2.52) and (2.57) can be calculated as given in (2.58).

$$S = \frac{U_{out}/U_0}{\Delta Z_1/Z_1} = \frac{m}{(1+m)^2} \tag{2.58}$$

After evaluating this result for the sensitivity of the bridge circuit we have some important outcomes.

The first thing that we have already seen is that the change of the source and indicator has no effect on the balance condition but… it causes a change in sensitivity as the Z_1/Z_2 gear can be different from Z_3/Z_4 gear.

We can also see from (2.58) that the sensitivity has its maximum when the absolute value of the gear is equal to one.

$$|m| = 1 \rightarrow S = S_{\max} \tag{2.59}$$

When m has a real value (resistive bridge) then S_{max} is 0.25 when m is simply imaginary the S_{max} is 0.5. Thus, the complex gear, i.e. applied reactive impedance elements in the bridge circuit, causes better sensitivity of the bridge.

2.3.2 Null Type (Balanced) Bridges

These circuits make use of a null-balance meter to compare two voltages; just like the laboratory balance scale compares two weights and indicates when they're equal. Bridge circuits can be used to measure all kinds of electrical values, like resistance (DC bridges) and impedance (AC bridges).

DC Impedance Bridges

The standard bridge circuit, often called a *Wheatstone bridge*, is shown in Fig. 2.41. It was invented by Samuel Hunter Christie and popularized by Charles Wheatstone and is mainly used for measuring resistance. It is constructed from four resistors, two of known values R_3 and R_4, one whose resistance is to be determined $R_1 = R_x$, and one which is variable and calibrated R_2. Two opposite vertices are connected to a source and a null-indicator device is connected across the other two nodes. The variable resistor is adjusted until the galvanometer reads zero.

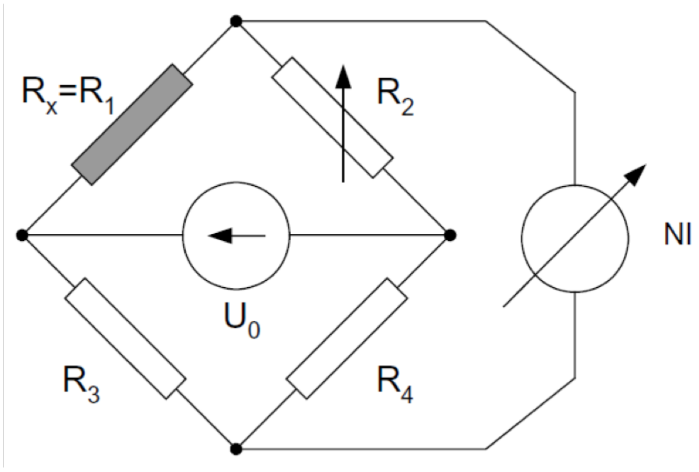


Figure 2.41 Wheatstone bridge

So, this bridge circuit can be used for resistance measurement according to the balance condition already discussed in this chapter. When $R_1 = R_x$ is the unknown resistance to be measured, from the balance condition of the bridge (2.48) the unknown resistance can be expressed as given in (2.60).

$$R_x = R_1 = R_2 \frac{R_3}{R_4} \tag{2.60}$$

According to this equation the measuring range of the resistance bridge can be set by the ratio of R_3/R_4 and the balance can be adjusted by R_2. The main advantage of this bridge is to be able to use a wide measuring range.

Unfortunately, the accuracy (or uncertainty) of the measurement depends on the accuracy (or uncertainty) of all three circuit elements. We have already seen that the uncertainty of the measurement calculated according to (1.34) gives the following uncertainty in this case.

$$\delta R_x = \sqrt{(\delta R_2)^2 + (\delta R_3)^2 + (\delta R_4)^2} \tag{2.60}$$

Additional uncertainties are caused by contact resistances, thermoelectric contact voltages, and the resistance of wires. The classical Wheatstone bridge, shown in Fig. 2.41 is used for resistance measurement between a few ohms and several mega-ohms.

When measuring small resistances, the three-wire connection or four-wire connection method can be applied to eliminate the effect of wire resistance that is comparable to the measured resistance.

The *three-wire connection* measuring method is presented in Fig. 2.42. If all three wires exhibit the same resistance (the same length) we can write the balance condition as the following.

$$R_4(R_x + r) = R_2(R_3 + r) \tag{2.61}$$

from which

$$R_x R_4 = R_2 R_3 + r(R_2 - R_4) \tag{2.62}$$

Thus, in the case of $R_2 = R_4$ the balance condition equation is the following.

$$R_x R_4 = R_2 R_3 \tag{2.63}$$

As we see in (2.63) the resistance (*r*) of the wire is eliminated.

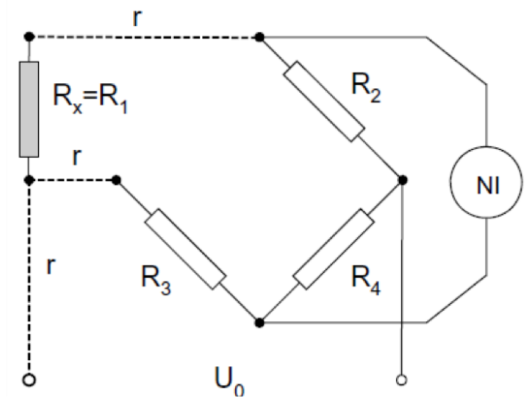


Figure 2.42 Three-wire connection resistance measurement

The enhanced solution is the *four-wire connection* that is an interesting variation of the Wheatstone bridge. It is the so-called *Thomson - Kelvin Double Bridge*, used for measuring very low resistances (typically less than 1/10 of an ohm). Its schematic diagram is shown in Fig. 2.43.

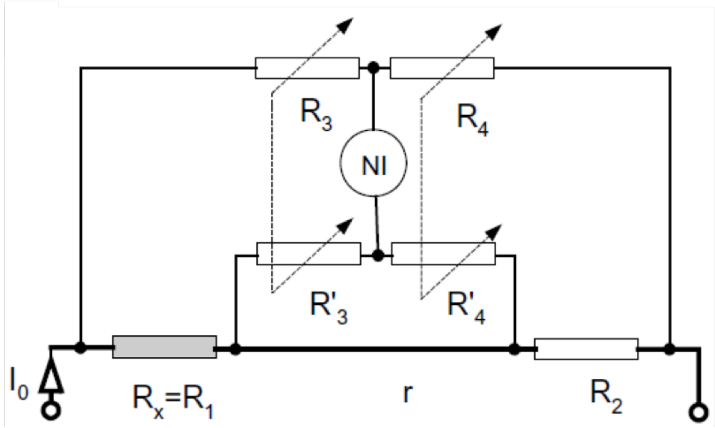


Figure 2.43 Thomson – Kelvin double bridge

The condition of the balance for this bridge is given in (2.64).

$$R_x = R_2 \frac{R_3}{R_4} + r \frac{R_3 R_4' - R_3' R_4}{R_4 (R_3' + R_4')} \tag{2.64}$$

The second term in the equation (2.64) is negligible if the following condition is fulfilled

$$R_3 R_4' = R_3' R_4 \rightarrow \frac{R_3}{R_3'} = \frac{R_4}{R_4'} \tag{2.65}$$

The condition (2.65) is relatively easy to achieve by mechanical coupling of the resistors *R*3 /*R*3' and *R*4/*R*4'. In such cases, the condition for balance of the Kelvin Bridge (2.64) is the same as for the classical Wheatstone bridge (2.60).

The Kelvin Bridge enables measurement of the resistances in the range of a few hundred micro-ohms.

AC Impedance Bridges

We have already seen in (2.50) and (2.51) that the balance of an *AC* bridge circuit has two conditions that need to be fulfilled: magnitude and phase conditions. This means that in order to balance such bridge circuits two independent adjusting elements are necessary. The process of balancing is therefore more complicated than in a *DC* bridge circuit.

In the case of *AC* bridge circuits another problem appears – it is difficult to eliminate influence of stray and earth capacitances. For that reason, it is necessary to *shield* all the elements in the *AC* bridge circuits. Another method is to use the *Wagner earth* (*Wagner ground*) with additional impedance elements to the bridge. These methods, however, are not discussed in detail here.

A huge number of various AC bridge circuits were designed and developed, like *Maxwell, Wien, Schering, Hay, Owen, Anderson, de Sauty,* etc. Moreover, all these bridges exist in various mutations and modifications. Historically the oldest and most known are the Wien Bridge and Maxwell Bridge circuits.

The circuit of the Wien Bridge is shown in Fig. 2.44. It can be used for capacitance measurement.

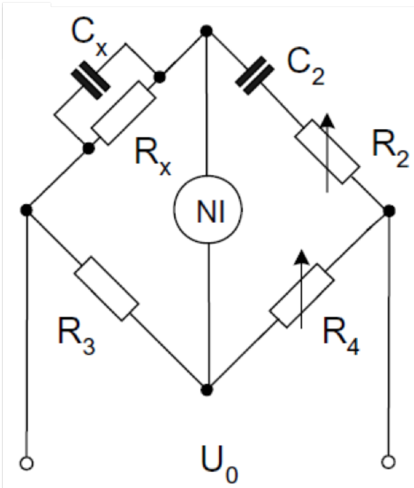


Figure 2.44 The Wien Bridge

From the balance condition we can write (2.66) and (2.67) that enables us to determine the value of capacitance and the value of loss resistance.

$$C_x = \frac{C_2 R_4}{R_3(1 + \omega^2 C_2^2 R_2^2)} \tag{2.66}$$

$$R_x = \frac{R_3(1 + \omega^2 C_2^2 R_2^2)}{\omega^2 R_2 R_4 C_2^2} \tag{2.67}$$

The equations show the frequency beside the bridge circuit elements which cause additional uncertainty in measurement. For this reason, this method is *seldom used in measurement*. A variant of this bridge, the so-called *Wien-Robinson Bridge* is used in oscillators as frequency determining circuit. The frequency can be set by a bridge as given in (2.68).

$$\omega^2 = \frac{1}{R_X C_X R_2 C_2} \tag{2.68}$$

A variant of the Maxwell Bridge is the *Maxwell – Wien Bridge* that is used for L_x inductance measurement when the loss of inductance is calculated as a series equivalent resistance R_x. The bridge circuit is shown in Fig. 2.45.

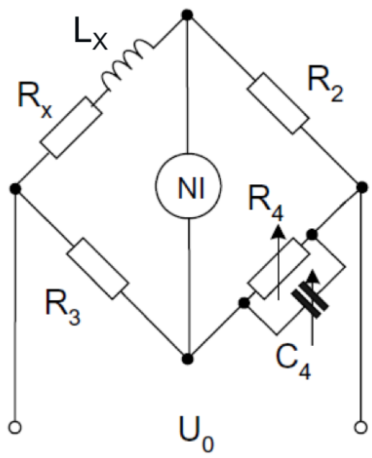


Figure 2.45 Maxwell-Wien Bridge

Balance conditions for this bridge are given in (2.69) and (2.70).

$$L_X = R_2 R_3 C_4 \tag{2.69}$$

$$R_X = R_2 \frac{R_3}{R_4} \tag{2.70}$$

For practical inductors the so called quality factor Q is used instead of the loss resistance R_x that can be calculated as follows.

$$Q = \frac{\omega L_x}{R_x} \tag{2.71}$$

For the capacitance measurement, the Schering Bridge is used widely. The circuit diagram is presented in Fig. 2.46. The measured capacitor model is a series C_x - R_x. The Schering Bridge is used in practice mainly for cable testing high voltage systems. From the balance condition equations, we can express C_x and R_x as given in (2.72) and (2.73).

$$C_X = C_3 \frac{R_4}{R_2} \tag{2.72}$$

$$R_X = R_2 \frac{C_4}{C_3} \tag{2.73}$$

For practical capacitors the *loss tangent* (dielectric loss) is used instead of loss resistance R_x that can be calculated as given in (2.74).

$$\tan\delta = \omega R_X C_X \rightarrow \tan\delta = \omega R_4 C_4 \tag{2.74}$$

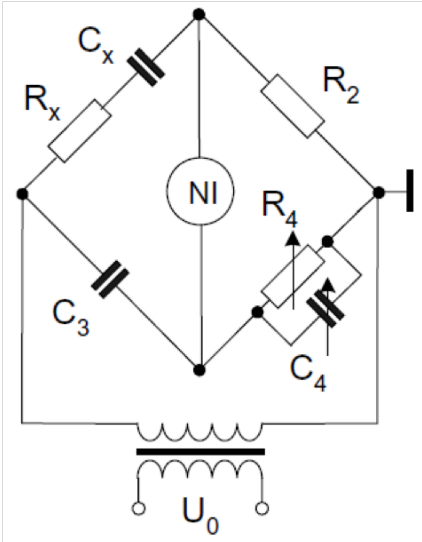


Figure 2.46 The Schering Bridge

Transformer Bridges

In bridge circuits the change of impedance during the balancing process results in the change of the voltage drop across this impedance. Thus, the bridge can be balanced by inserting an additional source of voltage instead of changing the value of the impedance. For example, in the circuit presented in Fig. 2.47 the condition of the balance is

$$R_x = R_w \frac{E_1}{E_w} \tag{2.75}$$

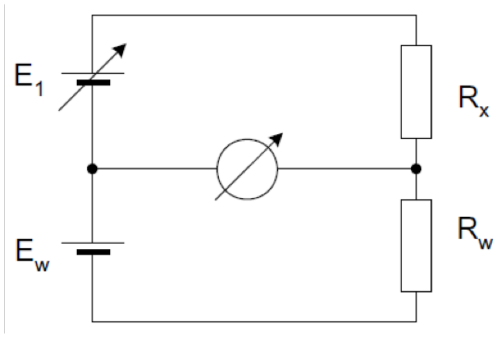


Figure 2.47 Source balanced bridge

Similarly, for the bridge circuit presented in Fig. 2.48 the condition of the balance can be written as in (2.76).

$$\frac{R_x}{R_w} = \frac{E_1}{E_w} = \frac{n_1}{n_2} \tag{2.76}$$

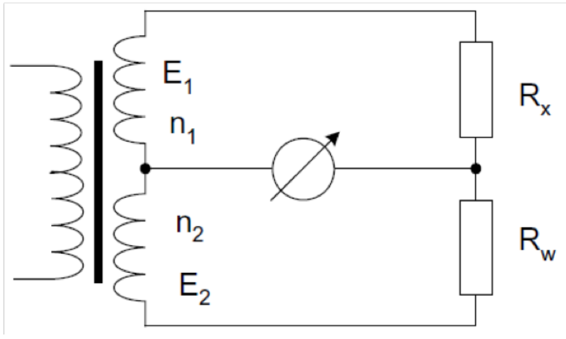


Figure 2.48 Sources substituted by transformer

The bridge can be balanced not only by the change of impedance or voltage but also by the change of the number of turns *n*. This is very convenient, because the number of turns can be precisely adjusted. Especially in the case of digital bridge circuits it is much easier to adjust the windings than to change the resistors or capacitors.

Fig. 2.49 presents the example of the transformer bridge circuit with two transformers. The output transformer acts in this circuit as the current comparator – the null indicator points to zero, when the resultant flux in the transformer is also equal to zero.

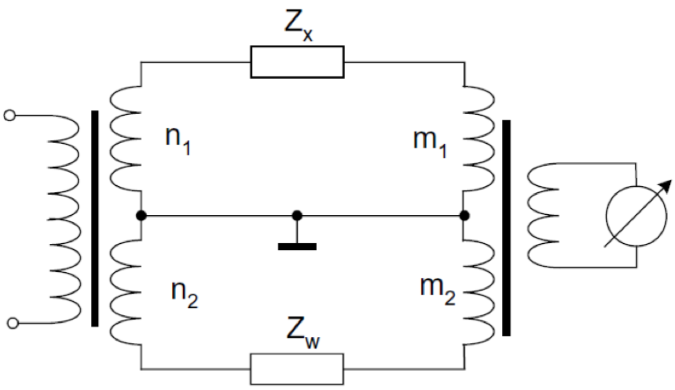


Figure 2.49 Transformer bridge with current comparator

The condition of the balance of this circuit is given in (2.77).

$$\frac{Z_x}{Z_w} = \frac{n_1}{n_2}\frac{m_1}{m_2} \tag{2.77}$$

Transformer bridge circuits have several advantages in comparison with impedance bridge circuits. First of all, as was mentioned earlier, balancing is possible by changing the number of turns. In transformer bridges the parasitic capacitances shunt the transformer turns and do not influence the conditions of the balance. Also, the sensitivity of transformer bridges are significantly better than impedance bridges. In order to make use of these advantages it is necessary to construct these transformers very precisely, with minimal stray fields. This is why transformer bridges are usually more expensive than classic circuits without transformer coupling.

2.3.3 Deflection Type (Unbalanced) Bridges

Unbalanced bridge circuits are used as transducers for converting the change of the resistance (and generally impedance) into the output voltage, given in (2.78).

$$U_{out} = SU_0 \frac{\Delta R_x}{R_{x0}} = SU_0 \varepsilon \qquad (2.78)$$

In (2.78) S is the sensitivity of the transducer and ε is the relative change of the resistance according to (2.79). R_{XO} is the resistance in the balanced state.

$$R_x = R_{x0} \pm \Delta R_x = R_{x0}(1 \pm \varepsilon) \qquad (2.79)$$

There are two kinds of symmetry in the design of unbalanced bridge circuits. Fig. 2.50 demonstrates the symmetry in diagonal output and Fig. 2.51 gives an example of symmetry in diagonal source.

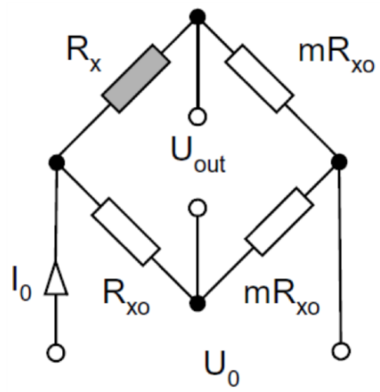


Figure 2.50 Symmetry in output diagonal

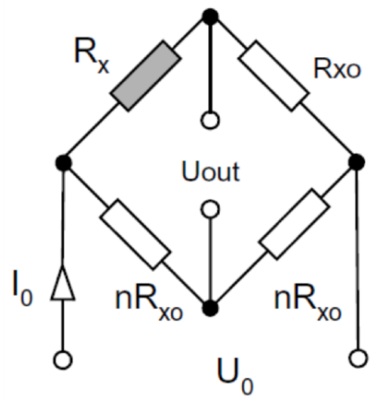


Figure 2.51 symmetry in source diagonal

It is important to note that the transfer functions of these circuits are nonlinear, and nonlinearity depends on the value of the resistance (coefficients *m* or *n*).

There are several methods for linearization, usually by the application of operation amplifiers. If a bridge is automatically balanced it uses a relatively small, i.e. linear part of the transfer characteristic. Fig. 2.52 demonstrates two examples for auto-balancing and therefore, for linearization. Because the output signal of the bridge is very small an additional amplifier on the output is necessary as seen in the second solution in Fig. 2.52.

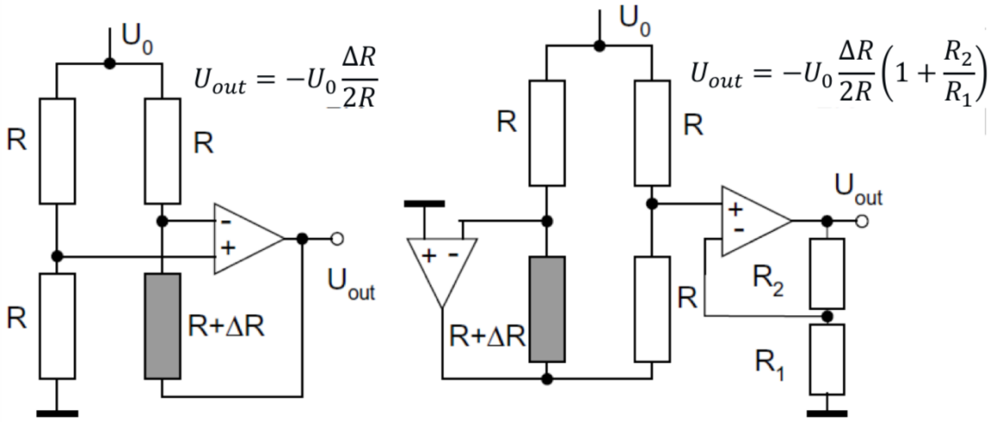


Figure 2.52 Linearization by auto-balancing

2.3.4 Alternatives to Bridge Circuits

The Wheatstone bridge is an old solution that has been used since 1843. In some cases, it can be successfully substituted by a differential amplifier. Circuits with differential amplifiers are presented in Fig. 2.53.

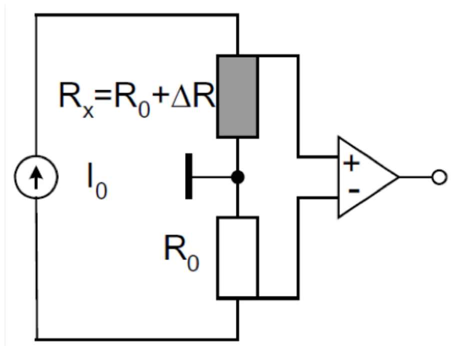


Figure 2.53 Bridge alternatives with applied differential amplifiers

The circuit, shown in Fig. 2.53, provides a similar performance to the bridge circuit, like compensation of the offset voltage (zero output signal for *R = 0*), linear conversion, compensation of interferences, such as changes of external temperature. A similar circuit was patented in 1994 by NASA (Anderson 1994, Anderson 1998) and today is sometimes called the *Anderson Loop*. Two examples of the Anderson loop are presented in Fig. 2.53.

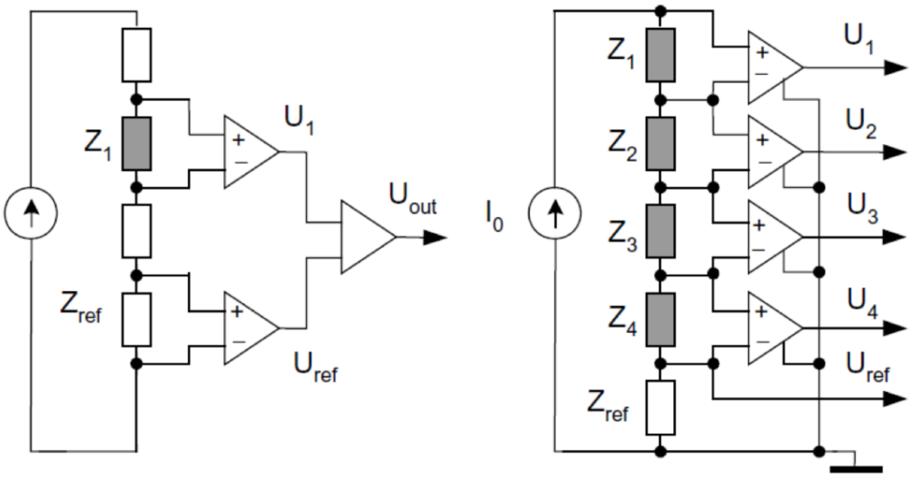


Figure 2.54 Anderson loops

The first solution applies a single sensor and the second one is established with four sensors.

The output voltage with one sensor solution is given in (2.80).

$$U_{out} = I_0 \cdot \Delta Z \tag{2.80}$$

The output signal of each sensor in the second solution can be determined as the difference between output voltage and reference voltage, as given in (2.81).

$$U_1 - U_{ref} = I_0 \cdot \Delta Z \tag{2.81}$$

It is also possible to determine the difference, like in a bridge circuit, the output signals of the sensors, for example

$$U_1 - U_2 = I_0 \cdot (\Delta Z_1 - \Delta Z_2) \tag{2.82}$$

In the case where several sensors are used the Anderson loop can be a good alternative for the bridge circuit.

2.4 Compensation Method and Comparators

It is possible to determine compensation by means of mutual neutralization of two voltages. The devices utilizing the idea of compensation of two voltages are called comparators.

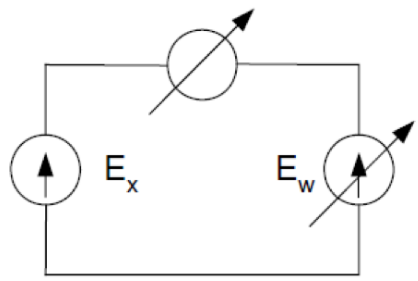


Figure 2.55 Voltage compensation

An example for voltage compensation is presented in Fig. 2.55. The balance condition is

$$E_X = E_W \tag{2.83}$$

Other values can also be compensated, for example the compensation of two currents is demonstrated in Fig. 2.56. The balance condition is given in (2.84).

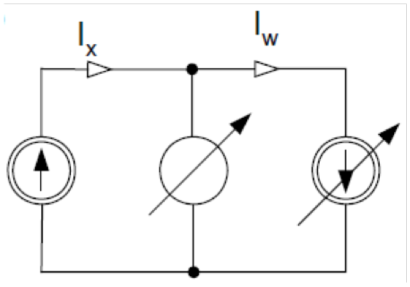


Figure 2.56 Current compensation

$$I_X = I_W \tag{2.84}$$

The compensation of two magnetic fluxes is illustrated in Fig. 2.57 with the balance condition in (2.85).

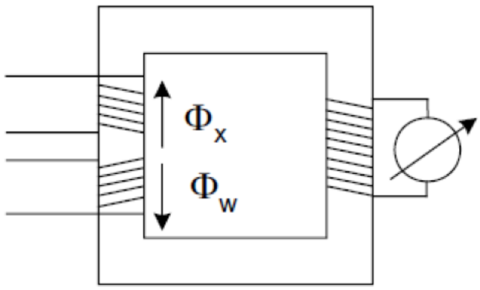


Figure 2.57 Magnetic flux compensation

$$\Phi_X = \Phi_W \tag{2.85}$$

Fig. 2.58 presents the typical design of the compensating transducer for the transfer characteristic of (2.86).

$$I_{out} = f(U_x) = R_S \cdot U_x \tag{2.85}$$

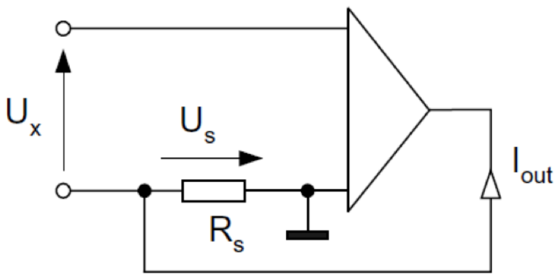


Figure 2.58 Voltage – Current transducer

The transducer presented in Fig. 2.58 has many important advantages of the compensation principle – very large input resistance (practically infinitively large) and excellent accuracy. With correct design of the transducer, the uncertainty of the signal processing depends only on the accuracy of the standard resistor R_S.

Finally, a simple but practically useful example of the current comparator used as the car lighting system tester is presented in Fig. 2.59. If both bulbs in the car lighting system work, then both currents are the same and the resultant magnetic flux in the yoke is equal to zero. The damage of one of the bulbs causes the unbalance of fluxes, which is detected by the magnetic field sensor placed in the gap.

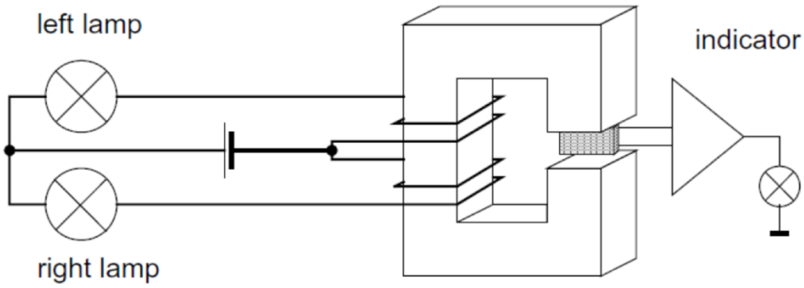


Figure 2.59 Car lighting system tester

3. Electrical Measurement Solutions

3.1 DC Power and Resistance Measurement

3.1.1 Ammeter-Voltmeter Method

This is the simplest method of measuring resistance and electric power. It uses one ammeter to measure current, I and one voltmeter to measure voltage, V and we get the value of resistance according to Ohm's law and electric power according to Joule's law, given in (3.1).

$$R = \frac{V}{I}, \qquad P = V \cdot I \tag{3.1}$$

Now we can have two possible connections of ammeter and voltmeter, shown in the figure below.

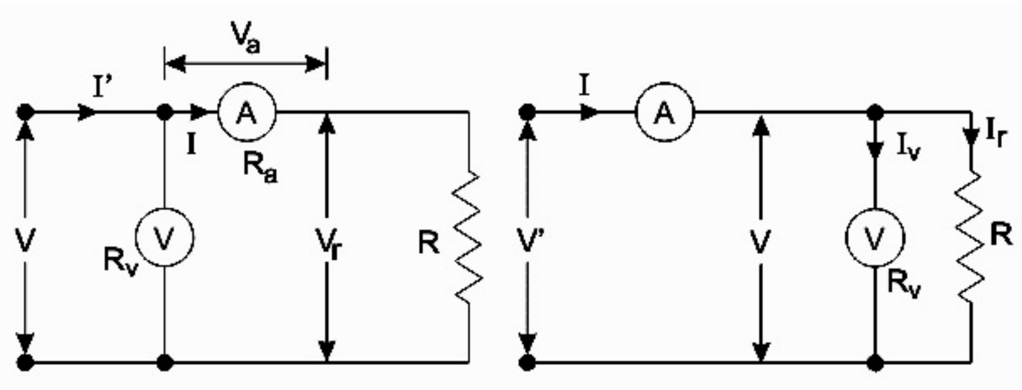


Figure 3.1 Two possible connections of the ammeter and voltmeter

On left side of Fig 3.1 the voltmeter measures voltage drops across the ammeter and the unknown resistance allowing us to calculate the resistance as in (3.2).

$$R_1 = \frac{V_a + V_r}{I} = \frac{I \cdot R_a + I \cdot R}{I} = R_a + R \tag{3.2}$$

The relative error of measurement in this case is given as

$$h_1 = \frac{R_1 - R}{R} = \frac{R_a}{R} \tag{3.3}$$

For connection on right side of Fig. 3.1, the ammeter measures the sum of the current through the voltmeter and resistance, hence

$$R_2 = \frac{V}{I_v + I_r} = \frac{V}{\frac{V}{R_v} + \frac{V}{R}} = \frac{R}{1 + \frac{R}{R_v}} \tag{3.4}$$

The relative error in this case is,

$$h_2 = \frac{R_2 - R}{R} \approx -\frac{R}{R_v} \tag{3.5}$$

It can be observed that the relative error is zero for R_a = 0 in first case and $R_v = \infty$ in the second case. Now the question arises of which connection should be used in which case. To find out this we equate both the errors

$$\frac{R_a}{R} = \frac{R}{R_v} \to R = \sqrt{R_a \cdot R_v} \quad (3.6)$$

Hence for resistances greater than that given by (3.6) we use the first method and for less than that we use the second method. This consideration for choosing measurement method is valid for both the resistance and power determination by measuring the current and the voltage.

3.1.2 Kelvin (Four-Wire) Method

This method is designed for measuring the resistance of some component located a significant distance away from the measuring device. Such a scenario would be problematic, because we can measure *all* resistance in the circuit loop, which includes the resistance of the wires (R_{wire}) connecting the measuring device to the component being measured ($R_{subject}$).

Usually, wire resistance is very small, only a few ohms over hundreds of meters, depending primarily on the gauge (size) of the wire, but if the connecting wires are very long, and/or the component to be measured has a very low resistance anyway, the measurement error introduced by wire resistance can be substantial.

In a situation like this we can determine the resistance of the subject component by Ohm's law if we measure the current going through it and the voltage dropped across it, as it is shown in Fig. 3.2.

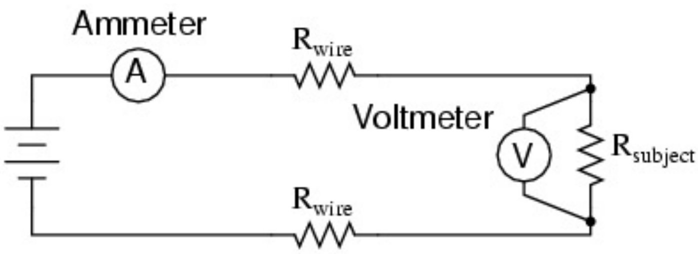


Figure 3.2 The measuring principle

Current is the same at all points in the circuit, because it is a series loop. Because we are only measuring voltage dropped across the subject resistance (and not the wires' resistances), though, the calculated resistance is indicative of the subject component's resistance ($R_{subject}$) alone.

Our goal was to measure this subject resistance *from a distance*, so our voltmeter must be located near the ammeter, connected across the subject resistance by another pair of wires containing resistance, as presented in Fig. 3.3.

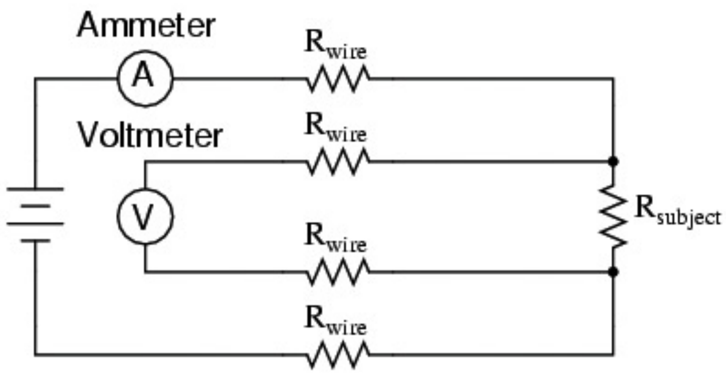


Figure 3.3 Voltmeter connected from a distance

At first it appears that we have lost any advantage of measuring resistance in this way, because the voltmeter now has to measure voltage through a long pair of (resistive) wires, introducing stray resistance back into the measuring circuit again. However, upon closer inspection it is seen that nothing is lost at all, because the voltmeter's wires carry miniscule current. Thus, those long lengths of wire connecting the voltmeter across the subject resistance will drop insignificant amounts of voltage, resulting in a voltmeter indication that is very nearly the same as if it were connected directly across the subject resistance:

Any voltage dropped across the main current-carrying wires will not be measured by the voltmeter, and so do not factor into the resistance calculation at all. Measurement accuracy may be improved even further if the voltmeter's current is kept to a minimum, either by using a high-quality (low full-scale current) movement and/or a potentiometric (null-balance) system.

This method of measurement, which avoids errors caused by wire resistance, is called the *Kelvin*, or *four-wire* method. Special connecting clips called *Kelvin clips* are made to facilitate this kind of connection across a subject resistance.

In regular, 'alligator' style clips, both halves of the jaw are electrically common to each other, usually joined at the hinge point. In Kelvin clips, however, the jaw halves are insulated from each other at the hinge point, only contacting at the tips where they clasp the wire or terminal of the subject being measured. Thus, current through the 'current' ('C') jaw halves does not go through the 'potential' ('P'), or 'voltage' jaw halves, and will not create any error-inducing voltage drop along their length. The connection of Kelvin clips is demonstrated in Fig. 3.4.

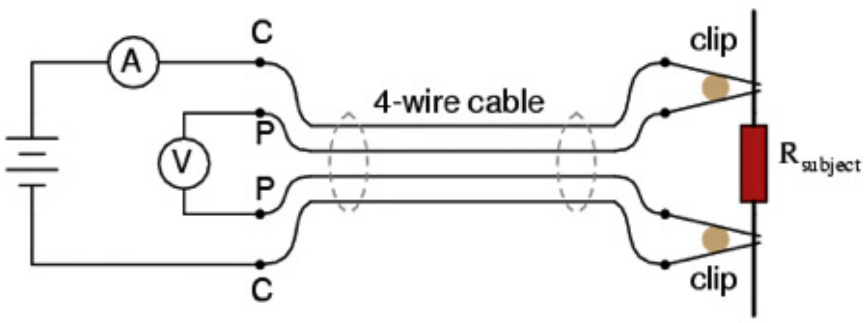


Figure 3.4 The connection of Kelvin clips

The same principle of using different contact points for current conduction and voltage measurement is used in precision shunt resistors for measuring large currents. Shunt resistors function as current measurement devices by dropping a precise amount of voltage for every amp of current through them, the voltage drop being measured by a voltmeter. In this sense, a precision shunt resistor ‘converts’ a current value into a proportional voltage value. Thus, current may be accurately measured by measuring voltage dropped across the shunt.

Current measurement using a shunt resistor and voltmeter is particularly well-suited for applications involving particularly large magnitudes of current. In such applications, the shunt resistor’s resistance will likely be in the order of milliohms or micro-ohms, so that only a modest amount of voltage will be dropped at full current. Resistance this low is comparable to wire connection resistance, which means voltage measured across such a shunt must be done so in such a way as to avoid detecting voltage dropped across the current-carrying wire connections, lest huge measurement errors be induced. In order that the voltmeter measures only the voltage dropped by the shunt resistance itself, without any stray voltages originating from wire or connection resistance, shunts are usually equipped with *four* connection terminals, as shown in Fig. 3.5.

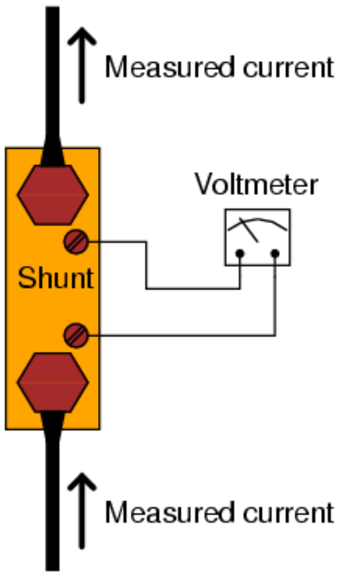


Figure 3.5 Current measurement with four-wire shunt

In metrological (*metrology = ‘the science of measurement’*) applications, where accuracy is of paramount importance, highly precise ‘standard’ resistors are also equipped with four terminals: two for carrying the measured current, and two for conveying the resistor’s voltage drop to the voltmeter. This way, the voltmeter only measures voltage dropped across the precision resistance itself, without any stray voltages dropped across current-carrying wires or wire-to-terminal connection resistances.

It should be noted that resistance measurement using *both* an ammeter and a voltmeter is subject to compound error. Because the accuracy of both instruments factors in to the final result, the overall measurement accuracy may be worse than either instrument considered alone. For instance, if the ammeter is accurate to +/- 1% and the voltmeter is also accurate to +/- 1%, any measurement dependent on the indications of both instruments may be inaccurate by as much as +/- 2%.

Greater accuracy may be obtained by replacing the ammeter with a standard resistor, used as a current-measuring shunt. There will still be compound error between the standard resistor and the voltmeter used to measure voltage drop, but this will be less than with a voltmeter + ammeter arrangement because typical standard resistor accuracy far exceeds typical ammeter accuracy. Using Kelvin clips to make connection with the subject resistance, the circuit is presented in Fig. 3.6.

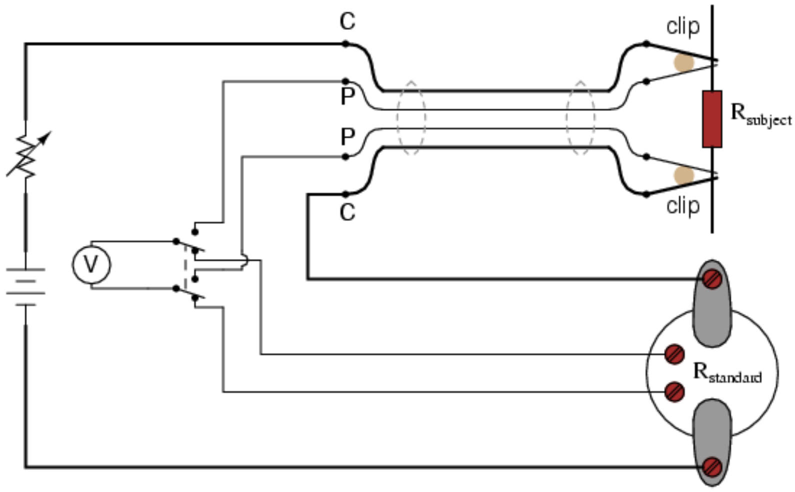


Figure 3.6 Measurement with a single voltmeter

All current-carrying wires in the above circuit are shown in 'bold', to easily distinguish them from wires connecting the voltmeter across both resistances ($R_{subject}$ and $R_{standard}$).

Practical Application

The Kelvin measurement can be a practical tool for finding poor connections or unexpected resistance in an electrical circuit. Connect a DC power supply to the circuit and adjust the power supply so that it supplies a constant current to the circuit as shown in Fig. 3.7.

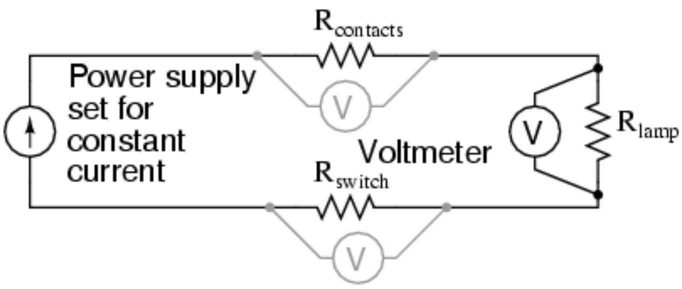


Figure 3.7 Finding unexpected resistances in a circuit

With a digital multi-meter set to measure DC voltage, measure the voltage drop across various points in the circuit. If you know the wire size, you can estimate the voltage drop you should see and compare this to the voltage drop you measure. This can be a quick and effective method of finding poor connections in wiring exposed to the elements, such as in lighting circuits.

3.2 AC Power and Impedance Measurement

3.2.1 Instrument Transformers

Instrument transformers are high accuracy electrical devices used to isolate or transform voltage or current levels. The most common usage of instrument transformers is to operate instruments or metering from high voltage or high current circuits, safely isolating secondary control circuitry from the high voltages or currents. The primary winding of the transformer is connected to the high voltage or high current circuit, and the meter or relay is connected to the secondary circuit. We distinguish between *current transformers* and *voltage transformers* depending on what we aim to test.

3.2.1.1 Current Transformer

Current transformers (CT) are a series connected type of instrument transformer. They are designed to present negligible load to the supply being measured and have an accurate current ratio and phase relationship to enable accurate secondary connected metering.

The CT is typically described by its *current ratio* from primary to secondary. A 1000:5 CT will provide an output current of 5 amperes when 1000 amperes are flowing through its primary winding. Standard secondary current ratings are 5 amperes or 1 ampere, compatible with standard measuring instruments. It is used to step down current for metering purposes for the safety of the equipment as well as operator.

A *current clamp* uses a current transformer with a split core that can be easily wrapped around a conductor in a circuit. This is a common method used in portable current measuring instruments, but permanent installations use more economical types of current transformer. Specially constructed wideband CTs are also used, usually with an oscilloscope, for measuring high frequency waveforms or pulsed currents within pulsed power systems. A photo of a practical current clamp is shown in Fig. 3.8.

Figure 3.8 The current clamp

3.2.1.2. Voltage Transformer or Potential Transformer

Voltage transformers (VT), also called potential transformers (PT), are a parallel connected type of instrument transformer. They are designed to present negligible load to the supply being measured and have an accurate voltage ratio and phase relationship to enable accurate secondary connected metering.

The PT is typically described by its *voltage ratio* from primary to secondary. A 600:120 PT will provide an output voltage of 120 V when voltage of 600 V is impressed across its primary winding. Standard secondary voltage ratings are compatible with standard measuring instruments.

3.2.1.3 Practical Application Example of Instrument Transformers

Fig. 3.9 demonstrates how to use instrument transformers to operate measuring instruments from high voltage or high current circuits. The load impedance (*Z*), supplied from a high voltage system, is measured with standard low voltage meters. The voltage terminals of the voltmeter, power factor ($\mathrm{pf} = \cos\varphi$) meter and the wattmeter are connected in parallel to the secondary coil (*a* – *b* terminals) of the voltage transformer (PT). The current terminals of the ammeter, pf meter and the wattmeter are connected in series to the secondary coil (*k* - *l* terminals) of the current transformer (CT) device.

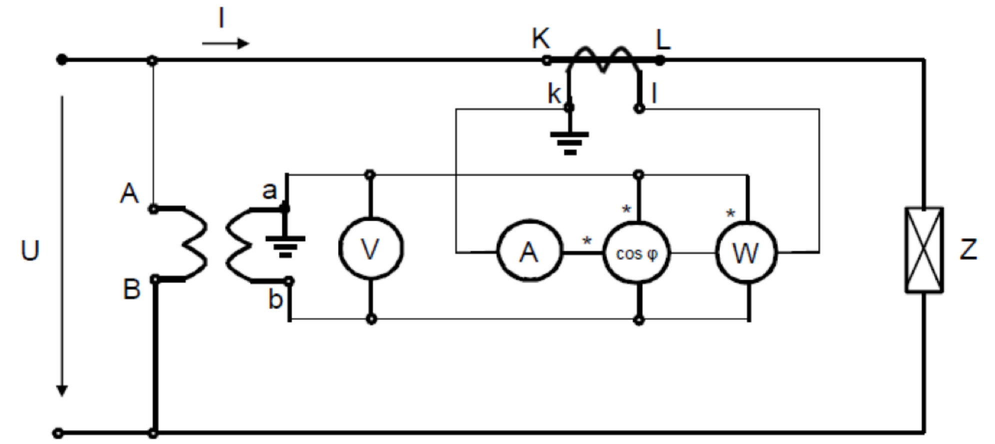


Figure 3.9 Application of instrument transformers

3.2.2 Three-Voltmeter Method

Usually, the wattmeters are used for measurement of AC power, but in some cases, it is not possible to use wattmeters, for example, because of their incorrect readings. In such cases the three-voltmeter (or three-ammeter) method is used for AC power measurement as demonstrated in Fig. 3.10.

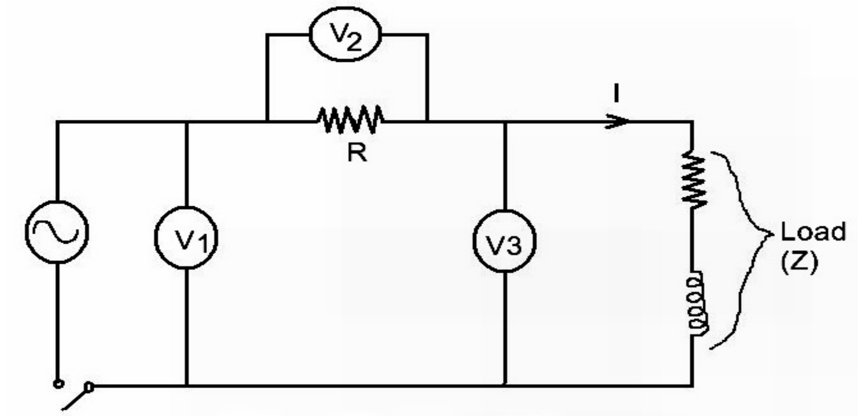


Figure 3.10 Tree-voltmeter method of measurement

V_1, V_2 and V_3 are the three voltmeters and *R* is a non-inductive resistance connected in series with the load. Supposing the inductive load, the phasor diagram of the V_1, V_2 and V_3 voltages is given in Fig. 3.11. In case of a capacitive load, the phasor diagram is 'similar' but the phase angle is negative.

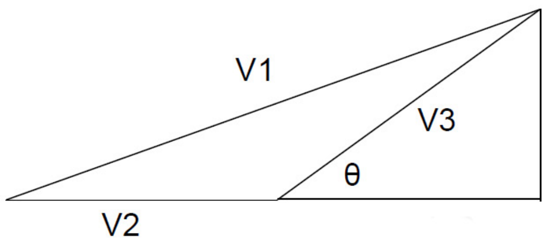


Figure 3.11 Voltage phasor diagram

Following the law of cosine applied to the phasor diagram we can write the following equation.

$$V_1^2 = V_2^2 + V_3^2 + 2 \cdot V_2 \cdot V_3 \cdot \cos\theta \tag{3.7}$$

The V_2 voltage is measured across the resistor, thus according to the Ohm's law,

$$V_1^2 = V_2^2 + V_3^2 + 2 \cdot (I \cdot R) \cdot V_3 \cdot \cos\theta \tag{3.8}$$

Regrouping the right term of the equation,

$$V_1^2 = V_2^2 + V_3^2 + 2 \cdot R \cdot (I \cdot V_3 \cdot \cos\theta) \tag{3.9}$$

In the brackets we see the P electric power of the load, thus,

$$V_1^2 = V_2^2 + V_3^2 + 2 \cdot R \cdot P \tag{3.10}$$

Or, substituting the power from (3.10), we get the following:

$$P = \frac{V_1^2 - V_2^2 - V_3^2}{2 \cdot R} \tag{3.11}$$

The power factor of the circuit can also be obtained from (3.7) as the following.:

$$pf = \cos\theta = \frac{V_1^2 - V_2^2 - V_3^2}{2 \cdot V_2 \cdot V_3} \tag{3.12}$$

From this measurement the (complex) load impedance can also be determined. Because of the common current flowing through the series elements,

$$V_2 = I \cdot R\,, \qquad V_3 = I \cdot Z \tag{3.13}$$

from which, the absolute value of the load impedance can be determined as in (3.14).:

$$\frac{V_2}{V_3} = \frac{I \cdot R}{I \cdot Z} \rightarrow Z = R\frac{V_3}{V_2} \tag{3.14}$$

The phase angle of the load impedance is given by the power factor

$$pf = \cos\theta \rightarrow \theta = cos^{-1}\frac{V_1^2 - V_2^2 - V_3^2}{2 \cdot V_2 \cdot V_3} \tag{3.15}$$

Thus, the series algebraic equivalent of the load impedance is

$$Z = (Z\cos\theta) + j(Z\sin\theta) \tag{3.16}$$

We assumed in this calculation that the current in the resistor R is the same as the load current. This assumption is acceptable when applying good quality voltmeters.

However, there are some disadvantages when using the three-voltmeter method. These include:

- when providing the same voltage V_3, across the load as its operational voltage, the supply voltage has to be higher than normal because the additional resistance *R*, is connected in series with the load *Z*.
- Even small errors in measuring voltages may cause serious errors in the value of power determined by this method, due to the subtractions in the calculation.

4. Sensor Theory

4.1 Positioning of Sensors

When positioning sensors in measurement, first of all, we have to recall what measurement is in general. When we measure something in a system (or in a process), we *collect information* about one or more system (or process) characteristics. On the other hand, measurement is always *comparison to etalon or a 'benchmark'*. This comparison can be *direct or indirect* according to the measurement method, i.e. we have the physical etalon in the measurement or we have a calibrated device that was originally calibrated to an etalon.

The general model of the measurement process using an electrical measuring device, is shown in Fig. 4.1.

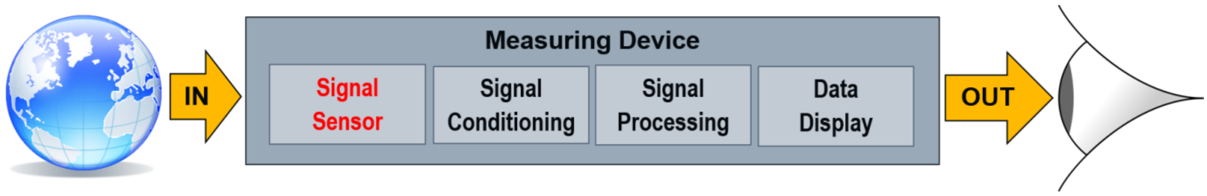


Figure 4.1 General model of electrical measurement process

With the measuring device, the first building block is a so-called signal sensor or sensor that converts a specific kind of physical parameter of the observed reality to an electrical signal. The second block is usually a signal conditioning circuit that is responsible for the appropriate gain, required noise reduction, wave form conditioning, etc. The next block is the signal processor block that converts the measured analogue signal to digital data, if necessary. This allows easier and improved post processing.

Fig. 4.2 gives an example for temperature measurement, based on the general measurement model.

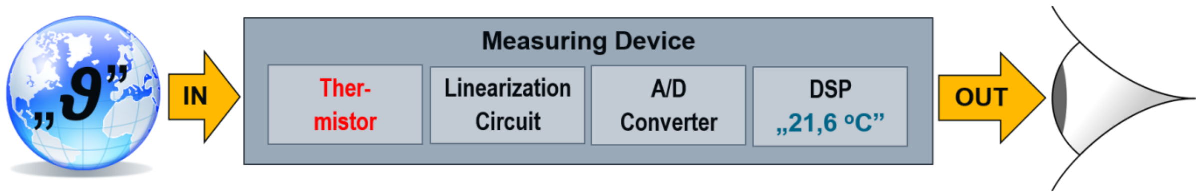


Figure 4.2 Temperature meter example

In this example the 'observed reality parameter' is the ambient temperature. The sensor element is a thermistor that converts the temperature to an electrical signal. Because this conversion has a linearity error, we need a linearization circuit to correct this measuring error. For display (or for data acquisition) we need a digital value of the measured temperature, that's why we apply AD conversion for the analogue electrical signal. This digital data can be either displayed, using a digital display or can be stored, and uploaded for later processing.

As highlighted in previous example, a sensor is a device that converts any physical parameter of an observed system, i.e. temperature, position, force, light intensity, etc., to an electrical signal i.e. to voltage or current. So, a sensor is a transducer that converts one form of energy to another. From this energy approach point of view, we distinguish between *sensors* and *actuators*.

- **Sensor** is a transducer that converts a physical parameter to an electrical signal.
- **Actuator** is a transducer that converts an electrical (control) signal to a physical output.

For example, in the acoustic 'world' a microphone, shown in Fig. 4.3(a) is a sensor, converting mechanical energy of SPL (sound pressure level) to electrical energy, usually to voltage. A loudspeaker is an actuator as it converts electrical energy to mechanical energy (SPL).

Figure 4.3 Microphone (a) and loudspeaker (b) as different transducers

In nature we have different kind of energies such as mechanical, thermal, chemical, magnetic, electrical, … and we know *'If we have a lot of things, they have to be systematized ...'*

The easiest, and therefore the best way for systematization is the energy approach. Introducing the following parameter pairs in nature we can calculate power in general as in (4.1)

$$p(t) = \sum_j f_j(t) \cdot h_j^*(t) \tag{4.1}$$

where f_j is the generalized force, h_j is the generalized charge and h_j^* is the generalized current. The time derivative of a variable is signed for short with an 'Asterix' in the top index for convenience as shown in (4.2)

$$(\quad)^* = \frac{d(\quad)}{dt} \tag{4.2}$$

We can interpret (4.1) for different types of energy in nature with the following results.

Table 4.1 Power in nature with derivative pairs

System	f_j	h_j^*
Electrical	Voltage (U)	Current (Q*→I)
Mech. (translation)	Force (F)	Velocity (x*→v)
Mech. (rotation)	Torque (M)	Angular velocity ($\boldsymbol{\varphi}$*→$\boldsymbol{\omega}$)
Flow	Pressure (P)	Volume current (V*→q)
Thermal	Temperature ($\boldsymbol{\vartheta}$)	Entropy current (S*)

4.1.1 Two-Port Sensor Models

A sensor as a generalized energy transducer can be interpreted as a two-port model, shown in Fig. 4.4.

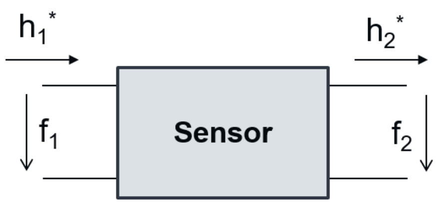


Figure 4.4 Two-port model of sensor

According to the two-port model we distinguish four types of transducers, as shown in Fig. 4.5.

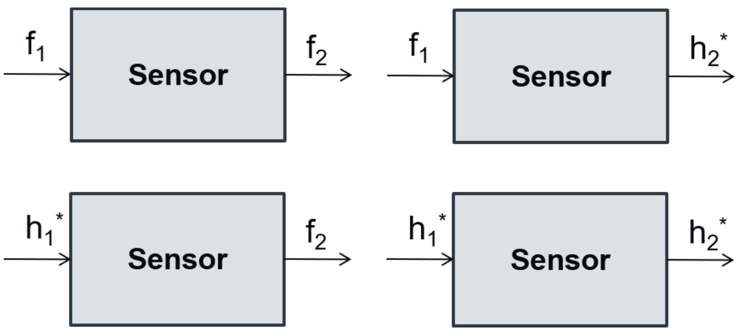


Figure 4.5 Force-force, force-current, current-force, current-current, transducers

4.1.2 Sensor Classification

According to operational principles we can distinguish two sensor groups. Sensors are *passive* if they need external energy for operation. Sensors, providing output energy without any external energy source, are called *active*. In sensor classification we can use the Miller-index - analogue to crystallography – as we can introduce 3 energy parameters (*x, y, z*) with the following meanings: x = input energy, y = output energy, z = external (auxiliary) energy

The six types of natural energy, with their abbreviations for using them in the Miller-index, are given in Table 4.2.

Table 4.2 Energy in nature with the Miller-index

Natural Energy	Miller-index (x,y,z)
Radiation energy	rad
Mechanical energy	mech
Heat energy	therm
Electrical energy	el
Magnetic energy	mag
Chemical energy	chem

We can determine the number of different types of sensors based on the Miller index classification. Because we can have six different input and six different output energies, and the auxiliary source can be any of them plus one (that is zero if no need of additional source i.e. the sensor is active), the total number of sensor variations is given in (4.3).

$$N_x \cdot N_y \cdot N_z = 6 \cdot 6 \cdot 7 = 252 \quad (4.3)$$

In practical life we do not have to use all of possible sensor variants, because the output energy is electrical, and we use an electrical support (external) source only. If both, the input and output are electrical we do not call that device a *transducer* or *sensor*, so x = 5, y = 1 (el), z = 2 (el or 0 depending if the sensor is active or passive) we have 10 types of sensors only.

The Miller-index based sensor classification can be graphically illustrated as given in Fig. 4.6.

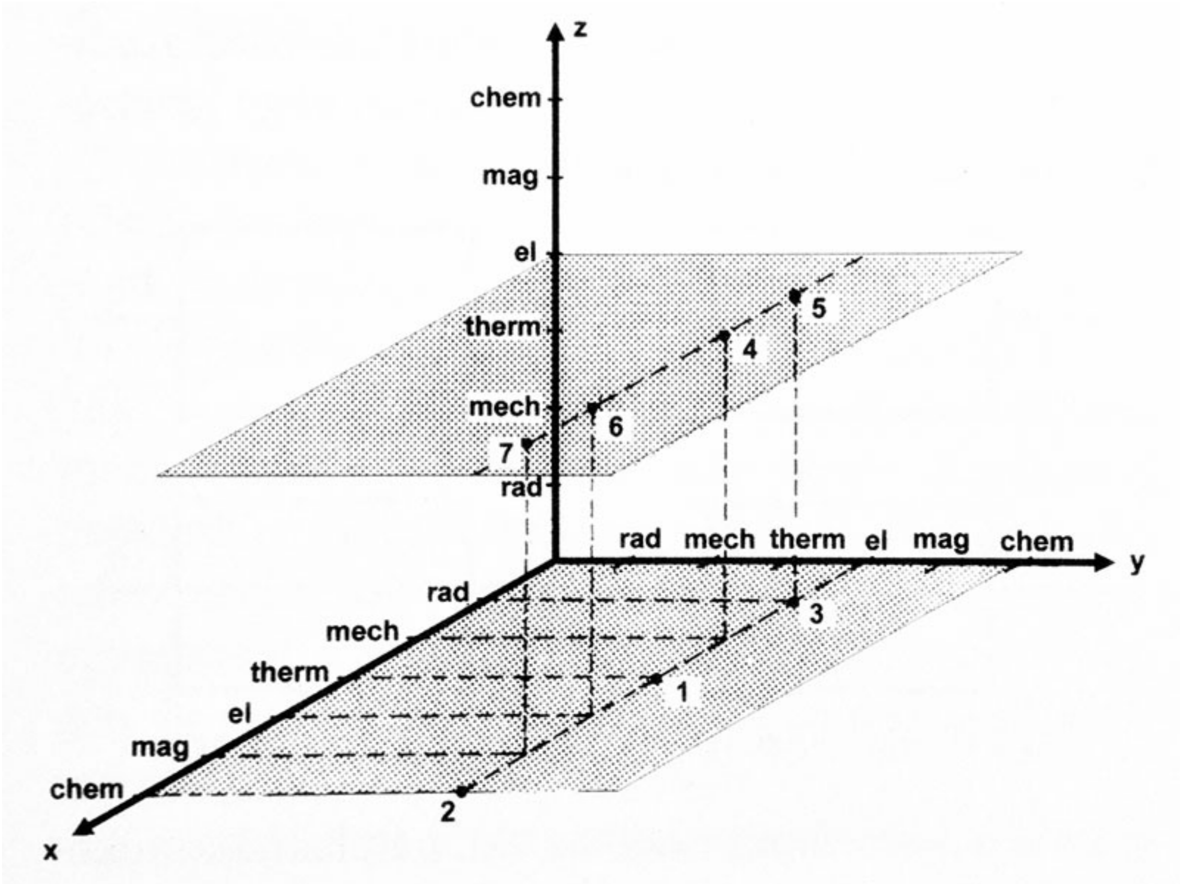


Figure 4.6 Graphical representation of sensor groups

We can see in Fig. 4.6 for example, the following transducers: 1) thermocouple (therm, el, 0), 2) pH-meter (chem, el, 0), 3) PV cell (rad, el, 0), 4) potentiometer (mech, el, el), 5) photoresistor (rad, el, el), 6) ‚pf' (*cos* φ) meter (el, el, el), 7) magneto resistor (mag, el, el).

4.2 Modelling of Sensors

The general block diagram model of a sensor is given in Fig. 4.7.

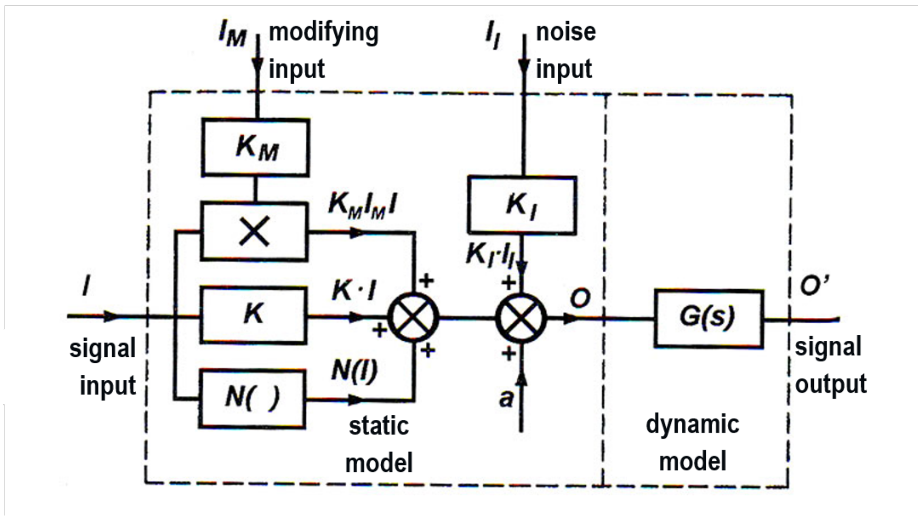


Figure 4.7 Block diagram model of a sensor

The model is divided into a static and dynamic part. According to Fig. 4.7 the output of the static part can also be described by the equation given in (4.4).

$$O = K \cdot I + N(I) + K_M \cdot I_M \cdot I + a + K_I \cdot I_I \tag{4.4}$$

where

- K: scalar factor of input signal;
- N(I): sensor transfer function;
- I_M: superimposed input signal of nonlinearity;
- K_M: scalar factor of nonlinearity;
- a: offset;
- I_I: noise input signal;
- K_I: scalar factor of noise.

The dynamic part of the model is represented by G(s) transfer function.

4.4.1 Practical Examples of Sensor Models

The following figures give practical examples for sensor applications. In Fig. 4.8 there is a block diagram of a strain gauge sensor based on the resistance model.

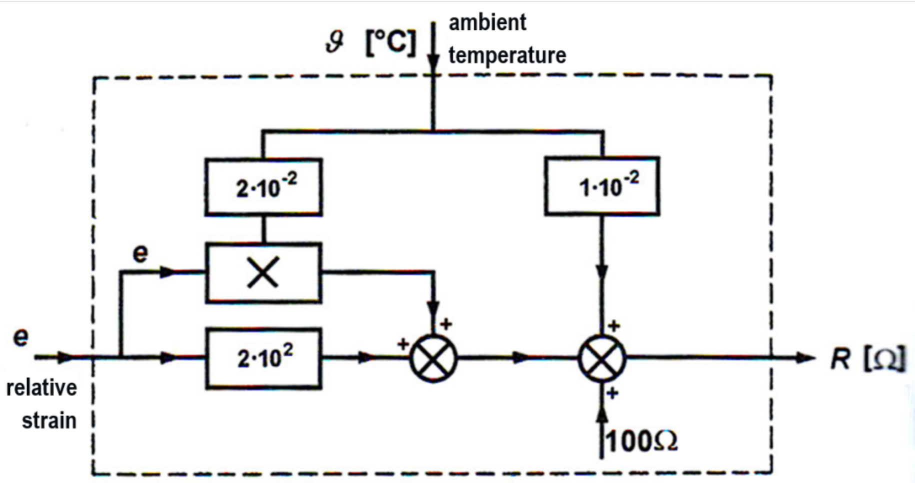


Figure 4.8 Strain gauging sensor

The block diagram of a Copper-Constantan thermocouple is shown in Fig. 4.9.

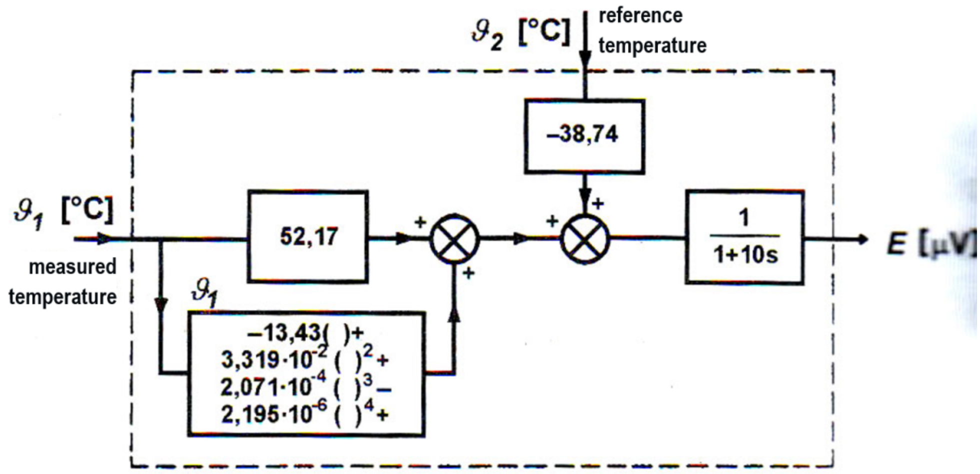


Figure 4.9 Copper-Constant thermocouple

and Fig. 4.10 represents the block model of a piezoelectric accelerometer.

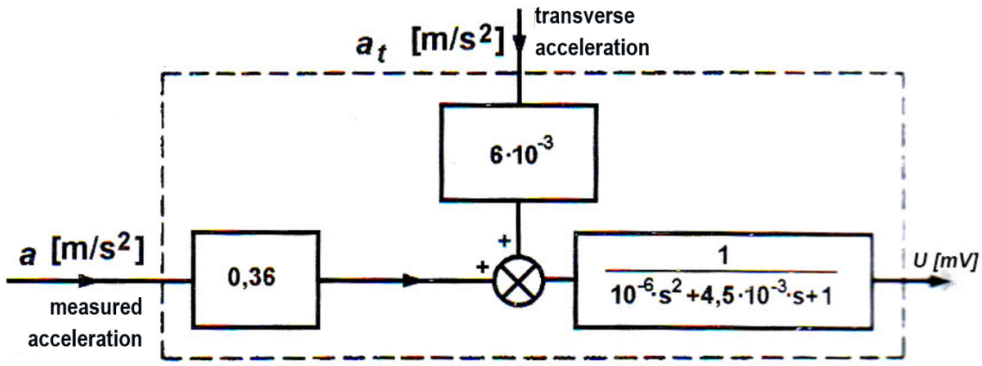


Figure 4.10 Block model of an accelerometer

4.3 Sensor Characteristics

Sensor characteristics, i.e. the relationship between the sensors input (B) and output (K) is defined at specified ambient conditions - *temperature, humidity, blast, etc.,* and given by the supplier as catalogue data. See Fig. 4.11.

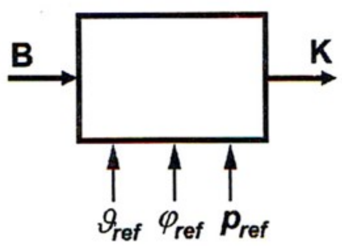


Figure 4.11 Sensor model with ambient parameters

The characteristic behaviour can be examined through static and dynamic conditions.

4.3.1 Static Sensor Characteristics

Typical static characteristics can be divided into three regions as shown in Fig. 4.12. The *Inoperative region* is the lowest part when the volume of input is below B_{min}. In this case the output is below the noise limit, so output cannot be interpreted. In the *operation region* the input is between B_{min} and B_{max}. This region is optimal for practical applications. When the input is over B_{max} we are in the *overload region.* This is a prohibited range of operation because of the unreliable output signal and possible damage to the sensor.

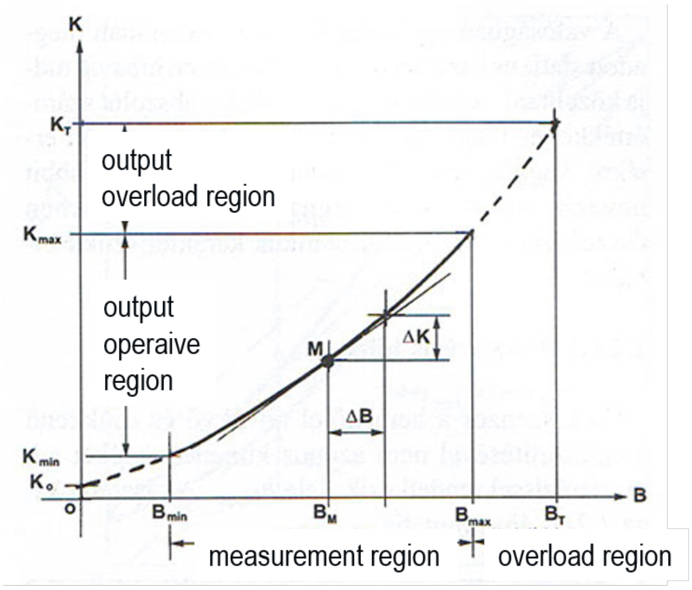


Figure 4.12 Characteristic regions of sensor

One of the main parameters of a sensor is the *static transfer factor* or *static sensitivity* that is defined in (4.5).

$$E_{B_M} = \left.\frac{\partial K}{\partial B}\right|_{B_M} = \left.\frac{\Delta K}{\Delta B}\right|_{B_M} \tag{4.5}$$

In general, the transfer characteristic is a non-linear function and the transfer factor has a working point dependent value. So, it has to be interpreted at a certain working point.

4.3.1.1 Static Characteristics Errors

A static sensor characteristics error is a deviation from the ideal characteristic curve. We can describe sensor errors with absolute error *H,* relative error *h* and measuring range referred accuracy class h_{op}, described in previous chapters. The following section notes the most significant error types from a practical point of view.

Hysteresis Error

The effect of hysteresis is shown in Fig. 4.13. In case of a hysteresis error, the output is different for the same value of input and output if the direction of change is different, i.e. input reaches its value from the bottom to the top or the top to the bottom.

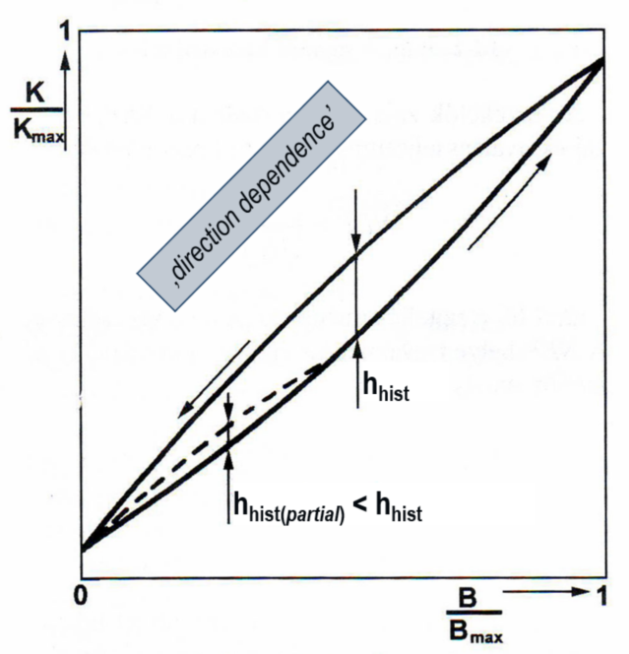


Figure 4.13 Hysteresis error

Repetition Error

Repetition errors occurred when the n-th measuring cycle produces a different result than the first cycle as shown in Fig. 4.14.

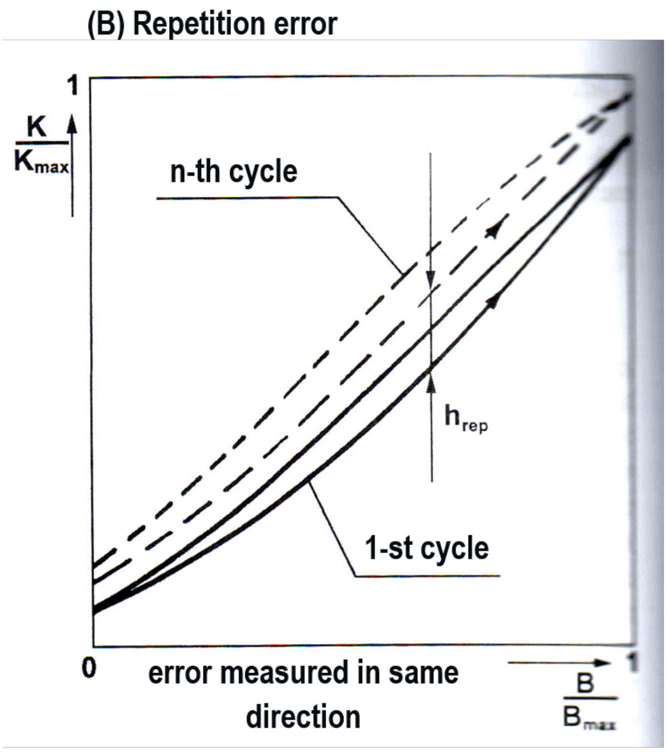


Figure 4.14 Repetition error

Linearity Error (Theoretical)

The effect of linearity error is shown in Fig. 4.15. The reference is the ideal (theoretical) line between 0 (B_0) and 1 (B_{max}). We have to mention that in some practical cases modified definitions of linearity errors are also used, like 'linearity error referred to B_{max}', 'independent linearity error', regression linearity error, etc. These alternative forms of defining linearity error are beyond the scope of this book.

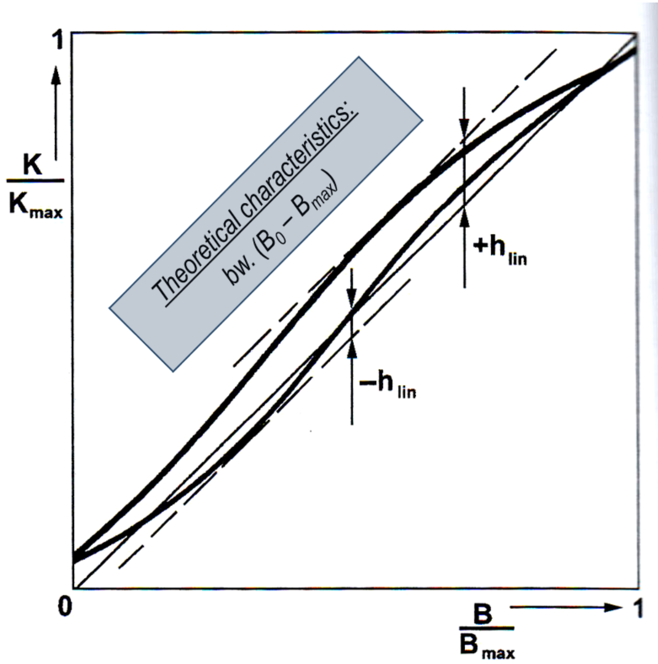


Figure 4.15 Theoretical linearity error

Resolution Error

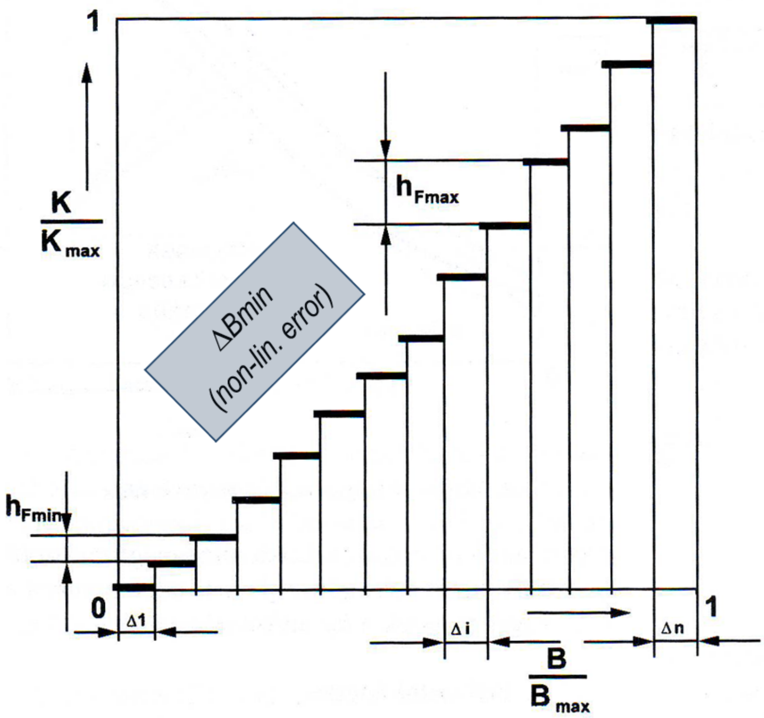


Figure 4.16 Resolution error

Resolution error is a non-linear error as the change step of response (output) depends not only on the change in input but also its absolute value. The effect can be interpreted in Fig. 4.16.

Creep

Creep is an instability in output as the output value changes in time even if there is no change in input value.

Null Point Shift

This parallel shifts occurs when the device produces a non-zero output for zero input.

Sensitivity Change

According to the aging process, sensitivity is not constant. Over a longer period of time its value changes, usually decreasing.

Temperature Dependency

Temperature dependency of the sensor can be described as an 'error range' belonging to the maximum and minimum temperature in the operation range. Fig. 4.17 gives an example of the temperature dependency of hysteresis error. The error range of maximum and minimum temperature values to the reference temperature value can also be defined as shown in Fig. 4.17.

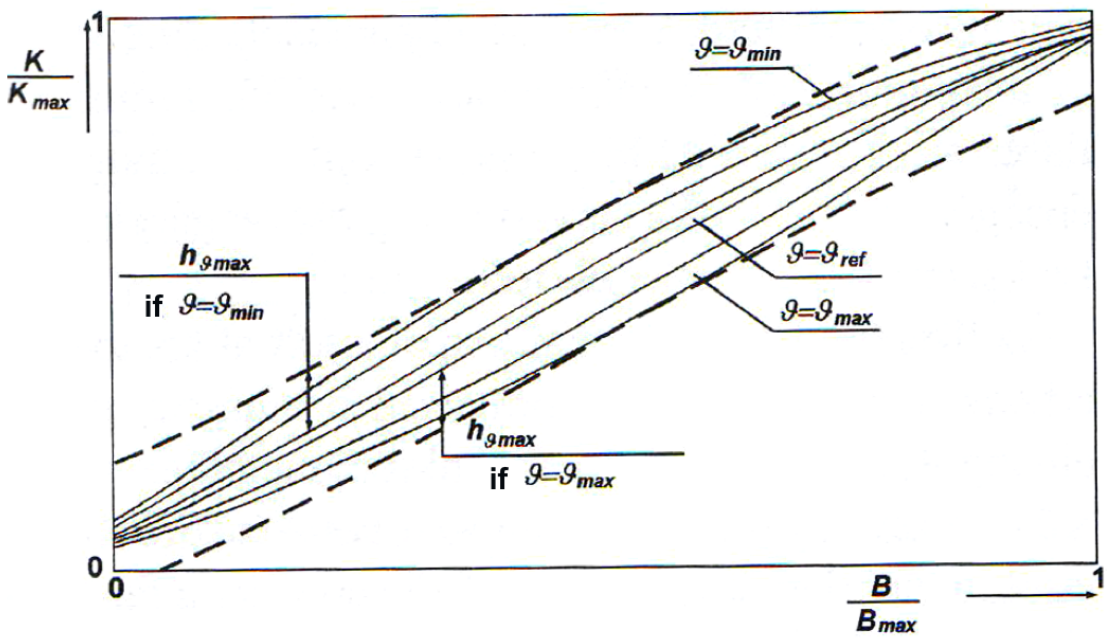


Figure 4.17 Temperature dependency of hysteresis error

Transient Temperature Error

The transient temperature step response is illustrated in Fig. 4.18, describing the effect of step temperature change on the output signal of a sensor. Depending of the transient behaviour of the device the transient response can be over-damped critically damped under damped or undamped according to 2nd order dynamic sensor models.

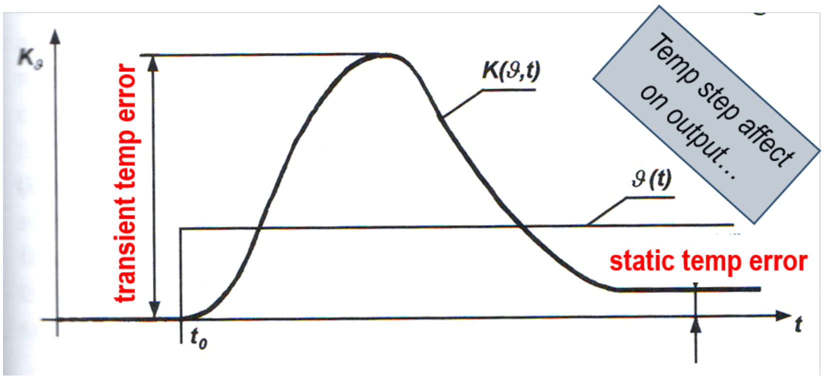


Figure 4.18 Transient temperature step response

4.3.2 Dynamic Sensor Characteristics

When inspecting varying characteristic behaviour over time i.e. finding *K(t)* response for *B(t)* excitation we use a differential equation as the transfer function of the sensor. In transfer function analysis, unit step function, designated as $u(t), \varepsilon(t)$ and unit pulse (or Dirac pulse) function, designated as $\delta(t)$ are used as excitation and the task is to find the *K(t)* response function. The dynamic characteristics in the time domain is the response to the unit step excitation. See Fig. 4.19.

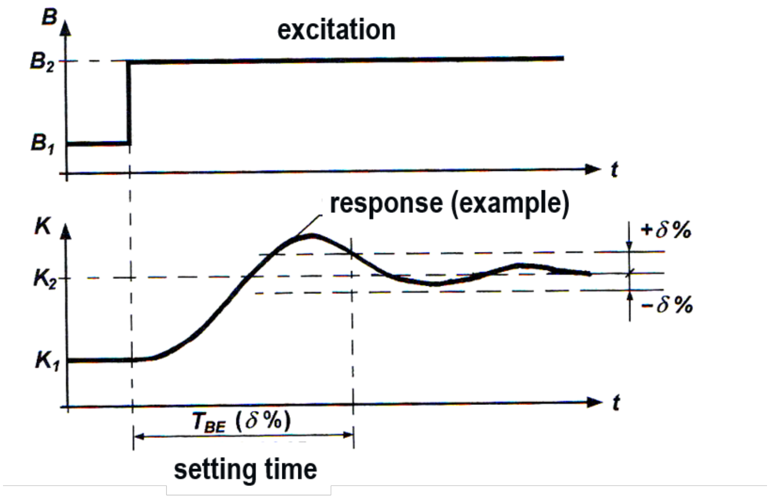


Figure 4.19 Dynamic characteristics over the time domain

Dynamic characteristics can also be found in the frequency domain. In this case the $\boldsymbol{W}(j\omega)$ transfer function is a complex function, a variation of its absolute value by frequency (amplitude characteristic in the Bode diagram) is shown in Fig. 4.20. An important characteristic parameter is *bandwidth* defined as frequency 'distance' between the lower and upper cut-off frequencies. Cut-off frequency belongs to a predefined $\pm\delta$ region as the maximum fluctuation of the output signal, usually 3 dB. In Fig. 4.20 a low-pass filter dynamic behaviour shows that the lower cut-off frequency is zero and the bandwidth is equal to upper cut-off frequency.

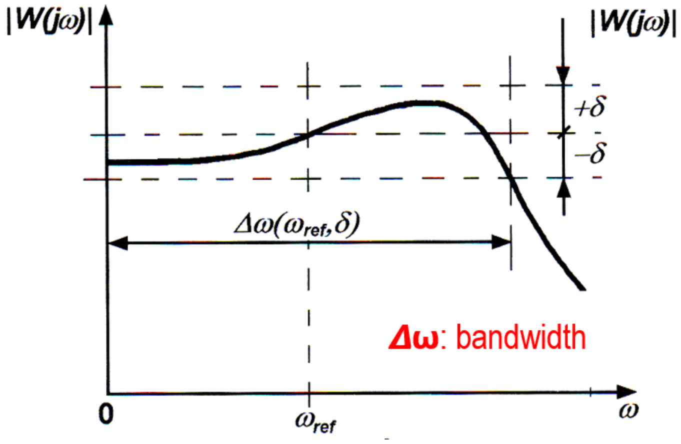


Figure 4.20 Dynamic characteristics in frequency domain

4.4 Sensor Groups Overview

In this section we have collected the most important sensor groups from a practical application point of view. These have been put into groups depending on the five types of energy to be measured, shown in Table 4.3. The headers of the columns give the Miller-indexes and in each group the most important sensors are given. (A Miller-index is given by *'x'* when it is undefined, i.e. it can be *'0'* or *'el'* according to whether the sensor is *active* or *passive*.)

More details about each element in Table 4.3 will be discussed in following chapters.

Table 4.3 Sensor groups

(THERM,EL,X)	(RAD,EL,X)	(MECH,EL,X)	*(MAG,EL,X)*	*(CHEM,EL,X)*
Metal Resistors	*Photo-meters*	*Force, Torque, Pressure*	*Magneto-resistors*	*Electro-chemistry*
Semiconductor NTC, PTC, pn	*Radiovawe Detectors*	*Position Detectors*	*Galvanic-Magnetic Det.*	*Humidity Detectors*
Thermo-couples	*IR Sensors*	*Flow Sensors*		*Ionselective Detectors*
Thermal Switches	*Quant Detectors*	*Speed Sensors*		*Gas Detectors*
		Accelero-meters		
		Acoustic Detectors		

4.5 Sensor signal transmission overview

The output signal of a sensor has to be transmitted to evaluating and processing elements. The signal transmission process is summarized in Table 4.4. More details about signal transmission methods will be discussed in the following chapters.

Table 4.4 Signal transmission methods

TRANSMISSION METHODS	TRANSMISSION MEDIUM	REMOTE TRANSMITTER	SMART SENSORS
Base Band	*Copper Wire*	*Analogue Transmitters*	*SCADA Systems*
Carrier Band	*Optical Fiber*	*Digital Transmitters*	*DCS Systems*
	Wireless Transm.		*Embedded Technologies*

4.6 Sensor Application Examples

Finally, in this chapter about sensors, here are some easy to understand operational principles and examples for sensing motion, acceleration, and light.

4.6.1 LVDT Displacement Sensor

Fig. 4.21a shows the principle operation of a motion sensor using a Linear Variable Differential Transformer (LVD). The output voltage magnitude and phase vs. displacement is illustrated in Fig. 4.21b. The seconder coils of the differential transformer are symmetric and are connected opposite so the output voltage is given by (4.6)

$$V_{OUT} = V_1 - V_2 \tag{4.6}$$

In case of the reference position of the core V_1 *is* equal to V_2 so the output voltage is zero. Displacement in both directions produces a positive voltage difference in output due to AC excitation but phase shifting holds information about the direction of motion as shown in Fig. 4.21b. More details about LVDT sensors are discussed in Chapter 6.

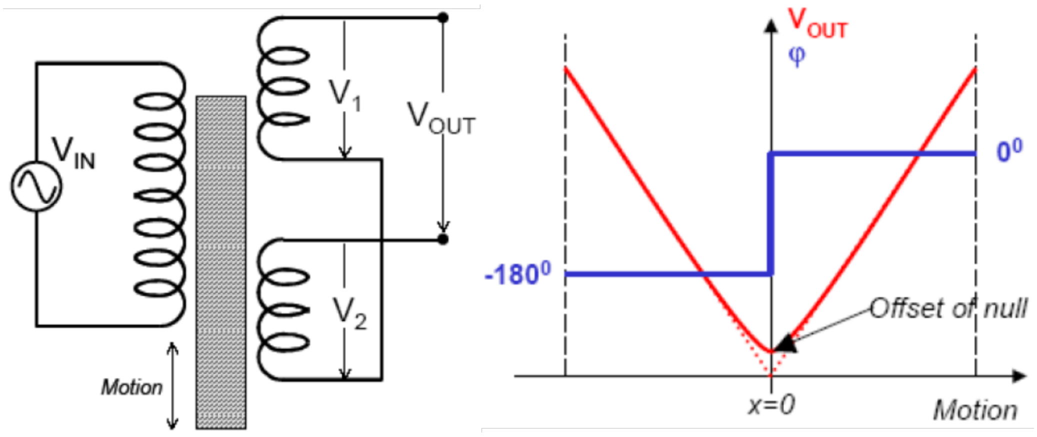


Fig 4.21 Motion sensor a) with LVDT, and b) its characteristics

4.6.2 Motion Sensor with Optocoupler

Motion sensing can also be established with an optocoupler, as is shown in Fig. 4.22. The position of the moving blocker controls the light intensity and subsequently the output current of the optocoupler.

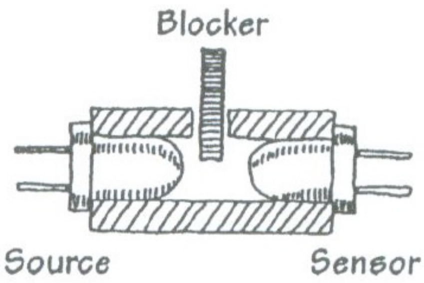


Fig 4.22 Motion sensor with optoisolator

4.6.3 Motion, Stress, Tension and Strain Sensors

A strain gauge, shown in Fig. 4.23, is a widely used sensor element for detecting and measuring motion, displacement, mechanical stress, pressure and strain. The metal layer on a plastic film has a structure like in Fig. 4.23. The longitudinal tension increases the electrical resistance of the metal layer but, because of the special geometric design, resistance is insensitive to lateral force. More details about strain sensors are discussed in Chapter 6.

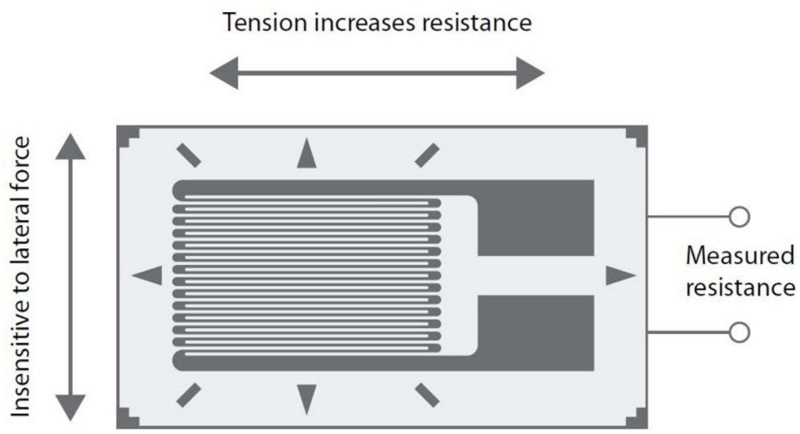


Fig 4.23 Strain gauge

4.6.4 Accelerometers

The principle of a piezoelectric sensor used for measuring acceleration is shown in Fig. 4.24. The device consists of a piezoelectric (crystal) element that produces an electric voltage under mechanical tension. The crystal is placed between a base element and a freely moving seismic mass. When acceleration occurs the seismic mass pushes (or pulls) the crystal which causes it to produce electrical voltage on its contacts. Because the internal impedance of the crystal is extremely high, an ICP (Integrated Circuit Piezoelectric) amplifier has to be applied for impedance matching and for a reasonable output voltage level.

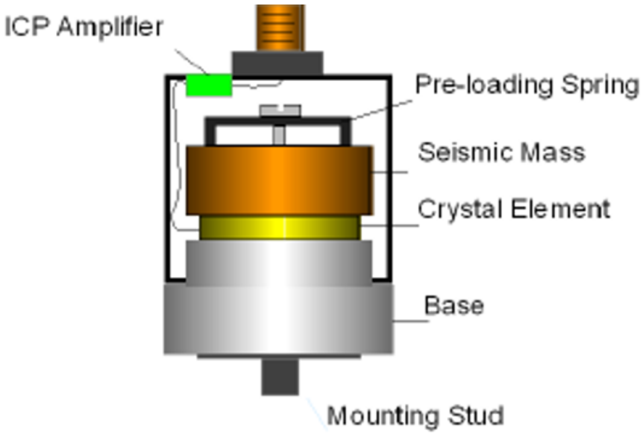


Fig 4.24 Piezoelectric accelerometer (sensor)

4.6.5 Photo Sensors

A Cadmium-Sulphur (CdS) sensor is only one of the possibilities of light-electrical transducers but it is widely used because of its simplicity and reliability. CdS cells have a feature where the resistance changes when subjected to light. A simplified representation is given in Fig. 4.25. More details about CdS cells and other photo-detectors will be discussed in detail in the following chapters.

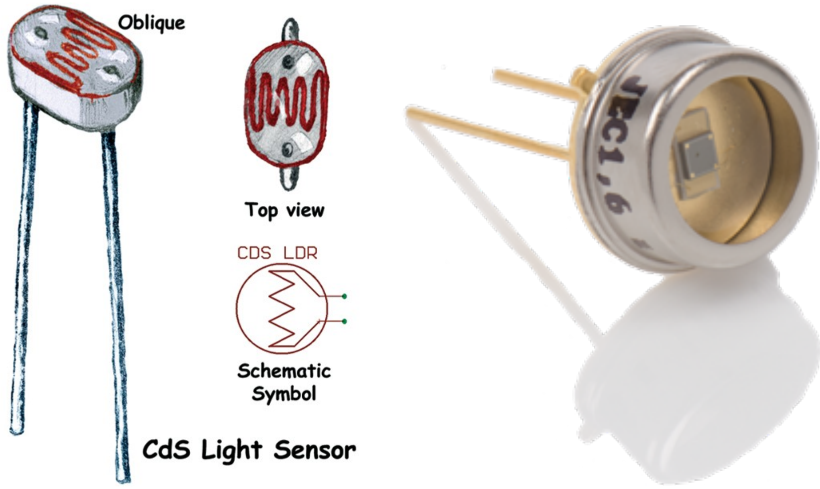


Fig 4.25 CdS photo sensor

5. Basic Sensor Networks

5.1 Measurement Network Architecture

In general, a measurement network is a circuit containing signal source, signal receiver and transmission medium between the source and receiver. The structure of this measuring network architecture is shown in Fig. 5.1.

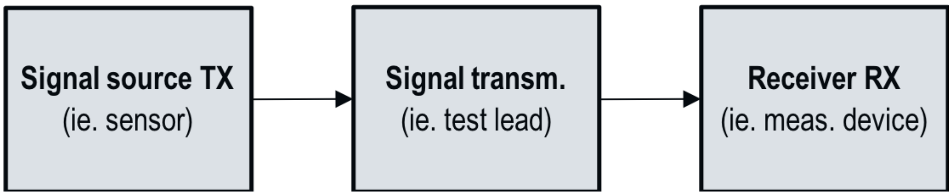


Figure 5.1 Measurement network

In more detailed analysis of measurement networks, we need to classify sources, i.e. *transducers or sensors,* indicated by a TX and receivers, which are indicated by an RX, which we signify the *measuring devices* from different aspects. *Impedance aspect* transducers and measuring devices can have symmetric and asymmetric constructions. With respect to grounding, both TX and RX can be *grounded*, *ground-independent* and *grounded with ground offset*. From a source model point of view, we can have a *voltage source model* or *current source model*. The following section looks more closely at signal sources, signal receivers and their possible interconnection.

5.1.1 Asymmetric Signal Sources

Asymmetric Grounded Source

The Thevenin model of an asymmetric grounded source is shown in Fig. 5.2. In this case one of the source terminals is grounded causing different impedances between ground and each source terminals.

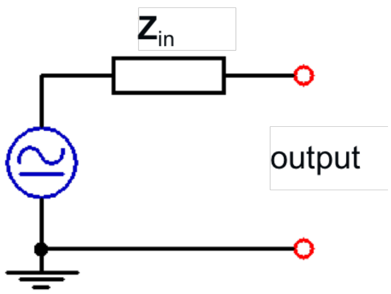


Figure 5.2 Asymmetric grounded source

Asymmetric Ground-Independent Source

The source model is shown in Fig. 5.3. Because each terminal is insulated from ground, one terminal can be grounded. The main feature of this asymmetric model is that different impedances can be measured between ground and each terminal.

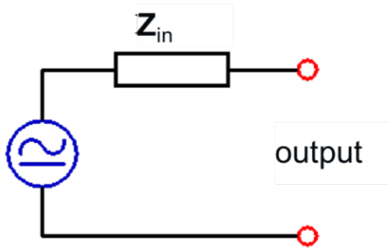


Figure 5.3 Asymmetric ground-independent source

Asymmetric Grounded Source with Ground Offset

The Thevenin model is shown in Fig. 5.4. The main feature of this model is that neither terminal can be grounded and impedances between ground and each terminal are different.

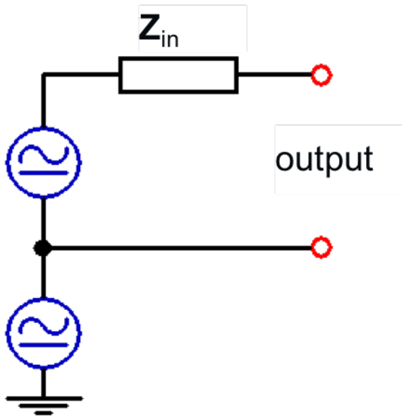


Figure 5.4 Asymmetric grounded source with ground offset

5.1.2 Symmetric Signal Sources

Symmetric Grounded Source

In this case neither terminal can be grounded, and impedances are the same between ground and each terminal. A symmetric grounded source model is shown in Fig. 5.5.

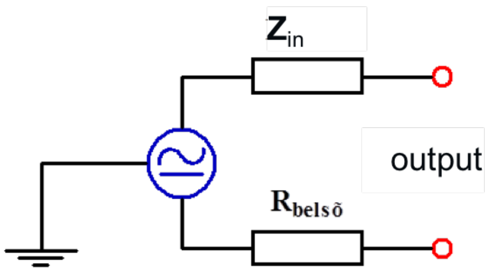


Figure 5.5 Symmetric grounded source

Symmetric Ground-Independent Source

In this case, shown in Fig. 5.6, one terminal or common point can be grounded. The same impedances are between the common point and the terminals.

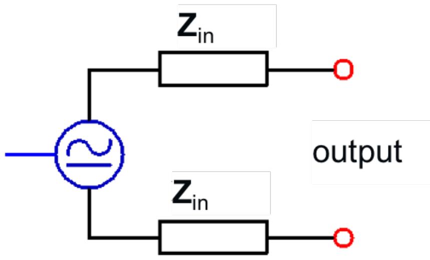


Figure 5.6 Symmetric ground-independent source

Symmetric Grounded Source with Ground Offset

In the model given in Fig. 5.7 neither terminal 1 nor terminal 2 can be grounded and the same impedances exist between ground to terminal 1 and ground to terminal 2.

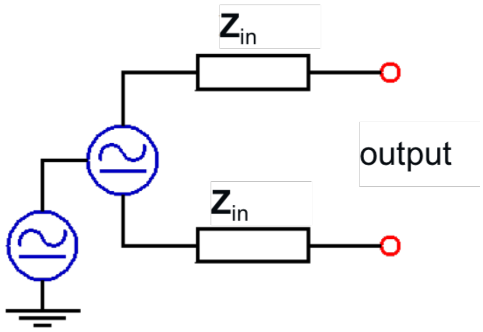


Figure 5.7 Symmetric grounded source with ground offset

5.1.3 Asymmetric Signal Receivers

Asymmetric Grounded Receiver

Fig. 5.8 shows that one input and one output terminal and also the housing are connected to ground. According to the asymmetric circuit topology, the impedances between ground and the input terminals are different.

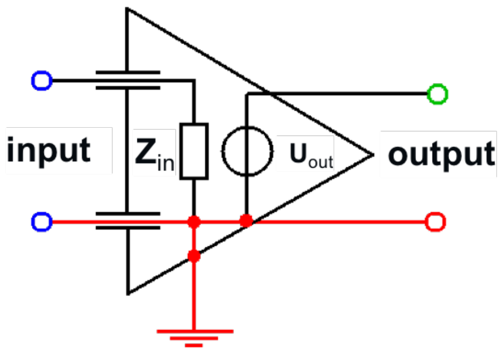


Figure 5.8 Asymmetric grounded receiver

Asymmetric Ground-Independent Shielded Receiver 1

According to the connection shown in Fig. 5.9, one output terminal and housing are connected to ground. One input is connected to an independent shielding. Accordingly, the internal shielding is independent from the housing, thus, inputs are independent from ground, output and also housing.

Because of asymmetric structure impedances are different between ground and input terminals. If one input is externally connected to ground, then different impedances can be measured between ground and each input terminal.

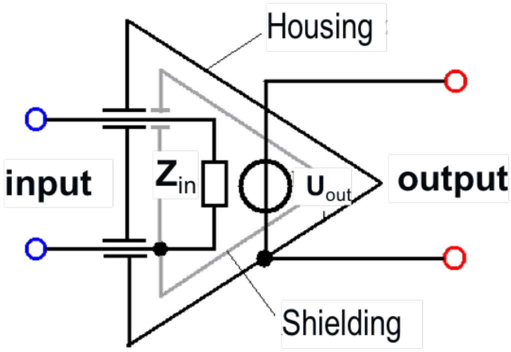


Figure 5.9 Asymmetric ground-independent shielded receiver 1

Asymmetric Ground-Independent Shielded Receiver 2

One of the output terminals and housing are grounded. Shielding, so called *guard,* (G) is floating i.e. ground independent. Features of this solution are, on one hand the high (parasitic) impedance between input H (high) and G (guard), and on the other hand the low (parasitic) impedance between input L (low) and G (guard). The connection of this solution with a parasitic capacitance is shown in Fig. 5.10.

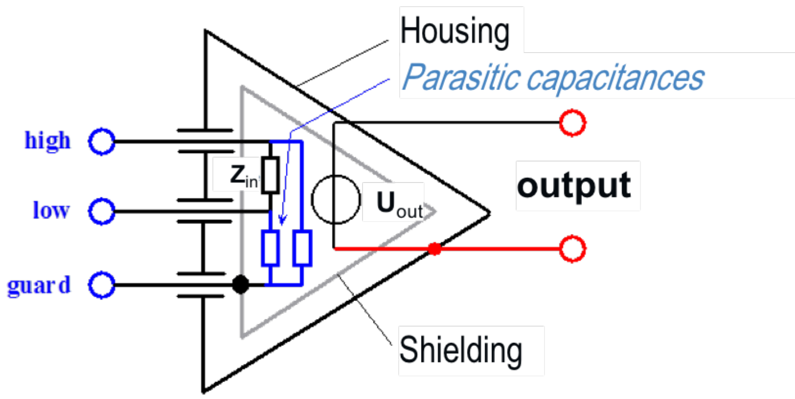


Figure 5.10 Asymmetric ground-independent shielded receiver 2

5.1.4 Symmetric Signal Receivers

Symmetric Grounded Receiver

The circuit model for this solution is shown in Fig. 5.11. The input common terminal, one output terminal and housing are connected to ground. The main feature is the same input impedance of both input terminals to ground.

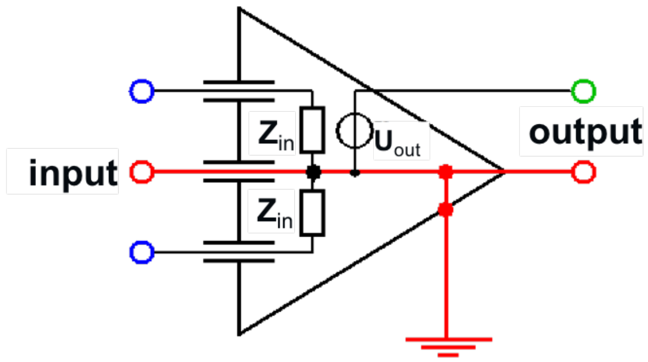


Figure 5.11 Symmetric grounded receiver

Symmetric Ground-Independent and Protective Shielded Receiver

The circuit structure is shown in Fig. 5.12. In this case one output terminal and housing are grounded and so, parasitic impedances between input terminals and shield are the same. Applying non-symmetrisation i.e. connecting the guard to one of input terminals will convert it to an asymmetric ground-independent receiver.

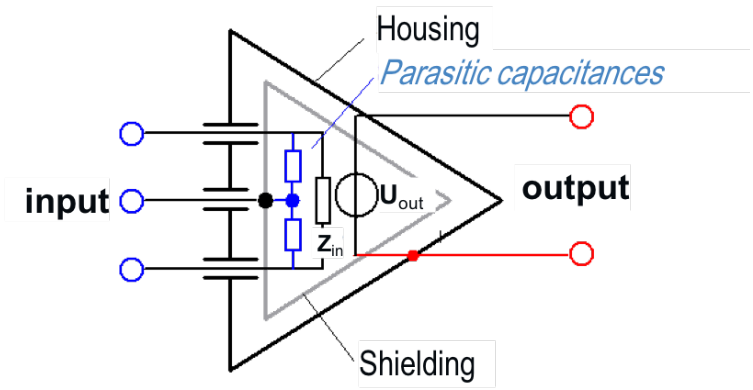


Figure 5.12 Symmetric ground-independent and protective shielded receiver

5.1.5 Signal Source to Receiver Interconnection

In practical applications, those sources and receivers mentioned so far have to work in interconnection. Because of the different types, the method of interconnection is not evident in some cases, i.e. interconnection of asymmetric sources and symmetric receivers, symmetric sources and symmetric receiver connections to reference ground, etc.

The signal transmission line is often shielded to reduce noise in the signal. The line shield has to be connected to a defined potential that is usually ground potential. This connection should be set up carefully paying attention to the following rules and examples.

Rules of Thumb for Interconnection

The first very important but hard to establish rule is that the *system must be grounded at a single point.* The difficulty of this rule is the connection of different systems that might already be separately grounded.

The second important rule is *to a build symmetric circuit* – 'if possible'. Doing this results in better noise suppression.

Fixing the shield potential is also required because of improved noise rejection. The single point of cable shield grounding should be at the shield of the receiver in order maintain low voltage between the test lead and the receiver's shield.

In case of a grounded source is applied the cold point (L) of the source has to be connected to ground.

The following interconnection examples illustrate the practical solutions according to the rules above.

Interconnection Example 1

In many cases, a ground independent asymmetric source (i.e. sensor) has to be connected to an asymmetric grounded receiver (i.e. instrument amplifier). The correct interconnection is demonstrated in Fig. 5.13. As can be seen, the system is grounded at a single point only. The shielding of the test lead is also grounded to fix its potential (to zero) at the receiver side.

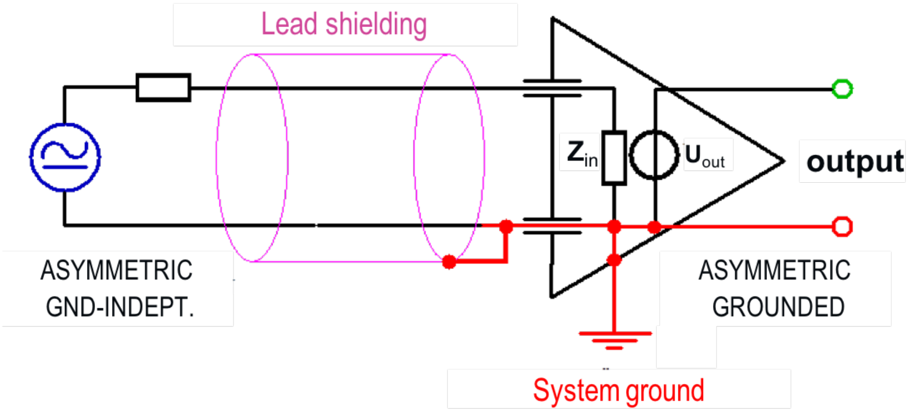


Figure 5.13 Interconnection for asymmetric elements

Interconnection Example 2

An asymmetric and ground independent source is connected to a symmetric and grounded receiver in Fig. 5.14. Symmetric connections result in a better Signal to Noise Ratio (SNR) than asymmetric ones. In this example, the system is single point grounded at the receiver side. Fixing test lead shielding potential is done by grounding at the receiver side. The system is grounded at a single point.

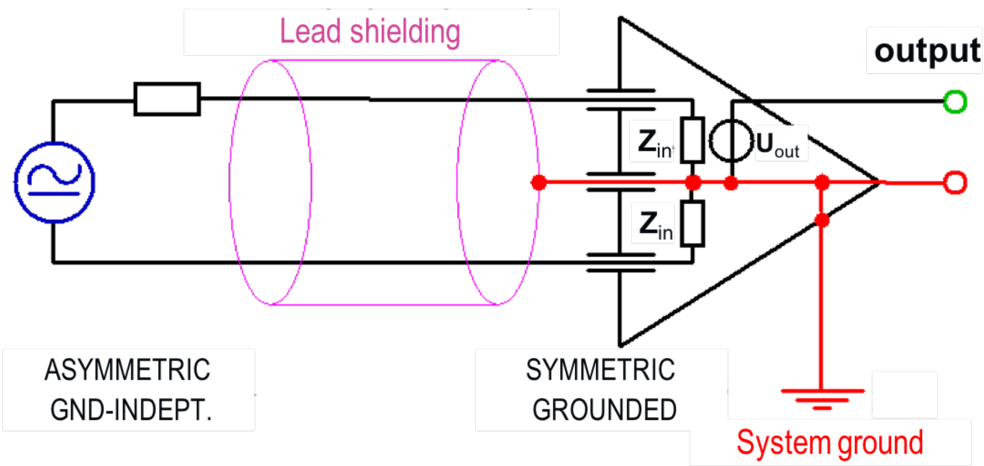


Figure 5.14 Asymmetric source connection to symmetric receiver

Interconnection Example 3

From a noise rejection point of view the best solution is to a apply symmetric solution for both the source and receiver side. Fig. 5.15 demonstrates the connection of a symmetric and ground independent source to the symmetric and grounded receiver. The applied rules of the 'single point ground' and 'fixed cable shielding potential at the RX side' can be seen in Fig. 5.15.

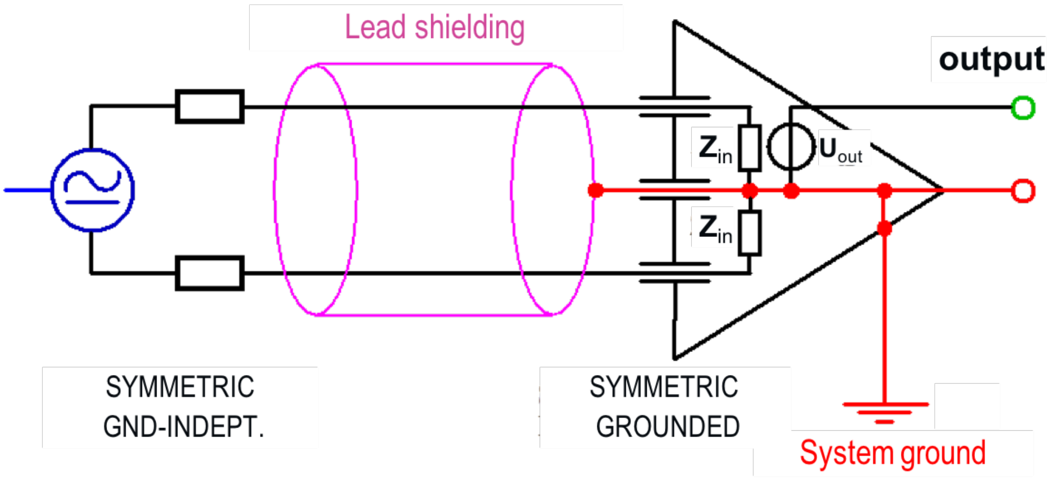


Figure 5.15 Interconnection of symmetric elements

Interconnection Example 4

When the signal of an asymmetric and grounded source has to be measured, it should be connected as shown in Fig. 5.16. Because the source is grounded the input part of the receiver must be ground independent to stay consistent with the rule of single point grounding. Shielding of a test lead is done by grounding on the source side where terminal 'L' is grounded. In this case the shielding of the test lead must not be connected to ground at the receiver side because of rule of single point grounding. The receiver is shielded for better noise rejection and the shield is connected to the 'L' input terminal of the receiver. The receiver is connected to ground at output only that has no galvanic connection to the receivers' input.

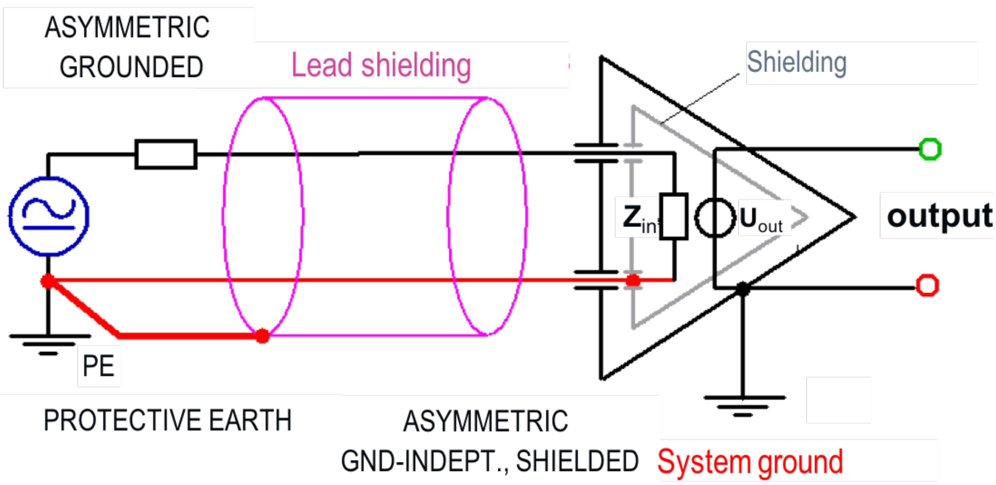


Figure 5.16 Asymmetric and grounded source connection to an asymmetric receiver

Interconnection Example 5

The right connection of an asymmetric and grounded source to a symmetric and ground independent, shielded receiver is shown in Fig. 5.17. Applying a grounded source, receivers' input must be ground independent. The potential of test lead shielding is fixed at the source side, but cable shielding is connected to guard at the receiver to reduce noise.

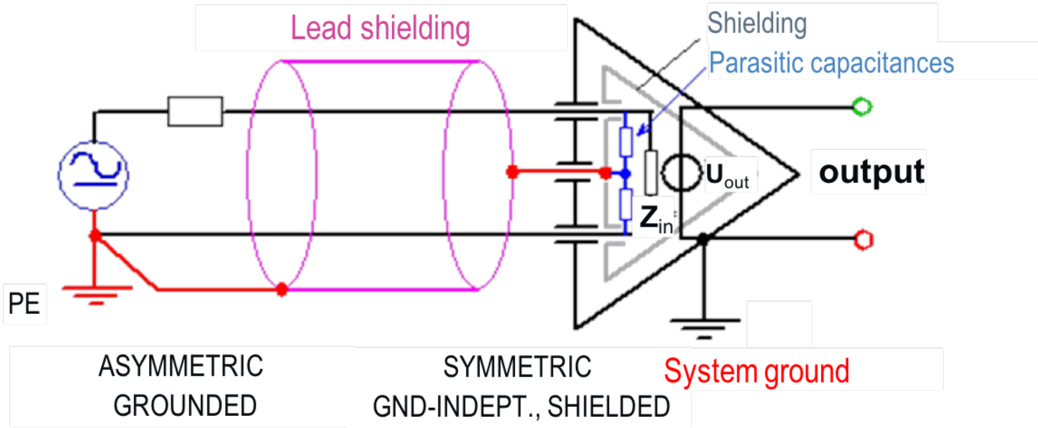


Figure 5.17 Asymmetric and grounded source connection to a symmetric receiver

Interconnection Example 6

In our last source to receiver example we have an asymmetric grounded source with ground offset. The source is not grounded directly in this case, but the offset voltage occurs between the low (L) terminal of the source and the ground, i.e. Protective Earth (PE). The source is connected to a symmetric and ground independent, shielded receiver as seen in Fig. 5.18.

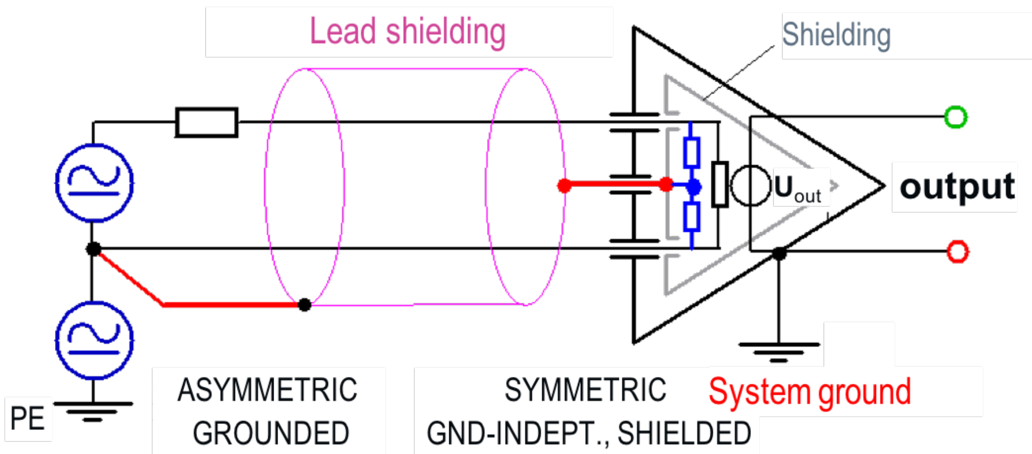


Figure 5.18 Asymmetric and offset grounded source connection to symmetric receiver

Cable shielding is connected to fix potential at the source side, but it can be terminal 'L' only. On the receiver side, this cable shielding is connected to the guard, i.e. the internal shielding of the receiver. The receiver is grounded only at the asymmetric output side.

Rules of Thumb for Lead Shielding

From the interconnection examples above, you can see that the planning of cable shielding is challenging with several factors having to be considered depending on the applied interconnection method. Here is a useful summary of the most important 'rules of thumb' to be considered in cable shielding.

When a ground independent source is applied, the cable shield has to be connected to ground at the receiver side. The reason is to maintain zero shielding potential at the receiver side, close to its sensitive input terminals.

When we have grounded the source without an offset voltage, then we have to connect the cable shielding to ground at the source 'L' terminal because this point has zero (ground) voltage. Grounding should never be done at the receiver side because single point grounding is a must to avoid ground loop that will be discussed later on.

When we have grounded the source with the ground offset voltage then we have to connect cable shielding to the 'L' terminal of the source.

In the case of applying a ground independent receiver (with internally applied shield 1 or shield 2 construction) the cable shielding has to be connected to guard i.e. to the internal shield 1 or shield 2 at receiver side. The only possible grounding of cable shielding should be on the source side.

Inadequate use of cable shielding, i.e. the negative effect of double grounding, is shown in Fig. 5.19. *The connection, shown in this figure must to be avoided.* The ground loop current, resulting from the possible potential difference between two separate grounding, causes an electromagnetic influence within the cable that is harmful from a noise rejection point of view.

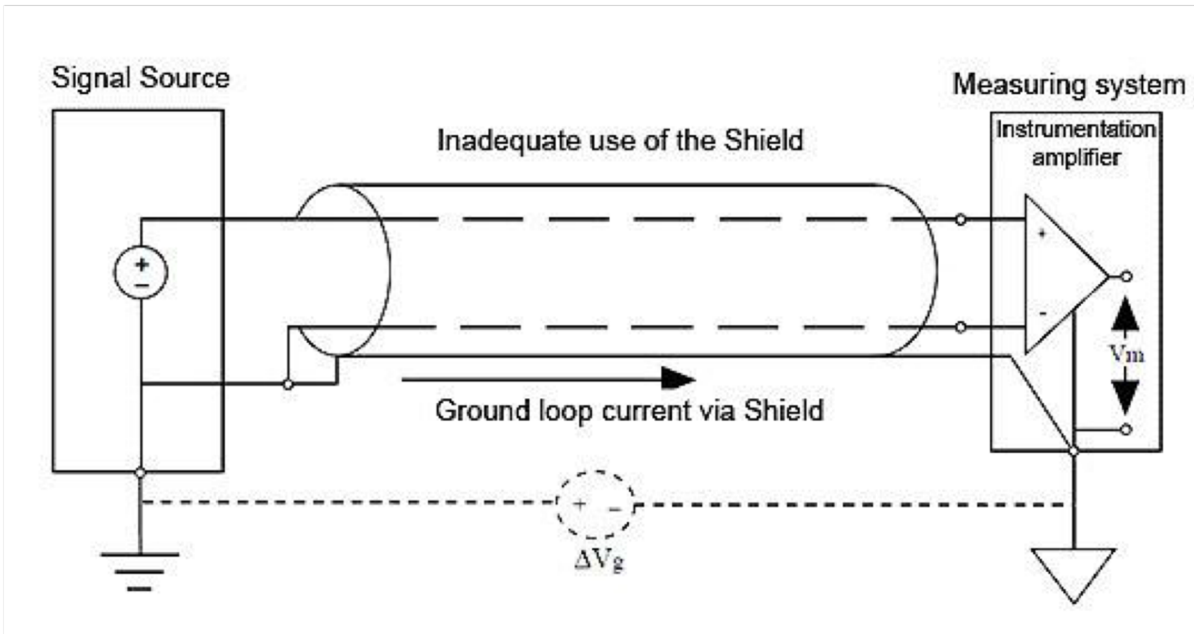


Figure 5.19 Ground loop caused by double grounding

The aspects of noisy signals and possible methods of noise rejection will be discussed in more detail in the following chapters.

5.2 Noise in Electrical Circuits

5.2.1 Noisy Signals and Noise Reduction

When measuring signals under real circumstances there is always some noise. This additional noise causes information loss or *distortion* in the measuring signal. One specific parameter of that distortion is the *signal to noise ratio* (SNR). Its level is usually measured in decibels (dB) as given in (5.1)

$$L_{SNR}^{(dB)} = 20 \log \frac{U_S}{U_N} \tag{5.1}$$

where U_S is the RMS value of the signal voltage and U_N is the RMS value of the noise voltage. Noise usually comes into the measuring system through coupling the transmission channel between the signal source (SRC) and receiver as shown in Fig. 5.20.

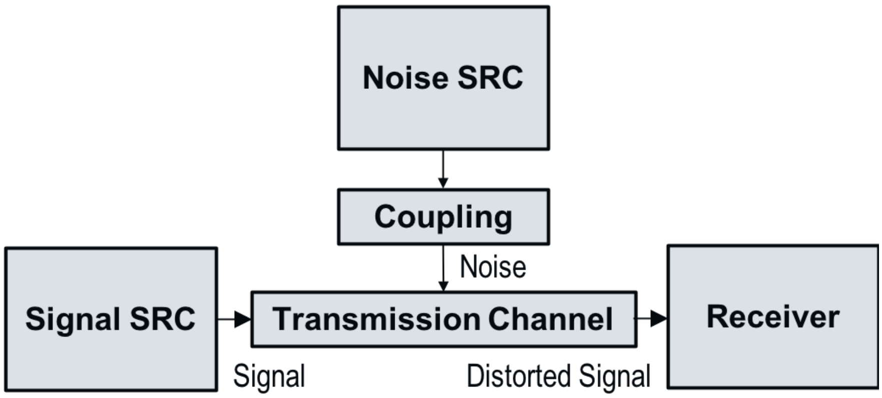


Figure 5.20 Noise coupling model

Our goal is to reduce information loss in the transmission channel i.e. to reduce noise. One possible way for noise reduction is to reduce coupling. The next section explores various coupling types providing practical noise reduction methods.

Coupling Types

The most probable cases for noise coupling, illustrated in Fig. 5.21, are the following.

In the case of a *conductive coupling* galvanic connection occurs between the signal transmission channel (as 'victim') and the noise source. In the case of *inductive coupling,* the connection is based on magnetic induction and the *capacitive coupling* comes to be through an electrostatic link. In some cases, a fourth type of noise coupling comes into the picture that is called the radiative link between the noise source and the 'victim'. In this group we can mention acoustic coupling caused by mechanical vibration or audio noise, for example.

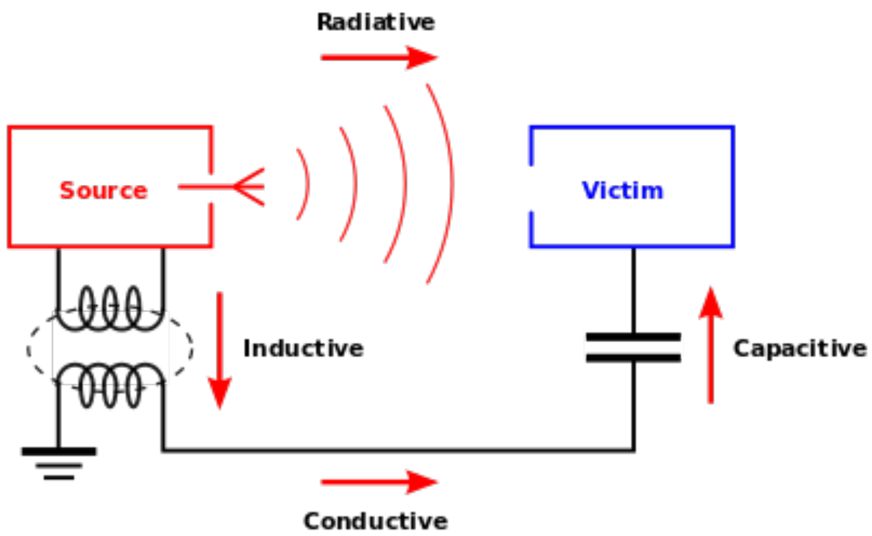


Figure 5.21 Coupling types

Noise reduction in the measuring environment is hard to do and needs thorough analysis, design and implementation in each unique case. However, some of general methods are given in the next section.

Noise Reduction Methods

First of all, we have to emphasize that the most efficient method for noise reduction is *eliminating the noise source* or, if it is not possible, *reducing its impact* at least.

Noise source is independent of us in many cases, so we cannot get rid of the harmful source. In such cases we try to eliminate – or at least reduce – the connection between the source of noise and the protected system.

The next method is filtering a noisy signal i.e. suppressing noise components in the measuring signal. Because of possible noise coupling from the test lead the best location for filtering is on the receiver side. Depending on the applied filtering method we distinguish between *analogue filtering* and *digital filtering*. An analogue filter is usually a low pass filter circuit, integrated into the receiver input section while a *digital filter* is usually done as signal post-processing by software.

As a rule of thumb, *frequency diversification* is often applied for noise reduction by using different signal frequencies from the noise frequency. Filtering can be done more efficiently if the noise components are in a different part of the frequency spectrum than the signal components.

Last, but not least we have to mention a method that is really innovative but has been in use for many years. *Active noise reduction* methods apply additional noise to the system but in opposite phase from the original noise to be suppressed. Fig. 5.22 demonstrates the method in acoustic noise reduction. Number of practical solutions are based on this method like reduction of machine vibration, reduction of vibrations in airplanes, etc.

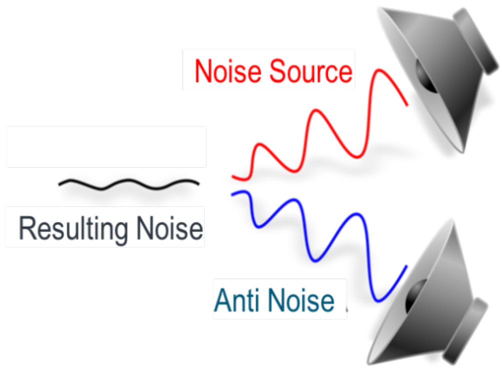


Figure 5.22 Active noise reduction

The noise reduction methods, mentioned above, are some of the most important and widely used techniques but we should bear in mind that this list is far from complete. Because noise reduction is a difficult job, in many cases it needs specific design and realization for each case. This collection of noise reduction methods can be used as the starting point for noise rejection.

5.2.2 Classification of Electric Noise

Because different kinds of noise need different intervention, we have to classify electric noise for better understanding.

According to their *time varying behaviour* we distinguish between *DC noise* and *AC noise*. DC noise is a time invariant noise that has amplitude as its single parameter. AC noise is a time variant noise, so it can be characterized by means of additional parameters like frequency, phase, in addition to its amplitude.

Additionally, we can also distinguish time variant noise such as *transient noise* and *stochastic noise*. While transient noise is a single quick run signal, i.e. switching on or switching off an inductive load, stochastic noise is a continuous signal over time with no periodic behaviour. Its frequency spectrum is not discrete but has a continuous and broadband spectrum.

According to the appearance of noise in measuring circuits we can distinguish between *differential mode noise (DMN)*, *common mode noise (CMN)*, and differential mode noise *caused by common mode noise*.

Differential Mode Noise (DMN)

DMN noise is also called *series noise* or *transversal noise* because of its series appearance in the measuring loop. An important feature of this noise is the counter phase control of the receivers input terminals. The effect diagram of differential mode noise is shown in Fig. 5.23.

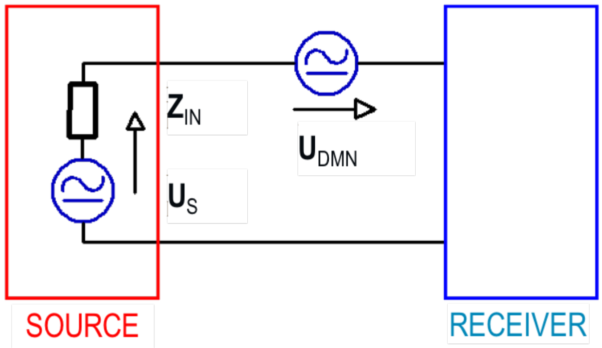


Figure 5.23 Effect diagram of DMN noise

Common Mode Noise (CMN)

Common mode noise is noise of the source with offset ground. An important feature of this noise is the common phase control of the receivers input terminals. The effect diagram of differential mode noise is shown in Fig. 5.24.

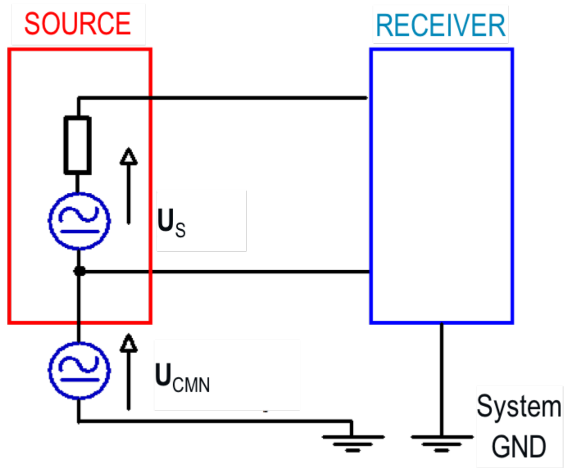


Figure 5.24 Effect diagram of CMN noise

Differential Mode Noise Caused by Common Mode Noise

Most of the problems in noise reduction solutions are caused by the common mode noise that is transformed to differential mode noise due to an incorrect grounding system. Double grounding causes differential mode noise even if only common mode noise was present originally. The harmful effect caused by double grounding is illustrated in Fig. 5.25. A receiver has an asymmetric grounded input and it is supplied by an asymmetric grounded source. Because the source and the receiver are distant there is likely to be a potential difference (voltage) between two grounds. This voltage conducts itself to a common mode voltage in the circuit of Fig. 5.25.

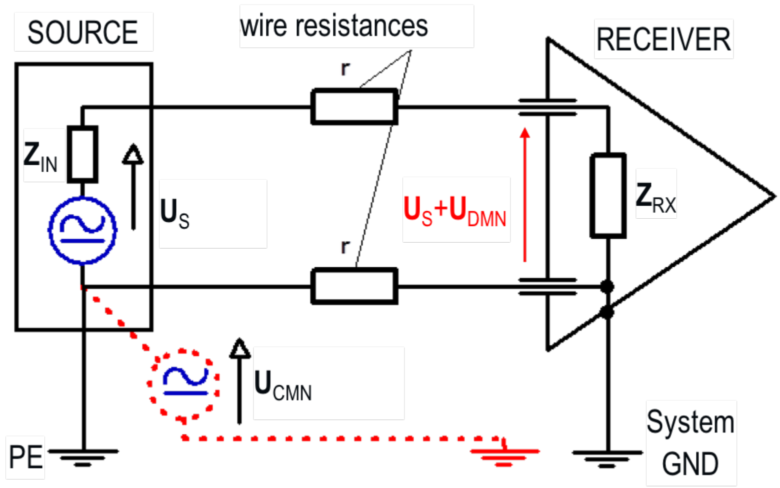


Figure 5.25 Effect of double grounding

Applying a voltage division for the circuit we can express the voltage component at receivers input as given in (5.2).

$$U_{DMN} = \frac{Z_{RX}}{Z_{RX} + Z_{IN} + r} U_{CMN} \tag{5.2}$$

In most cases the receivers' input has a high (extremely high) input impedance compared to low lead resistance and low impedance of voltage source (Thevenin generator model). Thus, a differential mode voltage caused by a common mode voltage can be approximated as in (5.3).

$$Z_{RX} \gg Z_{IN} + r \rightarrow U_{DMN} \approx U_{CMN} \tag{5.3}$$

Thus, the common mode noise voltage (U_{CMN}), measured between the system ground and PE ground, is transformed into differential mode noise (U_{DMN}), measured between the input terminals of the receiver. This harmful noise is caused simply by double grounding. Our goal is to build a circuit to suppress the transformation of common mode noise to differential mode noise.

Common Mode Rejection

To describe the circuit behaviour of suppressing the common mode signal we introduce Common Mode Rejection (CMR), measured in dB (decibel), given in (5.4)

$$CMR^{(dB)} = 20 \log \frac{U_{CMN}}{U_{DMN}} \tag{5.4}$$

We see the common mode rejection of the asymmetric connection of the source and the receiver, illustrated in Fig. 5.25, is close to the worst value of 0 dB. That is in accordance with the rule of thumb, discussed in section '5.1.5 Signal source to receiver interconnection'.

Common Mode Rejection of Symmetric Input

For the circuit analysis we have the connection, shown in Fig. 5.26. The receiver has a symmetric input and source is grounded with offset.

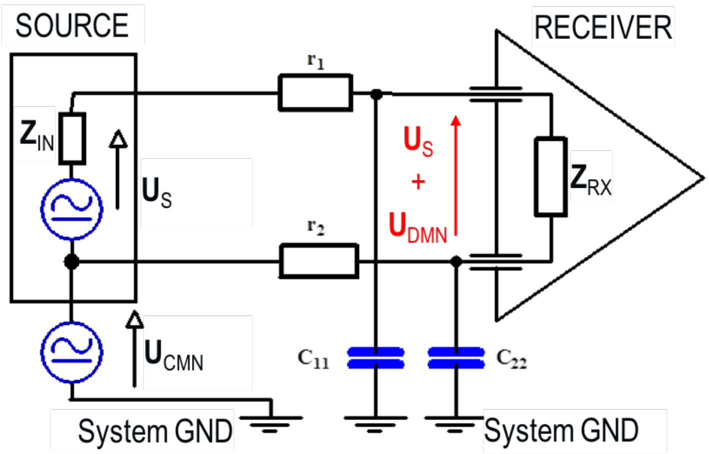


Figure 5.26 CMR of symmetric input

We can introduce the following equivalent impedances, described in (5.5) and (5.6) for convenient calculation.

$$\boldsymbol{Z}_1 = \boldsymbol{Z}_{IN} + r_1, \qquad \boldsymbol{Z}_2 = r_2 \tag{5.5}$$

$$\boldsymbol{Z}_{11} = \frac{1}{j\omega C_{11}} \approx \boldsymbol{Z}_{22} = \frac{1}{j\omega C_{22}} \tag{5.6}$$

With these impedances the equivalent circuit of Fig. 5.26 from the effect of a common mode voltage point of view can be seen in Fig. 5.27.

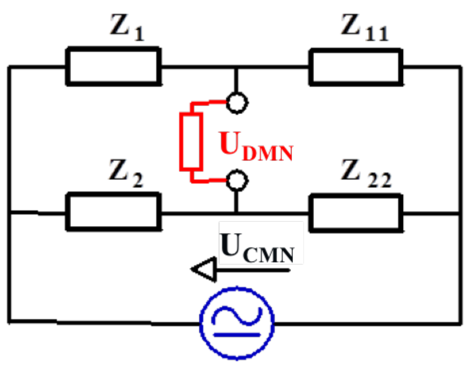


Figure 5.27 The bridge equivalent

With further simplification we introduce Z and ΔZ impedances as defined in (5.7).

$$\boldsymbol{Z}_{11} = \boldsymbol{Z}_{22} = \boldsymbol{Z}, \qquad \Delta\boldsymbol{Z} = \boldsymbol{Z}_2 - \boldsymbol{Z}_1 \tag{5.7}$$

where Z is so called *common mode impedance* and ΔZ is *impedance of wire asymmetry*.

According to the bridge circuit in Fig. 5.27 we can write (5.8)

$$\frac{\boldsymbol{U}_{DMN}}{\boldsymbol{U}_{CMN}} = \frac{\Delta\boldsymbol{Z}}{\boldsymbol{Z}} \tag{5.8}$$

thus, the common mode rejection is

$$CMR^{(dB)} = 20\log\frac{U_{CMN}}{U_{DMN}} = 20\log\frac{Z}{\Delta Z} \tag{5.9}$$

Methods for Increasing Common Mode Rejection

Because our goal is to decrease the differential mode noise caused by common mode noise we have to examine this effect to find out what kind of interventions are necessary for increasing common mode noise rejection. First of all, the best way for this is to *eliminate double grounding*. The second idea is from (5.9), as a decrease in ΔZ will increase CMR. Because $\Delta Z = 0$ means symmetric realization, our goal is to design and apply symmetric receiver input as much as possible. Also from (5.9) we see that increasing the Z impedance results in a higher CMR. So, our goal is to design and apply a receiver with high common mode input impedance, that can be established with a low parasitic capacitance, as seen in Fig. 5.26. One possible method for decreasing each parasitic capacitance between input terminals and ground is shown in Fig. 5.28. In this case protective shielding is applied that is connected to Guard. The C_{11} and C_{22} parasitic capacitances are formed between input terminals and internal shielding. There is an additional C_3 parasitic capacitance between shielding and system ground (signed RF in Fig. 5.28).

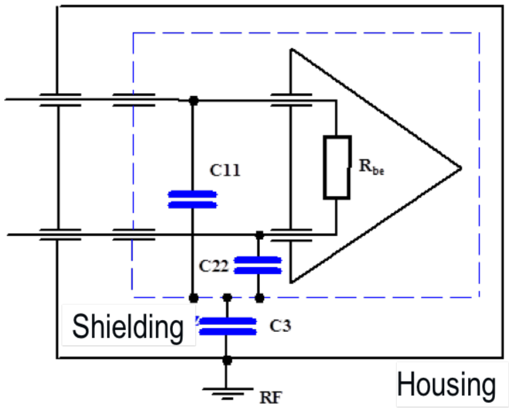


Figure 5.28 Effect of protective shielding

Because C_3 is connected in series with both input parasitic capacitances, as shown in Fig. 5.29, the equivalent parasitic capacitance will be lower and so the common mode input impedance will be higher when applying shielding in the receiver.

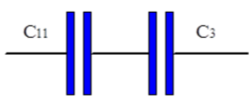


Figure 5.29 Series capacitors

5.2.3 Noise Sources

The type of intervention we use against electric noise, strongly depends on the source or ‘reason’ for the noise. The most common sources of noise are summarised below.

When different kinds of conductive materials are in galvanic connection, a so-called *contact potential* is formed at the deflection surface as local ‘Galvan cell’. The protection against noise voltage, caused by contact potential, is to minimize the contact of different metals thus preventing corrosion.

The second common source of noise is the *thermal potential* that comes from contact of different metals. This contact causes temperature dependent noise as differential mode noise in the system and can be reduced by minimising the number of contacts of different metals.

Transmission resistance is a very common source or reason for electric noise in systems. It is caused by unsafe contacts in switches. The method for prevention is to apply reliable materials for switch contacts and clean them regularly.

We use the common name of *electromagnetic disturbances* to describe the coupling channel from any external (disturbing) electromagnetic field. To decrease the effect of coupling we can use a twisted pair of wires for the test cable, a proper grounding system, and also magnetic shielding of the protected system.

Electric disturbances are caused by parasitic capacitances in a system creating common mode and/or differential mode noise. The best prevention is to separate the signal channel from the source of noise as far as it is possible. One solution is to apply grounded electrostatic shielding.

The common name of *switch ON/OFF disturbances* is used for the transient noise caused by conductive, capacitive, or inductive coupling when opening an inductive circuit or short circuiting a capacitive circuit. To avoid this, either overvoltage protection or current limitation can be applied as protection.

Due to the piezoelectric effect in insulations *cable bending disturbances* have to be taken into consideration as an additional source of electric noise. This noise occurs in relation to cable bending as well as shield bending. To avoid the harmful effects of cable bending disturbances, the design and application must be in compliance with bending requirements.

Power line origin disturbances are caused when, for example, parasitic capacitances cause coupling between the primer and the seconder coils of the main transformer. As the result of this coupling, noise is superposed to the measuring signal. An example of such superposition is shown in Fig. 5.30.

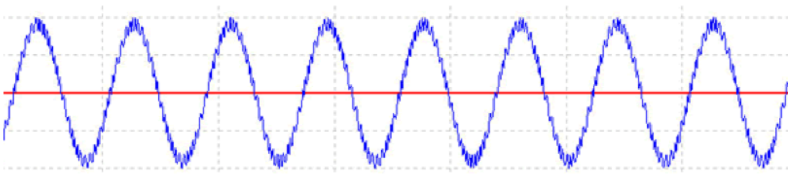


Figure 5.30 Signal with superposed noise component

The classical prevention against this disturbance is an electrostatic shielding, by applying a grounded row of copper between the primary and secondary windings.

5.3 Groundings and Earths

5.3.1 Definitions

Electric circuits need a common reference potential to be able to work together in their interconnection. *Ground*, often abbreviated with GND, is a *voltage reference point* in electronics and electrical engineering. Ground potential is supposed to be zero, i.e. 0 V.

Technically, the term 'ground' refers to a physical connection to the Earth. It can be asked, why this common reference point is connected to Earth? The answer is; the Earth's potential at a certain point is almost constant because of the high electrical capacity of the Earth.

The general electrical symbol for ground is shown in Fig. 5.31. It is a bit confusing this symbol is used for both earth ground, and common ground. While *common ground* is interpreted as the reference point in an electrical circuit from which voltages are measured, *earth ground* is a direct physical and electrical connection to the earth by applied copper, aluminium or an aluminium alloy conductor.

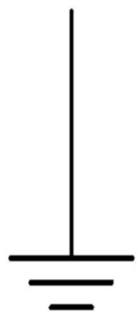


Figure 5.31 General symbol for ground

According to the definition by the National Electrical Code (NEC), an earth ground is a conductive pipe or rod that is physically driven into the earth to a minimum depth of 8 feet (approx. 2.5 m). A simple implementation of an earth ground is shown in Fig. 5.32.

Figure 5.32 Earth ground implementation

Besides the common reference point, ground potential is also used in electric power networks for life protection. Connection of the Protective Earth (PE), even in a single-phase power transmission system, can have different forms as shown in Fig. 5.33. While the solution on the left demonstrates sockets in USA, the picture on the right shows the EU standard socket with protective earth.

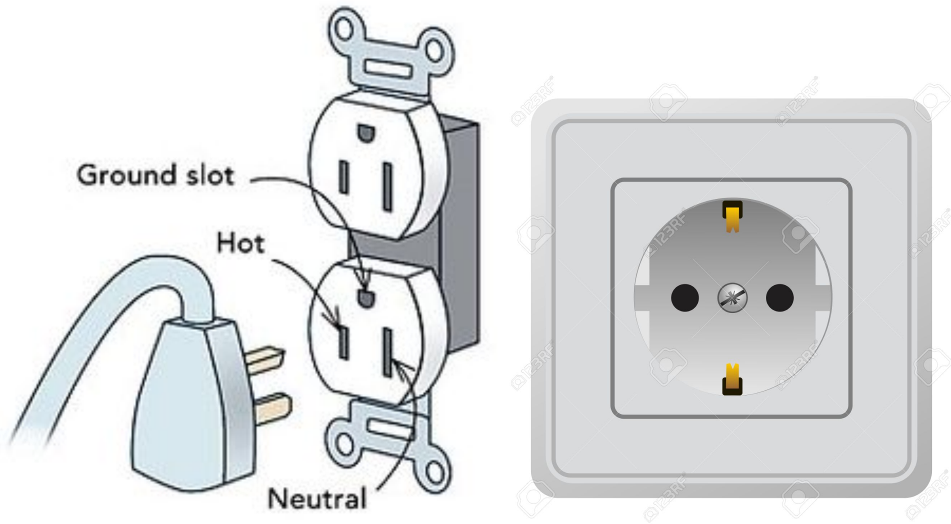


Figure 5.33 Connection of the PE conductor in different countries

It is important to mention, that not all voltage measurements are about grounding in electrical circuits. Fig. 5.34 illustrates the voltage across the upper resistance is measured by applying, a so called *ground independent voltmeter*, meaning the symmetric input of this device.

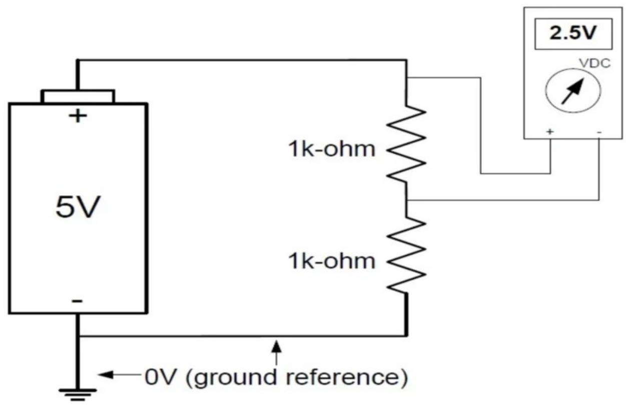


Figure 5.34 Ground independent voltage measurement

5.3.2 Ground Types

The following section distinguishes between different types of grounding in more detail.

Signal Ground

This is a reference point from which the signal is measured when, due to inevitable voltage drops when the current flows within a circuit, some 'ground' points will be slightly different to others. There may be several signal grounds in a circuit. Imagine if you had an amplifier with a voltage gain of 100 and you were amplifying a tiny signal, if the ground for the signal was elevated by just 0.01V the output would be wrong by 1V. Typically a signal ground would be a connected to the same stage of the circuit as the signal was connected. Two main groups of signal grounds are *Analogue Signal GrouND* (A-GND) and *Digital Signal GrouND* (D-GND) as reference points for analogue and digital systems.

Chassis Ground

This is the box or frame in which a circuit is built. Typically, it is earthed to create a barrier between the user and the circuits inside to prevent electric shock or to shield against interference or radiation. In some high current applications, it is used as a conductor to carry current, for example in a vehicle where running many thick wires to the battery would be impractical, making a connection to the chassis is relatively easy almost anywhere. Chassis grounds should be connected to the other grounds; usually done at a point close to where the power source arrives.

Earth Ground

This is a theoretical zero. It is the potential of perfectly conducting Earth beneath your feet but obviously it varies widely depending on where you are. The connection to Earth is normally through the power network or by a rod driven into the ground or sometimes both. It is supposed to be a return path for current in the event of a short circuit on your AC power lines, but it is also used in radio applications as a 'zero' reference to read antenna signals against.

Protective Earth

Protective earth (PE) is a grounded conductor connected to the normally non-voltage metal parts of electrical equipment. It has a life protection function and is forbidden to use in normal operation. No normal operation return current is allowed to flow through the PE conductor.

Power Ground

Power ground or common ground is the common point of a power supply. This ground line is designed for conducting normal operation return current.

System Ground

System GrouND (S-GND) is a single common point of other grounds to be connected to the earth potential. It is important to know that different types of grounds must be connected to each other to have a common reference and combining grounds must be done at a single point as illustrated in Fig. 5.35.

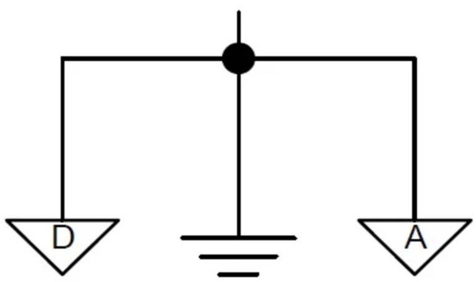


Figure 5.35 Common point of different grounds

As a rule of thumb, to avoid a '*noisy ground system*', noisier digital return currents have to be separated from less-noisy analogue return currents. The common point of different grounds should be a *star ground* or *grounded bus bar* at least, even if a star ground is hard to implement.

Neutral Wires, Earth and Ground Wires in a Power Line System (Example)

According to the rules of correct ground connections, discussed in the previous section, the following example gives the possible implementation of grounds in a power line system. The so-called TN-S connection, shown in Fig. 5.36, is widely used in power transmission systems. Note that detailed discussion of terminology standards recommended by the International Electrotechnical Commission (IEC) goes beyond the scope of this book but a short explanation of the TN-S abbreviation is the following. In TN systems the Earth connection is supplied by the electricity supply network, either as a separate protective earth (PE) conductor or combined with the neutral (PEN) conductor. In TN-S systems PE and N are separate conductors that are connected together only near the power source. In Fig. 5.36 a single-phase load with an exposed conductive part is connected to the power line system i.e. to the line (L), neutral (N) and protective earth (PE) conductors.

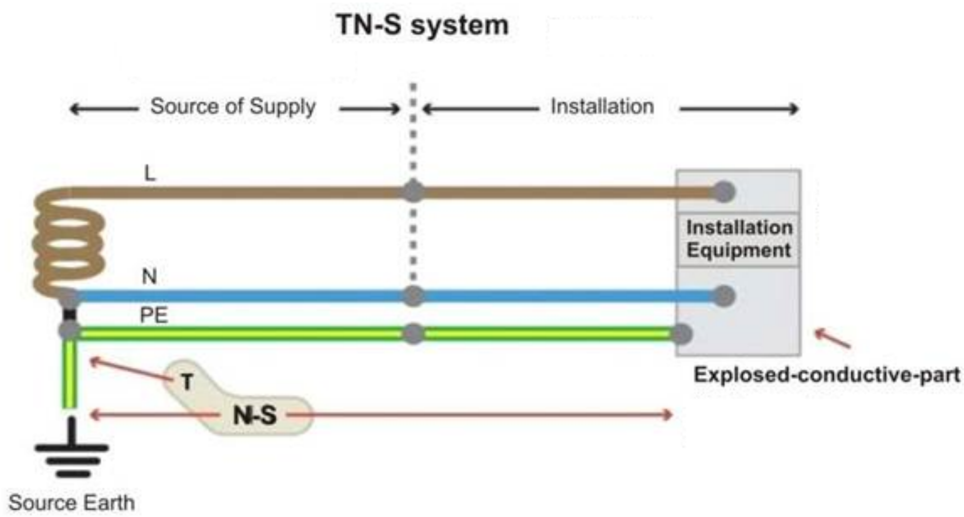


Figure 5.36 TN-S connection system

The detailed explanation of the TN-S system is out of the scope of this material, but the main rules of proper grounding can be easily followed in Fig. 5.36. The neutral wire as the system ground and protective earth ground are connected in single point at the generator side. This common point as a 'star point' is connected to the earth ground. Metal housing of the installation equipment is connected to the protective earth line with no connection to the operational part of the equipment. No operational return current runs through the PE line.

The neutral line is the return path for an AC circuit, so it is designed for carrying return current under normal operational conditions. The neutral wire is always assumed to be charged but has zero potential hence is connected to ground.

Protective earth is used for *safety* reasons against leakage or residual currents through the path of least resistance. It may be connected to the body of the installation equipment and in case of insulation failure it is supposed to carry a minor current, for short period.

5.3.3 Grounding Symbols

The general symbol of ground has already been shown in Fig. 5.31. But, because of the previously mentioned different types of ground, standardized distinctions in definitions and how they are symbolised are also necessary. For example, the most commonly applied ground symbols are given in Fig. 5.37. More details about grounds in general, low-noise grounds, safety and protective grounds, chassis or frame connections, are given in Standards of IEEE 315 and IEC 60417.

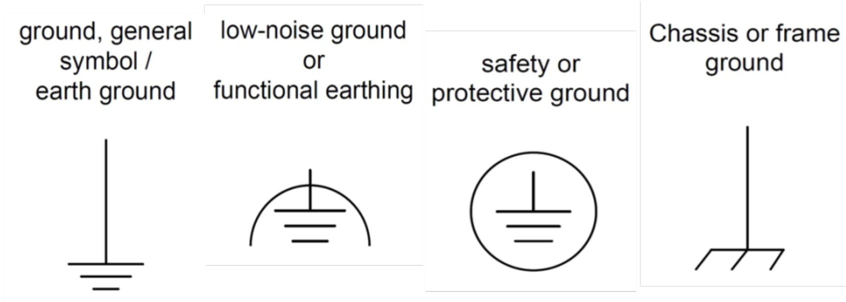


Figure 5.37 The most commonly used ground symbols

In many cases the ground symbol, given in Fig. 5.38, is used for specifying a common potential that is not zero, or simply for clear distinction, according to referred IEEE 315 standard.

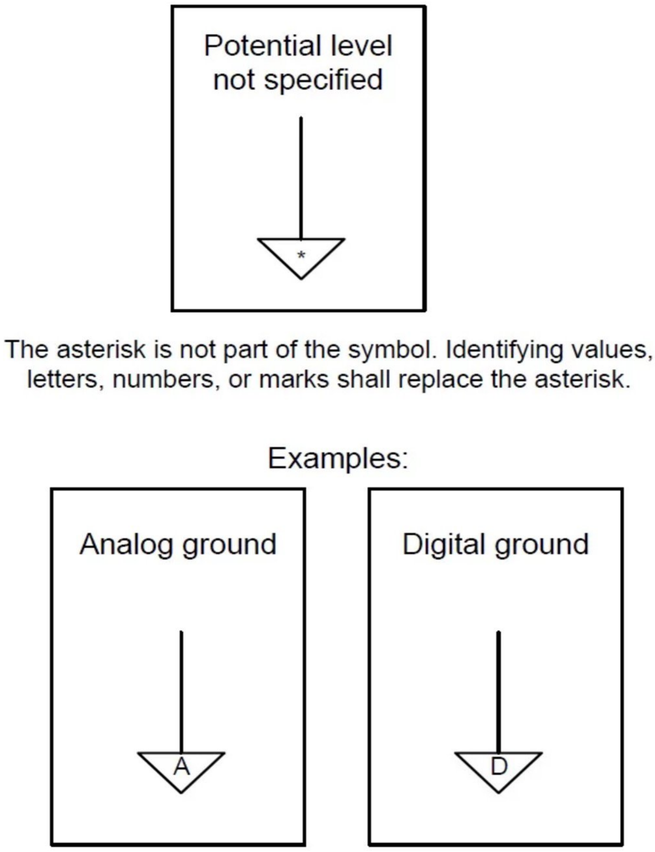


Figure 5.38 Ground symbol with specified ground type or potential level

5.3.4 The Ground Loop

The ground loop, as the harmful effect of double grounding, has already been mentioned previously in this chapter. We have seen that ground loop transforms common mode noise to differential mode noise, causing a significant deterioration in noise suppression.

Fig. 5.39a illustrates the ground loop effect when the sensor (signal source) is in galvanic connection to the ground. The solution is simple in this case, as seen in Fig. 5.39b. The sensor is insulated by using an insulated mount allowing the grounded source and grounded receiver to be connected to each other. This means that no differential mode noise is caused by common mode noise due to different ground potentials.

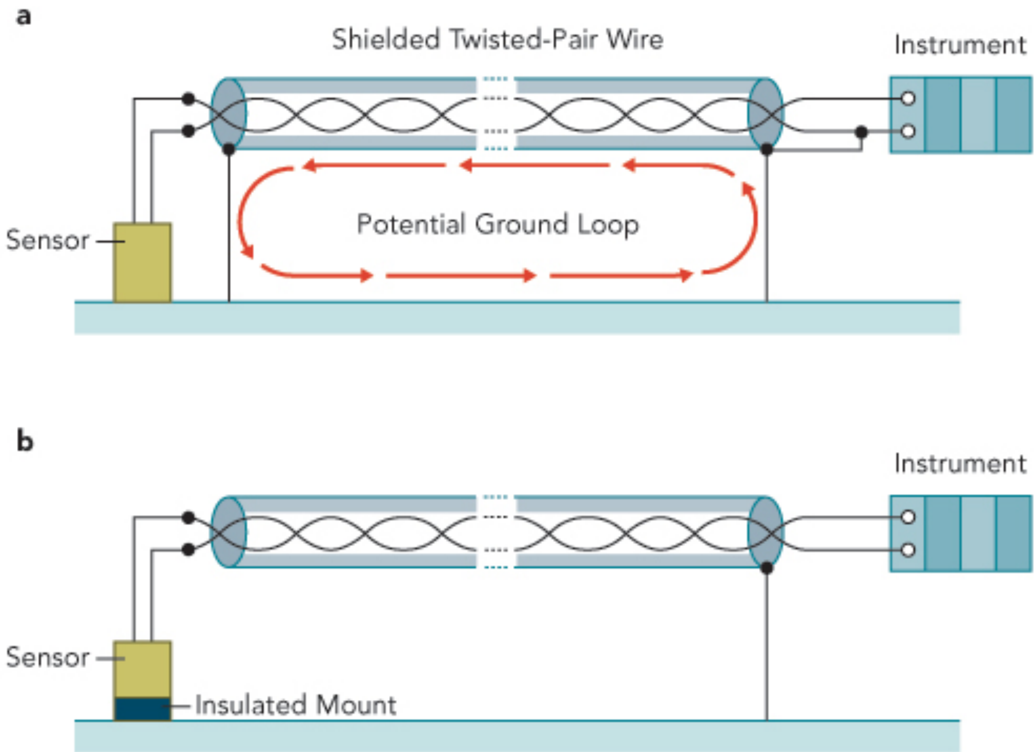


Figure 5.39 Ground loop caused by incorrect sensor mounting

To eliminate the ground loop is not always as easy as it was in the previous example. In the next example, shown in Fig. 5.40, two electronics i.e. asymmetric amplifiers (*C1* and *C2*) are cascade-connected by a shielded (coax) lead. Terminals *L* (low) are connected to the housing and to PE (protective earth). The goal is to 'open' ground loop, while neither cutting the PE conductor nor the L line of the signal connection.

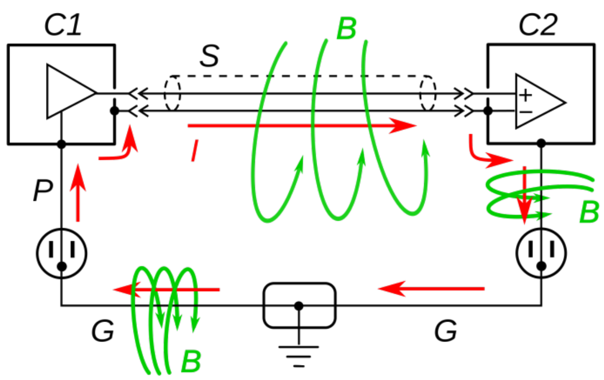


Figure 5.40 Ground loop caused by protective earth connections

The only possible way to apply an additional element in signal connection is to eliminate the galvanic connection in the ground loop with the preserved signal connection. Without the need for completeness some of the practical methods are given in the following.

Flying Capacitor

The capacitor is 'analogue memory' as it remembers the voltage it was connected to previously. When the terminals of a capacitor are connected to the source, as shown in Fig. 5.41, the capacitor will be charged to U_S voltage (through r_1, and r_2 small resistors). Sudden switching of the capacitor terminals to the receiver input, will result in an input voltage of U_S. By repeating this switch function continuously, the source voltage is copied to the receiver input with no galvanic connection between the source and receiver. In case of low additional parasitic capacitances, a very high common mode rejection is achieved and the is no ground loop in the circuit.

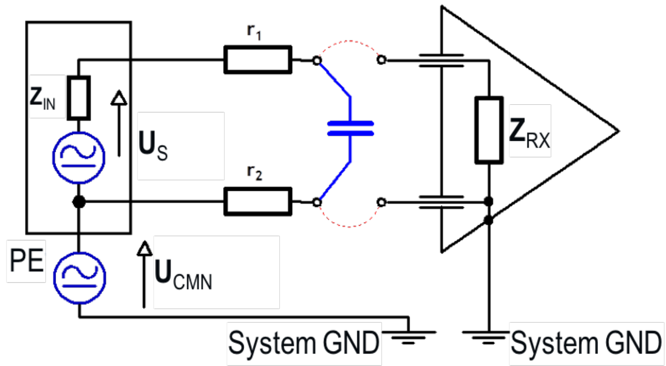


Figure 5.41 Insulation principle by applied flying capacitor

Isolation Transformer

Another way to 'open' the ground loop is to apply an isolation transformer in the signal connection as shown in Fig. 5.42. Because this solution is relatively cheap and simple it is widely used for eliminating ground loops if the possible for nonlinearity of the transformer is acceptable.

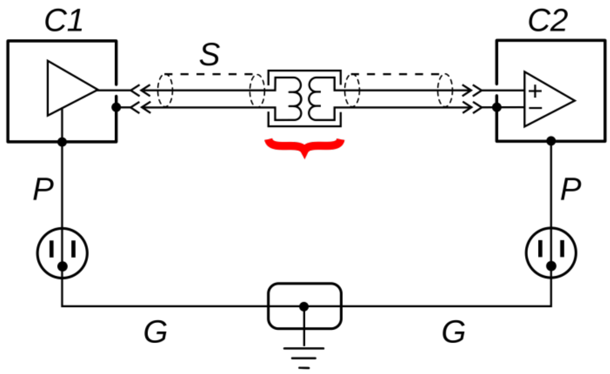


Figure 5.42 Isolation transformer in signal connection

Optical Insulation

In electronics, an optical isolator, also called an opto-isolator, optocoupler, or photocoupler, is a component that transfers electrical signals between two isolated circuits by using light. A common type of opto-isolator consists of a LED-photodiode and a phototransistor in the same opaque package. Other types of source-sensor combinations include LED-photodiode, LED-Darlington circuit. See Fig. 5.43. Usually opto-isolators transfer digital (on-off) signals, but some techniques allow them to be used with analogue signals.

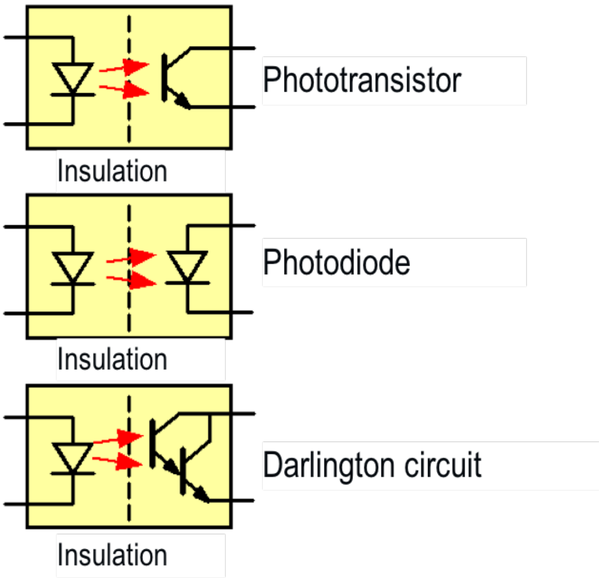


Figure 5.43 Different type of opto-isolators

Because opto-isolators have significant linearity error or low dynamic band in case of analogue signal transmission an improved solution is applied as shown in Fig. 5.44. The source signal is connected to a voltage-frequency converter to control the opto-isolator using frequency instead of amplitude, eliminating its nonlinearity. On the receiver side the signal has to be converted back using a frequency-voltage converter.

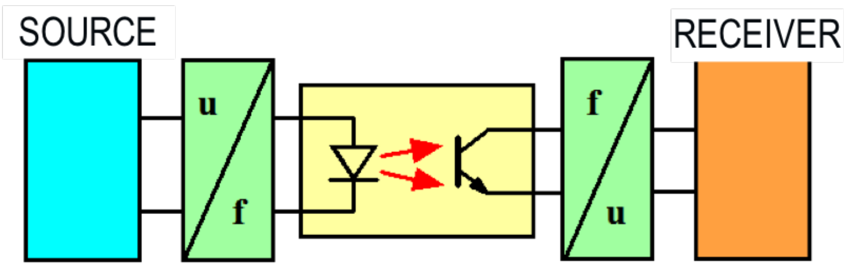


Figure 5.44 Improved opto-isolator circuit for analogue signal transmission

6. Sensor Measurement Solutions

6.1 Temperature Sensors

Temperature is defined as the energy level of matter, which can be evidenced by some change in that matter. Temperature measuring sensors come in a wide variety and have one thing in common: they all measure temperature by sensing some change in a physical characteristic.

Measuring methods can be divided into those using the principle of thermal conductivity, called *contact sensors*, and sensors using the principle of thermal radiation, *pyrometers*. Thermal sensors can also be classified according to the necessity of additional energy sources. Active sensors do not need an external source for their operation because they produce an electrical voltage (or current) that is proportional to the temperature they measure.

Passive sensors produce a change in an electrical parameter i.e. in resistance, proportional to the measured temperature. To transform this resistance to electric voltage or current we need an external source to supply energy.

The basic types of temperature measuring sensors discussed here are thermocouples, Resistive Temperature Devices (RTDs), thermistors, silicon diodes, and infrared detectors or pyrometers. Different sensors have different features in their sensitivity, linearity and also in operation range. For example, Fig. 6.1 demonstrates the different temperature dependent behaviour of thermocouples, resistive temperature devices and thermistors with negative temperature coefficients.

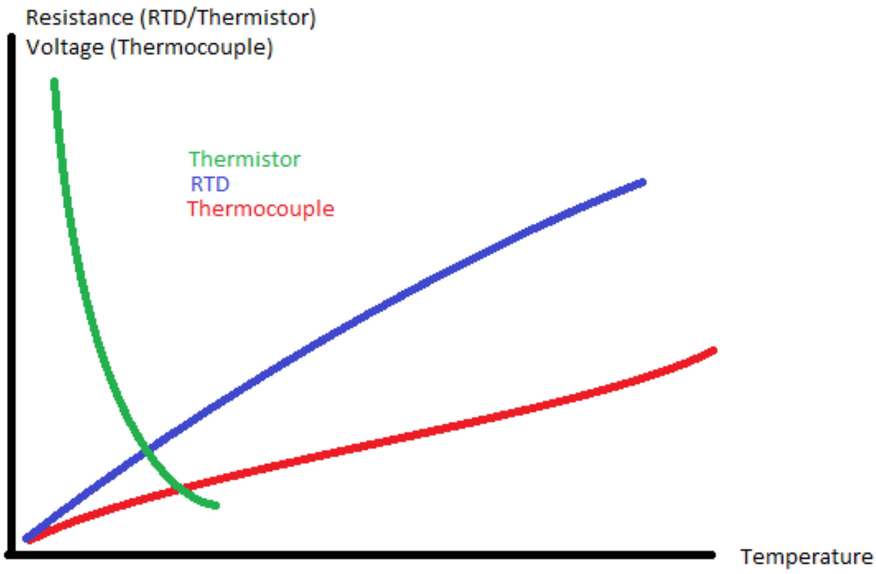


Figure 6.1 Comparison of thermistors, RTDs and thermocouples

6.1.1 Thermocouples

Thermocouples are voltage devices that indicate temperature with a change in voltage. As temperature goes up, the output voltage of the thermocouple rises – but not necessarily linearly.

Often thermocouples are located outside a metal or ceramic shield that protects it from exposure to a variety of environments. Metal-sheathed thermocouples also are available with many types of outer coatings, such as Teflon, for trouble-free use in acids and strong caustic solutions. Thermocouples measure voltage change and signify temperature.

The principle of operation is based on the temperature dependent contact potential, given in (6.1) and formed between two connected conductive materials, i.e. metal

$$U = k \cdot T \tag{6.1}$$

According to the principle of operation, thermocouples are active sensors because they provide an electrical voltage with no need of an additional energy source.

Table 6.1 shows an example of the temperature dependency of two practical thermocouples. We can see the output voltage is measured in millivolts which is why thermocouples are usually connected to amplifier for further signal processing. Although thermocouples are active sensors, so theoretically, they do not require external energy to work, in practice we usually apply electronic amplifiers which require an external source of energy.

Table 6.1 Contact potentials for Fe-CuNi and NiCr-Ni thermocouples

T (°C)	$U_{Fe\text{-}CuNi}$ (mV)	$U_{NiCr\text{-}Ni}$ (mV)
-200	-8.15	
-100	-4.75	
0	0	0
100	5.37	4.10
200	10.95	8.13
300	16.56	12.21
400	22.16	16.40
500	27.85	20.65
600	33.67	24.91
700	39.72	29.14
800		33.30
900		37.36
1000		41.31

If we have connected the leads, made of different metals to the voltmeter, measurements would be unreliable because of additional contact potentials at the voltmeter connectors, because of the unknown and uncontrolled ambient temperature. The solution for this problem is given in Fig. 6.2, showing the 'relocated' unavoidable additional contact potentials to the controlled, using a known reference temperature. This reference point can be achieved by melting ice, or by the application of an electronic compensation device.

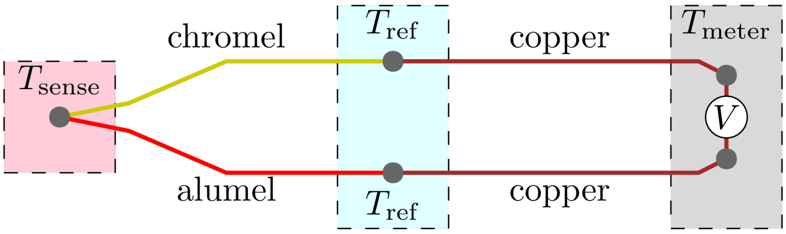


Figure 6.2 Temperature measurement with thermocouples

The contact potentials of the sensors and reference points are given by (6.2).

$$U_{sen} = k_{sen}T_{sen}\,, \qquad U_{ref1} = k_{ref1}T_{ref}\,, \qquad U_{ref2} = k_{ref2}T_{ref} \tag{6.2}$$

The voltage measured by the voltmeter is

$$U_{meter} = U_{sen} - \left(U_{ref1} + U_{ref2}\right) = k_{sen}T_{sen} - (k_{ref1} + k_{ref2})T_{ref} \tag{6.3}$$

Because the *k* coefficients of the contacted metals and the reference temperature are known, the relationship between the T_{sen} temperature to be measured and the U_{meter} voltage is well defined. See (6.4).

$$U_{meter} = k_{sen}T_{sen} - k_{ref}T_{ref} \tag{6.4}$$

6.1.2 Resistive Temperature Measuring Devices

Resistive temperature measuring devices are also electrical. Rather than using a voltage as the thermocouple does, they take advantage of another characteristic of matter which changes with temperature - its resistance. According to the principle of operation this sensor is passive because it needs an additional energy source to transform resistance change to voltage (or current) change. In general, RTDs are more linear than are thermocouples. They increase in a positive direction, with resistance going up as temperature rises. RTDs have *Positive Temperature Coefficients*, so they are also referred to as PTC sensors.

The most popular RTD sensor is the Pt-100 platina resistor that is (more or less) the 'world standard' in electrical temperature measurement. The Pt-100 can be used in a wide temperature range with excellent linearity. Its static characteristic is given by the function in (6.5) and (6.6), where ϑ is the temperature in °C, i.e. the temperature difference from 0 °C. The equation in (6.5) is applicable between 0 and 850 °C and (6.6) is for temperatures between -200 and 0 °C.

$$R(\vartheta) \cong 100(1 + 3.9\;10^{-9}\,\vartheta - 0.58\;10^{-6}\,\vartheta^2)\;\Omega\,, \qquad 0\;°\text{C} \le \vartheta < 850\;°\text{C} \tag{6.5}$$

$$R(\vartheta) \cong 100\begin{bmatrix} 1 + 3.9\;10^{-9}\vartheta \;\; - 0.58\;10^{-6}\,\vartheta^2 \\ -4.27\;10^{-12}(\vartheta - 100)\,\vartheta^3 \end{bmatrix}\;\Omega\,, \qquad -200\;°\text{C} \le \vartheta < 0\;°\text{C} \tag{6.6}$$

In practice the Ni-100 nickel resistor is also used for the same measurement purposes but at a lower temperature range because of lower linearity. Their characteristic is approximated as the following.

$$R(\vartheta) \cong 100\begin{pmatrix} 1 + 5.4\,10^{-3}\,\vartheta + 6.6\,10^{-6}\,\vartheta^2 \\ +2.8\,10^{-11}\,\vartheta^4 \end{pmatrix}\Omega\,, \qquad -60\,^{\circ}\mathrm{C} \leq \vartheta < 180\,^{\circ}\mathrm{C} \quad (6.7)$$

Although, the Pt-100 has a lower sensitivity than the Ni-100, the platina resistor is more commonly used due to its better linearity and its greater measurement range.

6.1.3 Thermistors

The thermistor is a special type of variable resistive element that changes its physical resistance when exposed to changes in temperature. It has an entirely different type of construction than the RTD resistor. It is an extremely nonlinear semi conductive device that can also decrease in resistance as temperature rises, based on its construction. The electrical symbol of the thermistor is given in Fig. 6.3.

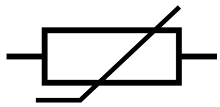


Figure 6.3 Symbol of the thermistor

While RTD elements are simple metal resistors, thermistors are semiconductor based passive sensors, made of metal oxides from the IV. main group of periodic system, like chromium, manganese, iron, cobalt, nickel, titanium, etc. Resistance of a thermistor can have a negative or positive temperature coefficient, depending on the applied materials. These heat-dependent resistors can operate in one of two ways, either increasing or decreasing their resistive value with changes in temperature which is why there are two types of thermistors available: Negative Temperature Coefficient (NTC) of resistance and Positive Temperature Coefficient (PTC) of resistance.

Negative Temperature Coefficient Thermistor

Negative temperature coefficient of resistance thermistors, or *NTC thermistors* for short, reduce or decrease their resistive value as the operating temperature around them increases. Generally, NTC thermistors are the most commonly used type of temperature sensors as they can be used in virtually any type of equipment where temperature plays a role.

NTC temperature thermistors have a negative electrical resistance versus temperature (*R*/*T*) relationship. The relatively large negative response of an NTC thermistor means that even small changes in temperature can cause significant changes in its electrical resistance. This makes them ideal for accurate temperature measurement and control.

We stated previously that a thermistor is an electronic component whose resistance is highly dependent on temperature so if we send a constant current through the thermistor and then measure the voltage drop across it, we can determine its resistance and temperature.

NTC thermistors are usually characterized by their base resistance (R_0) at room temperature that is 25 ^{o}C, as this provides a convenient reference point.

Another important characteristic is the 'B' value. The B value is a material constant, which is determined by the ceramic material from which it is made and describes the gradient of the resistive (*R/T*) curve over a particular temperature range between two temperature points. Each thermistor material will have a different material constant and therefore a different resistance versus temperature curve. However, a 'B' material constant between 2000 K and 5000 K is used for practical NTC thermistors.

The characteristic equation of a thermistor is given in (6.8).

$$R(T) = R_0 e^{B\left(\frac{1}{T}-\frac{1}{T_0}\right)} \tag{6.8}$$

Where:

- T_0 is the first temperature point in Kelvin;
- T is the second temperature point in Kelvin;
- R_0 is the thermistors resistance at temperature T_0 in Ohms;
- R is the thermistors resistance at temperature T in Ohms.

Thermistors change their resistance exponentially with changes in temperature, so their characteristic curve is nonlinear. For example, the characteristics curve for the 10 kΩ NTC thermistor which has a B-value of 3455 is shown in Fig. 6.4.

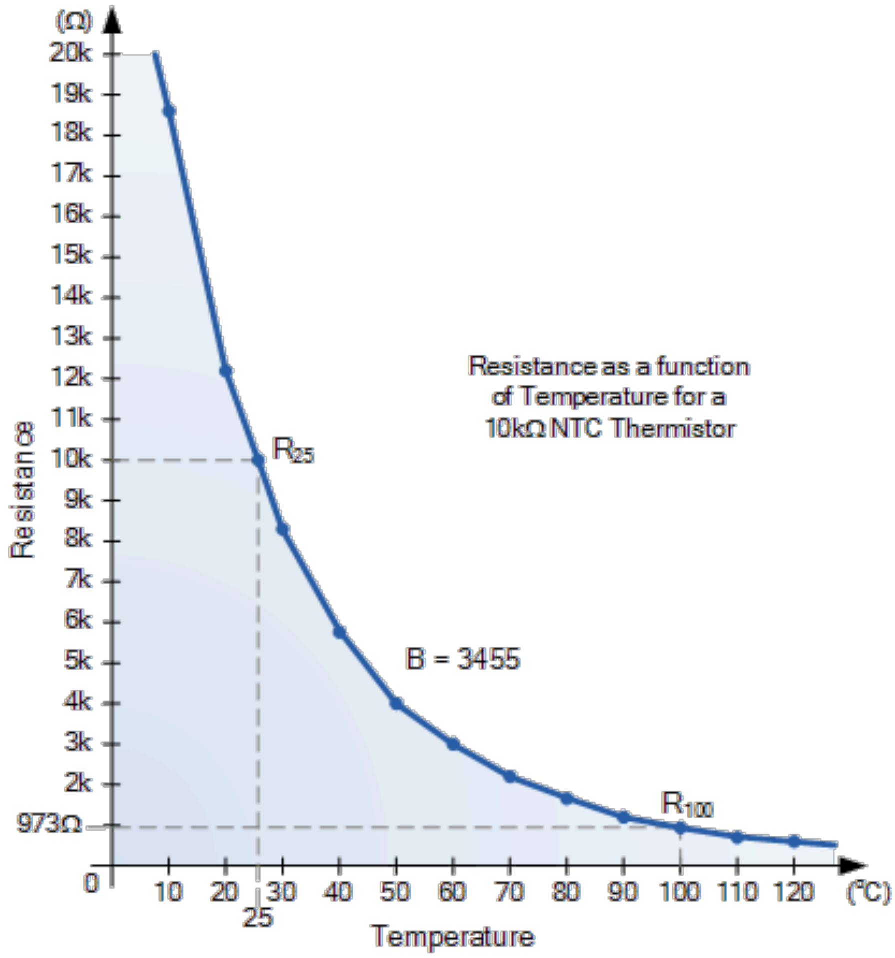


Figure 6.4 Temperature dependent resistance of an NTC thermistor

Temperature *sensitivity* of an *NTC thermistor* is calculated in (6.9) by substituting *R*(*T*) from (6.8).

$$S = \frac{\partial R(T)}{\partial T} = \frac{\partial R_0 e^{B\left(\frac{1}{T}-\frac{1}{T_0}\right)}}{\partial T} = -\frac{B_{Th}}{T^2} R \qquad (6.9)$$

We can see that the lower the temperature, the higher the sensitivity for NTC thermistors.

Both, the PTC and NTC thermistor have much better sensitivity than RTD resistors. Because of the possible PTC and NTC temperature dependent behaviour they are widely used in electrical and electronic circuits i.e. for temperature compensation in generators (oscillators), for temperature measurement and for overcurrent or overvoltage protection instead of using fuse. A practical realization of a temperature measuring device is shown in Fig. 6.4.

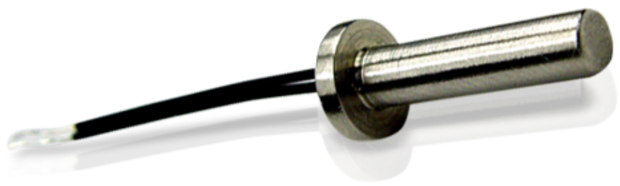


Figure 6.5 Thermistor for temperature measurement

So how can we use a thermistor in a practical circuit to measure temperature? Hopefully by now we know that a thermistor is a resistive device and therefore according to Ohms law, if we pass a current through it, a voltage drop will be produced across it. As a thermistor is a passive type of a sensor, that is, it requires an excitation signal for its operation, any changes in its resistance as a result of changes in temperature can be converted into a voltage change.

The simplest way of doing this is to use the thermistor as part of a potential divider circuit as shown in Fig. 6.5. A constant voltage is applied across the resistor and thermistor series circuit with the output voltage measured across the thermistor.

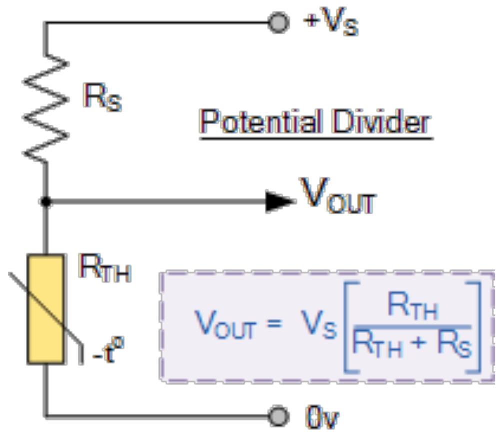


Figure 6.5 Temperature measurement by applying the thermistor

If for example we use a 10 kΩ thermistor with a series resistor of 10 kΩ, then the output voltage at the base temperature of 25 °C will be half the supply voltage.

When the resistance of the thermistor changes due to changes in temperature, the fraction of the supply voltage across the thermistor also changes producing an output voltage that is proportional to the fraction of the total series resistance measured between the output terminals.

Thus, the potential divider circuit is an example of a simple resistance to voltage converter where the resistance of the thermistor is controlled by temperature with the output voltage produced being proportional to the temperature. So, the hotter the thermistor gets, the lower the voltage.

If we reversed the positions of the series resistor, R_S and the thermistor, R_{TH}, then the output voltage will change in the opposite direction, that is the hotter the thermistor gets, the higher the output voltage.

We can use NTC thermistors as part of a basic temperature sensing configuration using a bridge circuit as shown in Fig. 6.6.

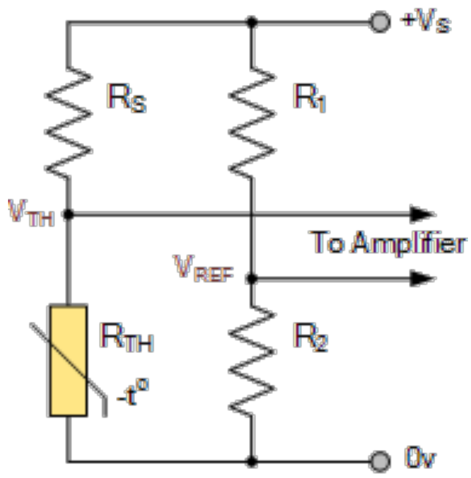


Figure 6.6 Temperature measurement using bridge circuit

The relationship between resistors R_1 and R_2 sets the reference voltage, V_{REF} to the value required. For example, if both R_1 and R_2 are of the same resistive value, the reference voltage will be equal to half of the supply voltage. That is $V_S/2$.

As the temperature and therefore the resistance of the thermistor changes, the voltage at V_{TH} also changes to either higher or lower than that at V_{REF} producing a positive or negative output signal to the connected amplifier. (The amplifier circuit used for this basic temperature sensing bridge circuit could act as a differential amplifier for high sensitivity and amplification, or a simple Schmitt-trigger circuit for ON-/OFF switching.)

The problem with passing a current through a thermistor in this way, is that thermistors experience what is called self-heating effects, that is, the $I^2 \cdot R$ power dissipation could be high enough to create more heat than can be dissipated by the thermistor affecting its resistive value producing false results.

Thus, it is possible that if the current through the thermistor is too high it would result in increased power dissipation and as the temperature increases, its resistance decreases causing more current to flow, which increases the temperature further resulting in what is known as *Thermal Runaway*. In other words, we want the thermistor to be hot due to the external temperature being measured and not by heating itself up.

Then the value for the series resistor, R_S above, should be chosen to provide a reasonably wide response over the range of temperatures for which the thermistor is likely to be used while at the same time limiting the current to a safe value at the highest temperature.

One way of improving on this and having a more accurate conversion of resistance against temperature (*R/T*) is by driving the thermistor with a constant current source. The change in resistance can be measured by using a small and measured direct current, or DC, passed through the thermistor in order to measure the voltage drop produced.

Thermistor Used for Inrush Current Suppression

We have seen that thermistors are primarily used as resistive temperature sensitive transducers, but the resistance of a thermistor can be changed either by external temperature changes or by changes in temperature caused by an electrical current flowing through them, as after all, they are resistive devices.

Ohm's Law tells us that when an electrical current passes through a resistance *R*, as a result of the applied voltage, power is consumed in the form of heat due to the $I^2 \cdot R$ heating effect. Because of the self-heating effect of the current in a thermistor, a thermistor can change its resistance with changes in current.

Inductive electrical equipment such as motors, transformers, ballast lighting, etc, suffer from excessive inrush currents when they are first turned-on. A series of connected thermistors can be used to effectively limit these high initial currents to a safe value. NTC thermistors with low values of cold resistance (at 25 ^oC) are generally used for current regulation.

Inrush Current Limiting Thermistor

Inrush current suppressors and surge limiters are types of series connected thermistors whose resistance drops to a very low value as it is heated by the load current passing through it. At the initial turn-on, the thermistors cold resistance value (its base resistance) is fairly high controlling the initial inrush current to the load.

As a result of the load current, the thermistor heats up and reduces its resistance relatively slowly to the point where the power dissipated across it is sufficient to maintain its low resistance value with most of the applied voltage developed across the load. The NTC thermistor, connected series to the load and its effect to the load current is shown in Fig. 6.7.

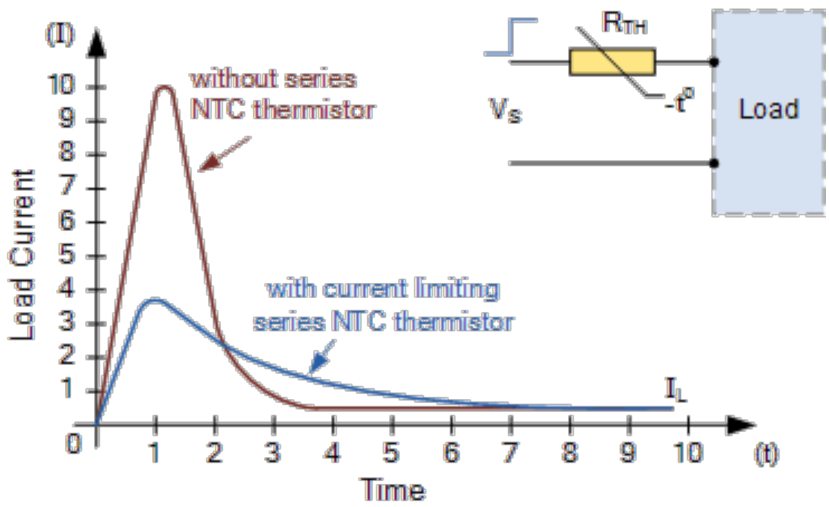


Figure 6.7 Inrush current limiting thermistor

Due to the thermal inertia of its mass this heating effect takes a few seconds, during which the load current increases gradually rather than instantaneously, so any high inrush current is restricted and the power it draws reduces accordingly. Because of this thermal action, inrush current suppression thermistors can run very hot in the low-resistance state so require a cool-down or recovery period after power is removed to allow the resistance of the NTC thermistor to increase sufficiently to provide the required inrush current suppression the next time it is needed.

Thus, the speed of response of a current limiting thermistor is given by its time constant. That is, the time taken for its resistance to change by 63% (i.e. 1 to 1/e) of the total change.

Thus, NTC thermistors provide protection from undesirably high inrush currents, while their resistance remains negligibly low during continuous operation supplying power to the load. The advantage here is that they able to effectively handle much higher inrush currents than standard fixed current limiting resistors with the same power consumption.

However, the operating current of the thermistor must be kept as low as possible to reduce any self-heating effects. If they pass operating currents which are too high, they can create more heat than can be quickly dissipated from the thermistor which may cause false results.

Thermistors are characterized by their base resistance and their B value. The base resistance, for example, 10 kΩ, is the resistance of the thermistor at a given temperature, usually 25 ^{o}C and is defined as: R25. The B value is a fixed material constant that describes the shape of the slope of the resistive curve over temperature (*R/T*).

We have also seen that thermistors can be used to measure an external temperature or can be used to control a current as a result of the I^2R heating effect caused by the current flowing through it. By connecting an NTC thermistor in series with a load, it is possible to effectively limit the high inrush currents.

6.1.4 Silicon Diodes

The silicon diode sensor is a device that has been developed specifically for the cryogenic temperature range. Essentially, they are linear devices where the conductivity of the diode increases linearly in the low cryogenic regions.

While the temperature-dependent characteristics of semiconductors are a big disadvantage in electronics this feature is useful in sensor technology i.e. in temperature measurement.

The forward and reverse voltage-current characteristics of a diode are shown in Fig. 6.5.

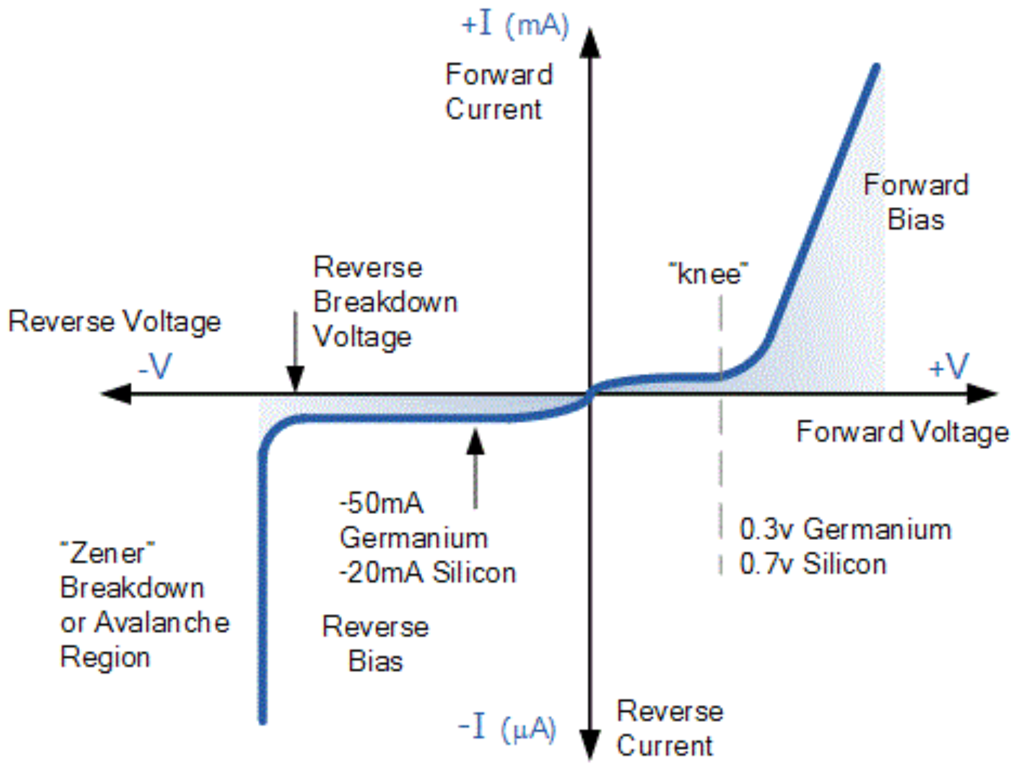


Figure 6.5 Voltage-current characteristic of the diode

The voltage-current characteristics of the PN-junction are temperature dependent and in general, we can state the following. The forward characteristics of a Si diode shifts to the left at the rate of 2.5 mV/°C and the reverse saturation current doubles for every 10 °C rise in temperature in reverse bias range. An example of the temperature dependent diode characteristics in both, forward and reverse bias region, Fig. 6.6 illustrates a Zener diode with a negative temperature coefficient and Fig. 6.7 illustrates an Avalanche diode with a positive temperature coefficient *(Source: ConceptsElectronics.com).*

In both cases, diodes show the same behaviour in the forward bias region. A rise in temperature causes a shift to the left i.e. the opening voltage decreases. In the reverse bias range the saturation current increases by the rising temperature, but Zener or breakdown voltage shows the opposite behaviour in the case of PTC and NTC elements.

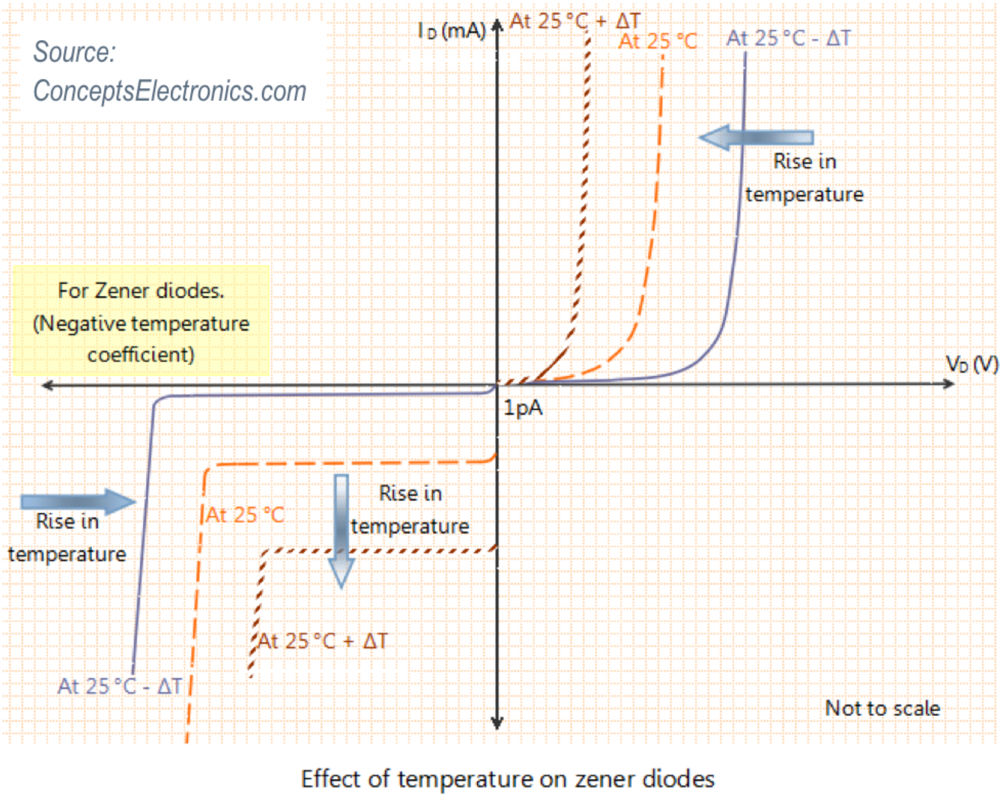


Figure 6.6 Temperature dependency of an NTC Zener diode

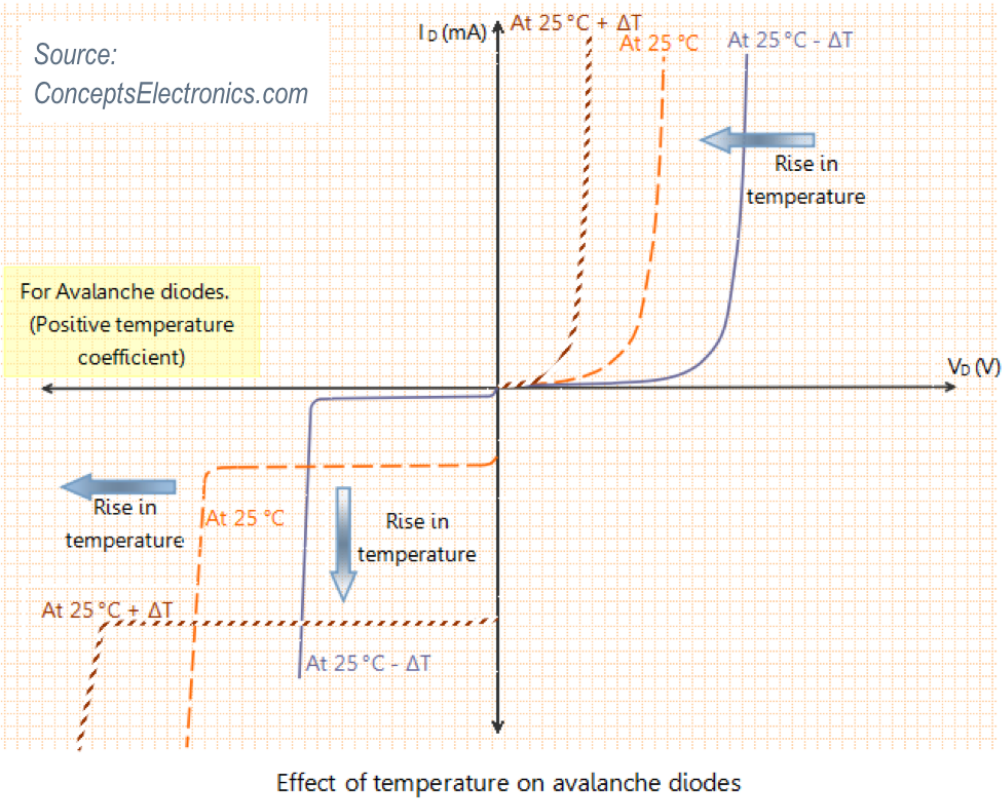


Figure 6.7 Temperature dependency of a PTC Avalanche diode

6.1.5 Pyrometers or Infrared Sensors

The Infrared sensors are non-contact sensors used for temperature measurement. As an example, if you hold up a typical infrared sensor to the front of your desk without contact, the sensor will tell you the temperature of the desk by virtue of its radiation - probably 22 °C at normal room temperature. In a non-contact measurement of ice water, it will measure slightly less than 0 °C because of evaporation, which slightly lowers the expected temperature reading. The applied method of measurement of pyrometers is based on radiation instead of thermal conductivity. This will be discussed later on in the chapter of radiation measurement. However, one example of a simplified operating diagram of a pyrometer is shown in Fig. 6.8. In this method, series connected thermocouples absorbs the energy of infrared radiation and the voltage output is proportional to the intensity of InfraRed (IR) radiation. An optical directing and focusing system should also be applied in such solutions but this part is not shown in Fig. 6.8.

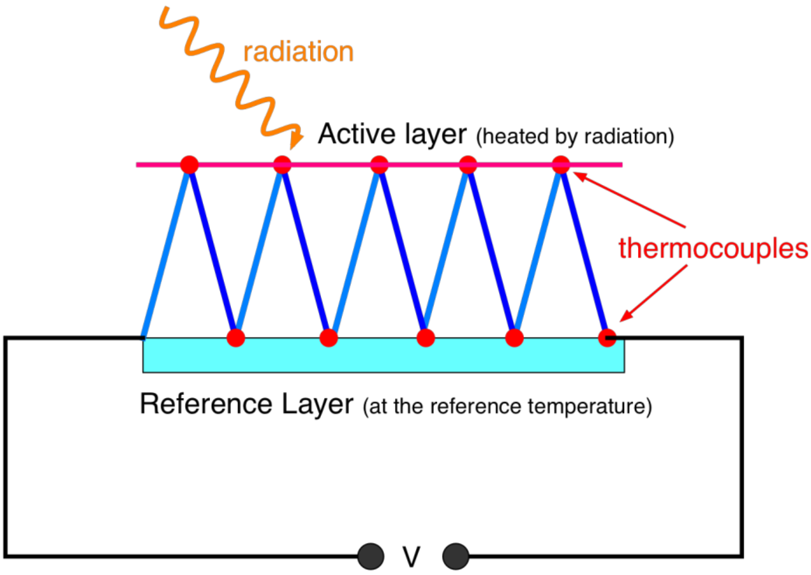


Figure 6.8 Operation principle of a radiation pyrometer

6.1.6 Other Temperature Detectors

Bimetallic Devices

Bimetallic devices take advantage of the expansion of metals when they are heated. In these devices, two metals are bonded together and mechanically linked to a pointer. The principle of operation can be followed on the Fig. 6.9. When heated, one side of the bimetallic strip will expand more than the other. And when geared properly to a pointer, the temperature is indicated.

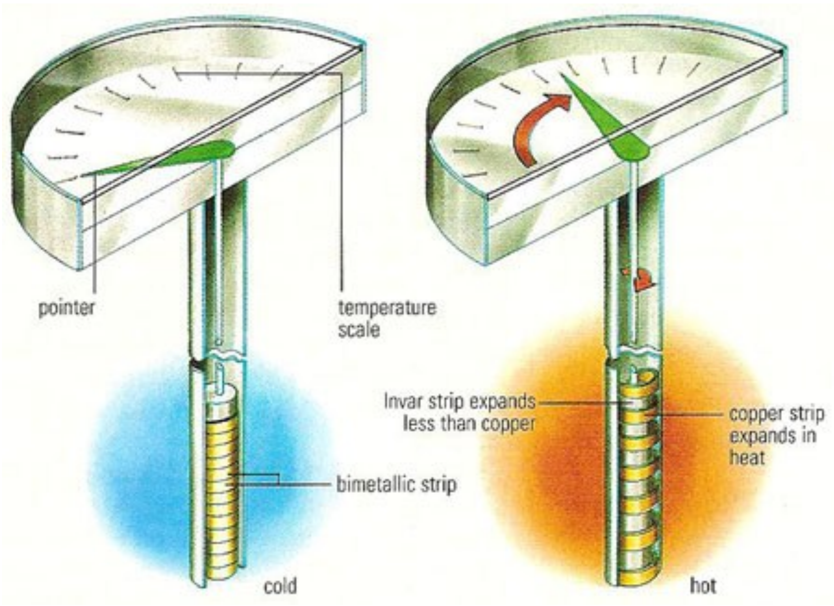


Figure 6.9 Bimetallic thermometer

The advantages of bimetallic devices are portability and independence from a power supply. However, they are usually not as accurate as electrical devices, and you cannot easily record the temperature value as with electrical devices like thermocouples or RTDs; but portability is a definite advantage for the right application. Bimetallic devices are portable but typically less accurate than electrical devices.

Thermometers

Thermometers are well-known liquid expansion devices. Generally speaking, they come in two main classifications: the mercury type and the organic, usually red, liquid type. The distinction between the two is notable, because mercury devices have certain limitations when it comes to how they can be safely transported or shipped.

For example, mercury is considered an environmental contaminant, so breakage can be hazardous. An example of a mercury thermometer is shown in Fig. 6.10.

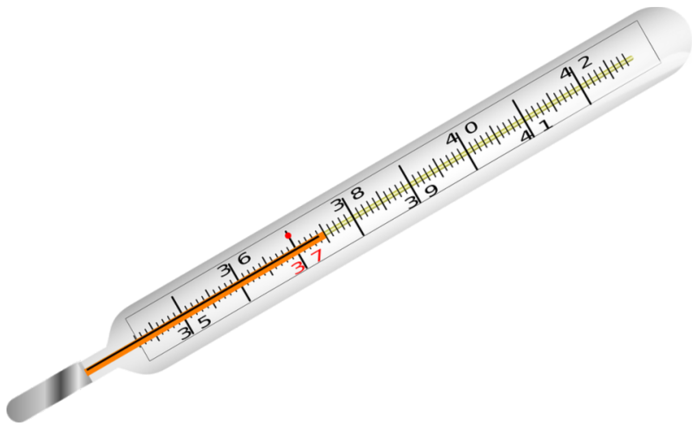


Figure 6.10 Mercury type thermometer

Change-of-State Sensors

Change-of-state temperature sensors measure just that - a change in the state of a material brought about by a change in temperature, as in a change from ice to water and then to steam.

Commercially available devices of this type are in the form of labels, pellets, crayons, or lacquers.

For example, labels may be used on steam traps. When the trap needs adjustment, it becomes hot; then, the white dot on the label will indicate the temperature rise by turning black. The dot remains black, even if the temperature returns to normal.

Limitations include a relatively slow response time. Therefore, if you have a temperature spike going up and then down very quickly, there may be no visible response. Accuracy is not as high as with most of the other devices more commonly used in industry. However, within their realm of application where you need a non-reversible indication that does not require electrical power, they are very practical.

6.2 Radiation Sensors

6.2.1 Measurement of Visible Light

The *visible spectrum* is the portion of the electromagnetic spectrum that is visible to the human eye. Electromagnetic radiation in this range of wavelength is called *visible light*. A typical human eye will respond to wavelengths with a frequency range from about 390 to 700 nm, this corresponds to a band about 430–770 THz. In this section we discuss the main features and measurement methods of the features of visible light.

6.2.1.1 The Electromagnetic Spectrum

The electromagnetic spectrum of frequency (f) versus wave length (λ) is shown in Fig. 6.11.

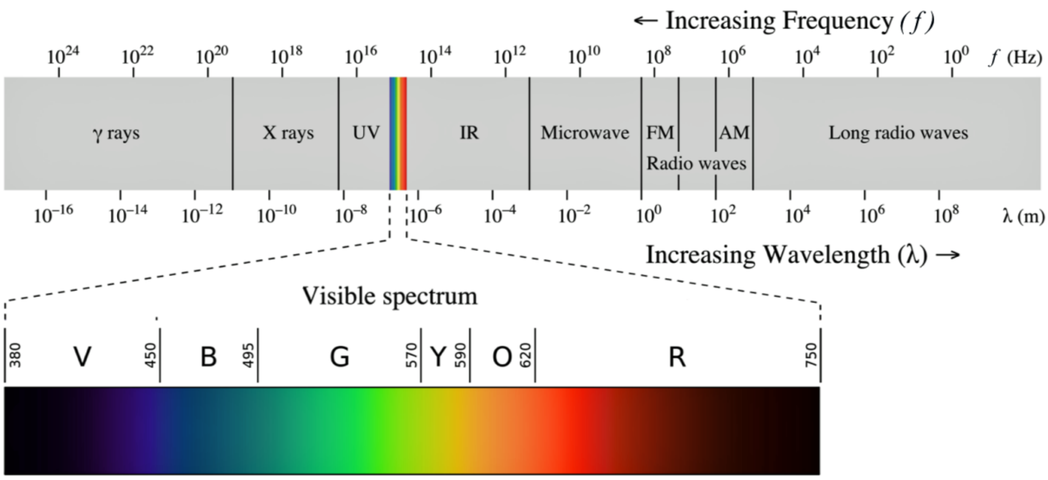


Figure 6.11 Electromagnetic spectrum

Wavelength is defined as the distance travelled by the electromagnetic wave during the period (T). We can write the relationship between the frequency and wave length as given in (6.10) supposing the speed of wave propagation is equal to the speed of light (c).

$$\lambda = c \cdot T = \frac{c}{f} \tag{6.10}$$

6.2.1.2 Human Eye

Because the most important 'sensor' of visible light is the *human eye* we discuss briefly how it reacts to light i.e. how we sense the intensity and colour of the light.

The light receptors in the eye are located around the fovea centralis as shown in Fig. 6.12. Two different type of receptors are *rods* and *cones* that each have a different task in light sensing. While rods are responsible for detecting the intensity of the light, cones perceive and recognize colours.

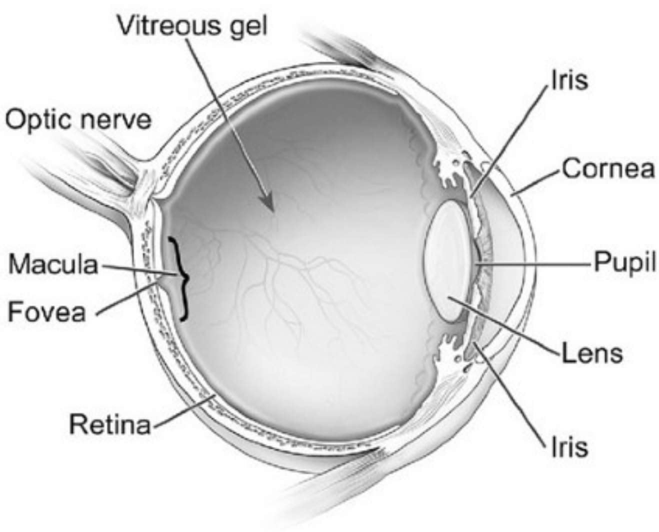


Figure 6.12 The human eye

The normalized spectral sensitivity of the receptors in the human eye is shown in Fig. 6.13. There are three types of cones and a single type of rod. While the spectral sensitivity of rods covers the full range of visible light, primarily the blue cones are sensitive to blue colours, green cones are sensitive to green colours and red cones are sensitive to red colours. Normalized sensitivity curves are shown in Fig. 6.13. Rods have higher sensitivity than cones.

An interesting 'feature' of human eye results from the fact that there are no rods at the fovea centralis, that is, the area responsible for the sharp vision. If you want to recognize a dim star at night, you are better off to look next to it to use the more sensitive rods on the peripheral of the fovea centralis instead of using the sharp vision area filled by cones with lower sensitivity.

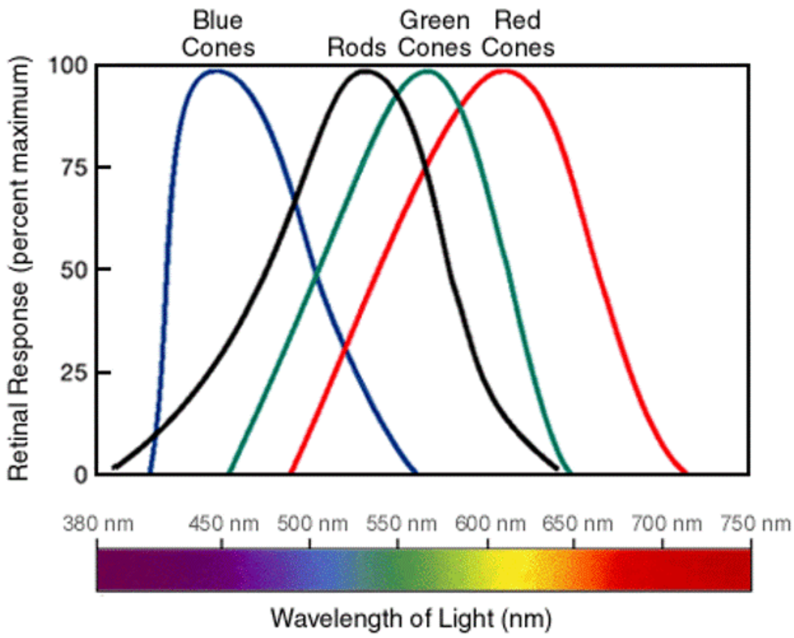


Figure 6.13 Spectral sensitivity of human eye

6.2.1.3 Colour Space

The spectrum, shown in Fig. 6.11, does not contain all the colours that the human eyes and brain can distinguish. Unsaturated colours such as pink, or purple variations such as magenta, are absent, for example, because they can be made only by a mix of multiple wavelengths. While colours, containing only one wavelength are also called pure colours or *spectral colours*, colours with mix of multiple wavelengths are the *mixed colours*. The full range of the visible colours i.e. the spectral colours and mixed colours can be represented in colour space, shown in Fig. 6.14. Spectral colours are located at the outer line ('colour horseshoe') and mixed colours are on the surface encompassed by the 'horseshoe'. The numbers in the outer line are the wavelengths of the spectral colours. The internal line in the colour space scaled with temperature in K (kelvin) is the *colour temperature* line. In photometry this scale is widely used because it refers to the temperature of the source of light in relation to its colour. Sunshine consists of wide spectrum of visible light beside of wide spectrum of *Ultra Violet* (UV) and IR lights. A reasonable part of this wide spectrum is absorbed by the atmosphere. This absorption, even in the visible part of the spectrum, strongly depends on the structure and thickness of the atmosphere. That's why sunshine is somewhat red in early morning and late afternoon and is more blue at noon. The colour of the light from an electric bulb is more yellow (red). This difference in colours is measured using the temperature of source. For example, sunshine has colour temperature of 5500-6000 K and electric bulbs have colour temperature of 2000-3000 K.

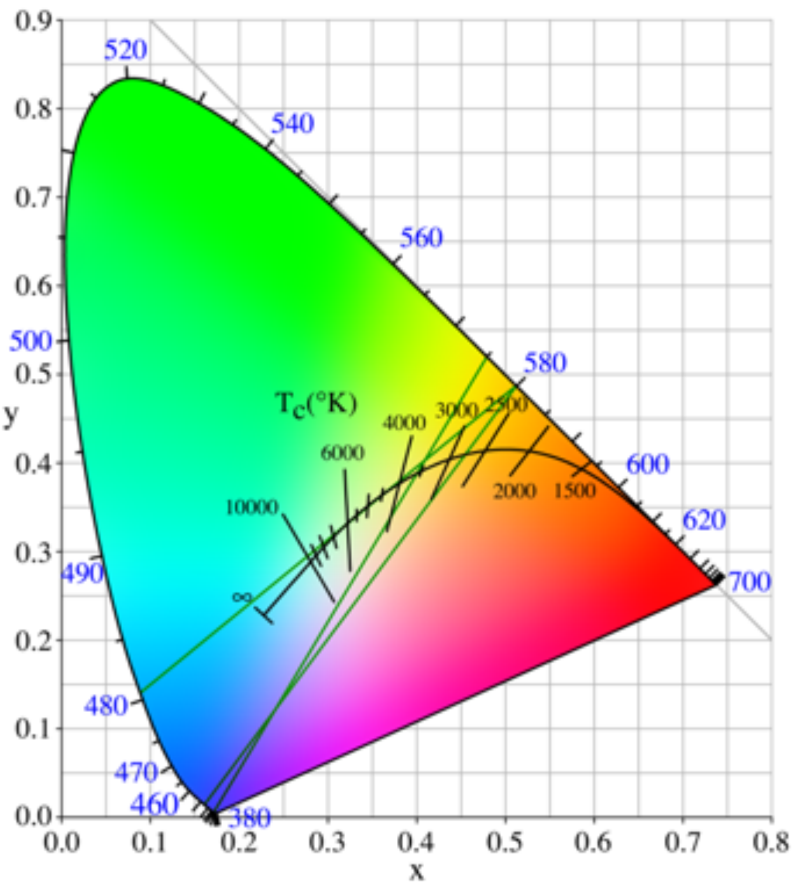


Figure 6.14 The colour space

We see the full range of mixed colours i.e. the full range of visible colours cannot be mixed by using three *base colours* or *colour channels*. Although, the area of the triangle determined by the three colour channels within the colour space can be high enough to produce (and reproduce) almost everything within the total colour space, it is very important to define the same colour channels for mixed colour to replicate the same result in colour production and reproduction.

Standard Colour Spaces

There is practically an infinite possibility to define colour channels to be used in colour mixing. Painters in their artistic works use many colours in their palette to have a high coverage of total colour space as possible. Even using three colour channels we have different possibilities in coverage of the total colour space as shown in Fig. 6.15.

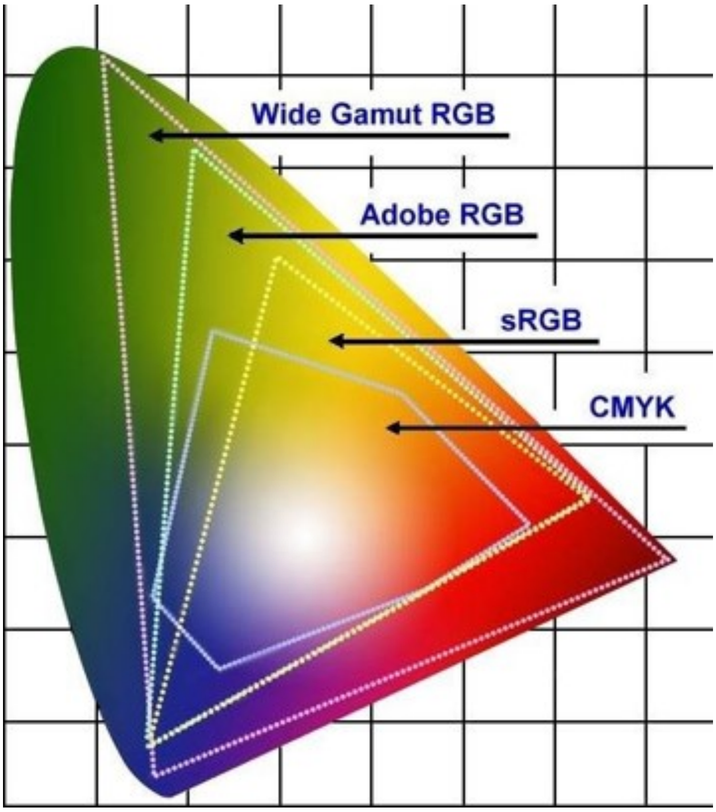


Figure 6.15 Different standard colour spaces

To avoid (or minimize) colour distortion, the production and reproduction of colours has to be interpreted in the same area so, the colour spaces applied in practice have to be standardized.

The mostly used colour spaces are shown in Fig. 6.15. The *sRGB* or *'standard RGB'* colour space uses red (R), green (G), and blue (B) colour channels as given in Fig. 6.15. This rather poor colour coverage is standardized for the Internet which means that it is the most commonly used. *Adobe RGB*, and *Wide Gamut RGB* are defined for professional use according to their rich colour coverage. There is an additional colour space in Fig. 6.15, that is CMYK using cyan (C), magenta (M), yellow (Y) and additional black (K). While RGB colour spaces are used for electronic colour mixing i.e. for displays and screens, CMYK colour space is used in publishing and printing.

Measurement devices and other equipment for the measurement and processing of visible light need to apply colour management that is illustrated in Fig. 6.16. The colour management assures the same colour space in each element or device of the system by application of the same set of data that characterizes a colour input or output of devices. The applied colour space is defined according to standards publicised by the International Colour Consortium (ICC). The ICC profiles are embedded in each device describing the colour attributes of a particular device or viewing requirement by defining a mapping between the device source or target colour space.

For example, if the source of the information, measured data, scanned image, etc. uses the Adobe RGB type of ICC profile, the same type of profile has to be used at the working place for post processing and also the display and printing device.

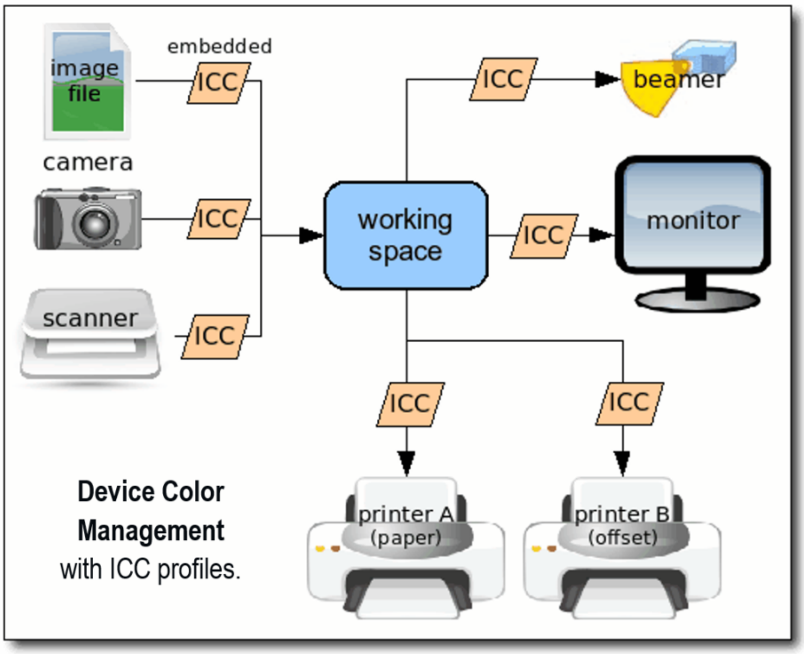


Figure 6.16 Colour space management

6.2.1.4 Colour Mixing

We mentioned that the usage of *Red-Green-Blue* (RGB) colour channels is not the only possibility, for example, Cyan-Magenta-Yellow-blacK (CMYK) channels are used in printing. To understands the reason for the different colour channel applications we need to examine two different ways of colour mixing that are additive and subtractive colour mixing. If we had three projectors with red, green and blue light sources and project their beam to white board like it is on the left side of Fig. 6.17 then we have mixed colours based on additive colour mixing. If we paint a white paper with cyan, magenta and yellow paints like it is on the right side of Fig. 6.17 then we have mixed colour based on subtractive colour mixing. Additive colour mixing, based on RGB colour channels is used in displays and screens and subtractive colour mixing based on complemented colours of RGB, that is, CMY colours, is used in printing. Please note that there is an additional 'colour' channel in subtractive colour mixing, that is black (K). It is applied not because of theoretical necessity but for better quality and for better efficiency in printing. Printers are used for solely black and white text printing in many cases so it is better to use a single black 'colour' instead of using a set of colour paints.

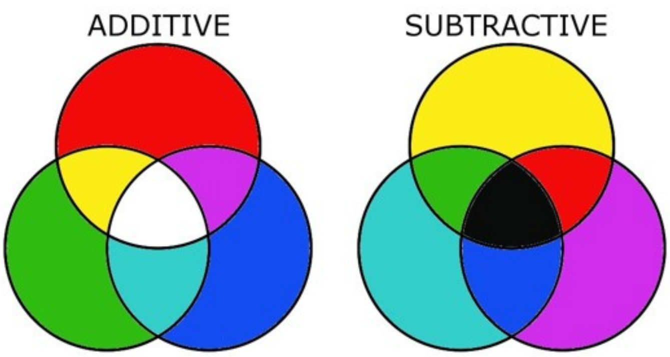


Figure 6.17 Additive and subtractive colour mixing

A painter's palette is shown in Fig 6.18 for illustration. Please note the number of applied base colours (colour channels) for their subtractive colour mixing.

Figure 6.18 Artistic colour palette

6.2.1.5 Light Sensors

Light sensors are transducers to convert the change of intensity of a visible light radiation to a change in an electric parameter like voltage, current or resistance. In measurements we usually want to know either the intensity of certain spectral components or the mean value over the whole visible spectrum. It is important to know that single wave light radiation with a single spectral colour is very rare in real life. This narrow band light, being close to spectral light, is called *monochrome light*. Because practical light sensors are broadband sensors having reasonable sensitivity over a full spectrum of visible light, colour filters must be applied to the sensor to be able to measure the monochrome component of the measured light. A possible yellow filter is shown in Fig. 6.19 to measure the yellow component of light.

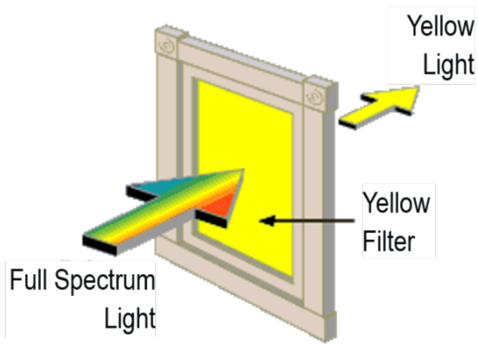


Figure 6.19 Colour filter applied to a light sensor

According to the principle of operation we can distinguish between photo-conductive and photo-electric sensors. While photo-conductive sensors are based on light radiation dependent resistors, photo electric sensors produce an electric voltage (or current) in their output.

Photo-Conductive Sensors (Photo-Resistors)

Photoconductivity in general is an optical and electrical phenomenon in which a material becomes more electrically conductive due to the absorption of electromagnetic radiation such as visible light, ultraviolet light, infrared light, or even X-ray or gamma radiation.

When light is absorbed by a material such as a semiconductor, the number of free electrons and electron holes increases and raises its electrical conductivity. To cause excitation, the light that strikes the semiconductor must have enough energy to raise electrons across the band gap, or to excite the impurities within the band gap. When a bias voltage and a load resistor are used in series with the semiconductor, a voltage drop across the load resistors can be measured when the change in electrical conductivity of the material varies the current through the circuit. Classic examples of photoconductive materials include Cadmium Sulphide (CdS), used in photography, Lead Sulphide (PbS), used in infrared detection. The symbol of photo-conductive sensors and a cadmium sulphide detector is given in Fig. 6.20.

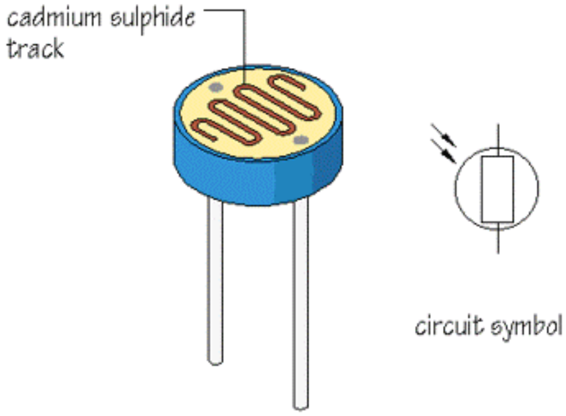


Figure 6.20 CdS detector and its symbol

A photoresistor (or Light-Dependent Resistor, LDR, or photo-conductive cell) is a light-controlled variable resistor. The resistance of a photoresistor decreases with increasing incident light intensity; in other words, it exhibits photoconductivity. A photoresistor can be applied in light-sensitive detector circuits, and light-activated and dark-activated switching circuits.

A photoresistor is made of a high resistance semiconductor. In the dark, a photoresistor can have a resistance as high as several megaohms (MΩ), while in the light, a photoresistor can have a resistance as low as a few hundred ohms. If incident light on a photoresistor exceeds a certain frequency, photons absorbed by the semiconductor give bound electrons enough energy to jump into the conduction band. The resulting free electrons (and their hole partners) conduct electricity, thereby lowering resistance. The resistance range and sensitivity of a photoresistor can differ substantially among different devices. Moreover, unique photoresistors may react substantially differently to photons within certain wavelength bands.

The principle of operation of the photoresistors is demonstrated by a simple example of a light-controlled switch, shown in Fig. 6.21. When the bulb lights up the resistance of the sensor falls and LED lights up. Please note that the controlled circuit in Fig. 6.21 has the features of both the switching (controlling) circuit and switched (controlled) circuit which are galvanically insulated i.e. there is no electric connection between two sub-circuits.

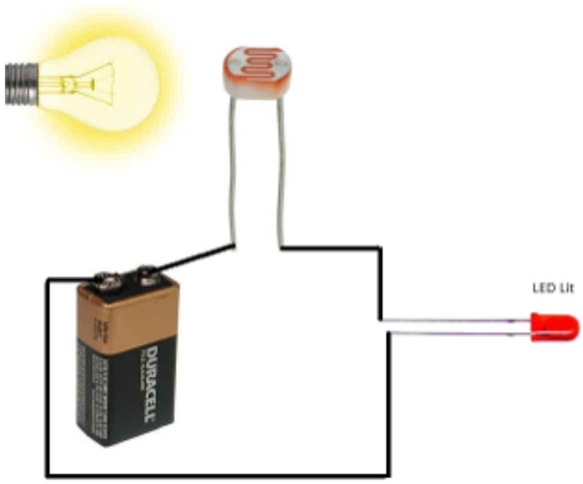


Figure 6.21 Light controlled switch

Although there are numerous benefits of photo resistors, like simplicity, reliability, and low cost, they also have drawbacks, primarily the slow operation due to significant parasitic capacitance.

Photo-Electric Sensors

Photo-electric sensors are active elements as they produce an electric voltage proportional to the light they detect. In principle there is no need for an additional source of electricity for their operation. Classic examples of photo-electric detectors include selenium-cell, phototube (photocell), photo diode and photo transistor.

Selenium Cell

Selenium (Se) semiconductor materials were applied in photometry in early television and photography. An illuminated selenium cell produces a voltage that can be measured as shown in Fig. 6.22.

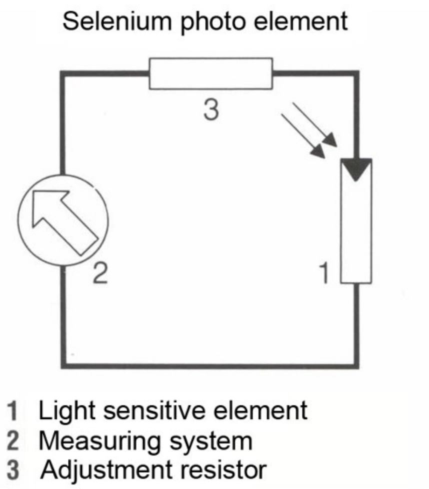


Figure 6.22 Selenium cell photoelement

The main benefits of the selenium photo element are that it is an active sensor, with no need of additional energy and its spectral sensitivity is very similar to the spectral sensitivity of the human eye. The main disadvantage in measurement applications is aging, i.e. there is a change in its sensitivity over time. Because of the need for frequent calibration this element is mainly significant from a historical point of view and has been replaced by alternative elements in measuring instruments.

Phototube (Photocell)

A phototube or photoelectric cell is a type of gas-filled or vacuum tube that is sensitive to light. Such a tube is more correctly called a 'photo-emissive cell' to distinguish it from photovoltaic or photoconductive cells. Phototubes were previously more widely used but have been replaced in many applications by solid state photodetectors. However, the photomultiplier tube is one of the most sensitive light detectors and is still widely used in physics research. The principle of operation can be followed in Fig. 6.23.

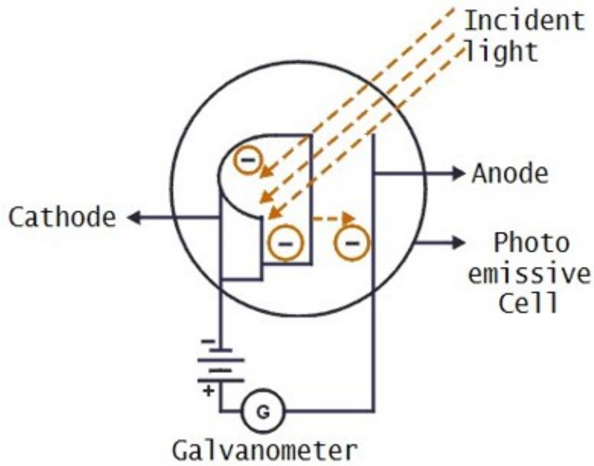


Figure 6.23 Phototube

The cathode of the tube is made of metal, like Cesium (Cz), with low emission energy. When the cathode is illuminated, electron emission occurs and due to the positive anode potential, the emitted electrons will flow from cathode to anode. This electron flow is proportional to the illumination assuming a constant anode voltage. The phototube is a sensitive photo detector but has certain drawbacks in practical applications since it is fragile and requires more than a hundred volts for anode voltage.

Photodiode and Phototransistor

A photodiode is a semiconductor device that converts light into an electrical current. The current is generated when photons are absorbed in the photodiode. Photodiodes may contain optical filters, built-in lenses, and may have large or small surface areas. Photodiodes usually have a slower response time as their surface area increases. The common, traditional solar cells used to generate electricity from the sun are large area photodiodes.

Photodiodes are like regular semiconductor diodes except that they may be either exposed (to detect vacuum UV or X-rays) or packaged with a window or optical fibre connection to allow light to reach the sensitive part of the device. Many diodes designed for use specifically as a photodiode use a PIN junction rather than a p–n junction, to increase the speed of response. The phototransistor is a semiconductor light sensor formed from a basic transistor with a transparent cover that provides much better sensitivity than a photodiode. The electrical symbol, the structure for illustration of light control and the V_{CE}-I_C characteristics versus illumination is shown in Fig 6.24.

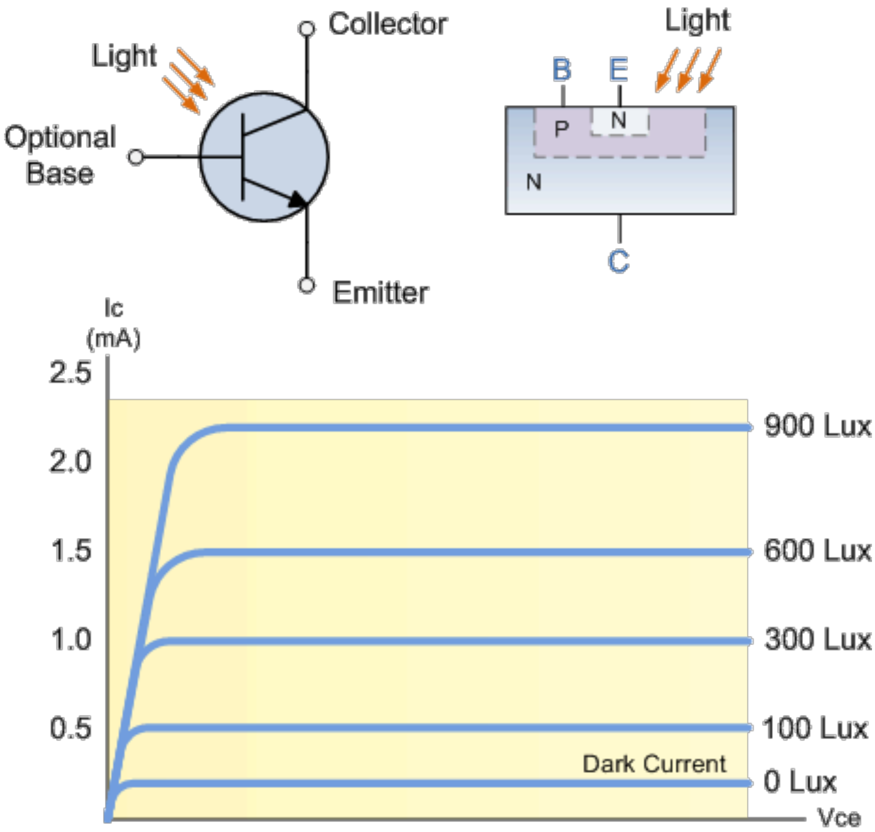


Figure 6.24 Phototransistor with its characteristics

A wide variety of phototransistor options are available for practical applications, such as ray sensors, consumption sensors, magnetic sensors, etc. For illustration, two phototransistors for radiation detection are shown in Fig. 6.25.

Figure 6.25 Phototransistor products

6.2.1.6 Examples of Application

Opto-Isolators

In electronics, an *opto-isolator*, also called an *optocoupler*, *photocoupler*, or *optical isolator*, is a component that transfers electrical signals between two isolated circuits using light. Opto-isolators prevent high voltages from affecting the system receiving the signal. Commercially available opto-isolators withstand input-to-output voltages up to 10 kV and voltage transients with speeds up to 25 kV/µs.

A common type of opto-isolator consists of an LED and a phototransistor in the same opaque package. Usually opto-isolators transfer digital (on-off) signals, but some techniques allow them to be used with analogue signals. A typical application circuit for an opto-isolator is shown in Fig. 6.26.

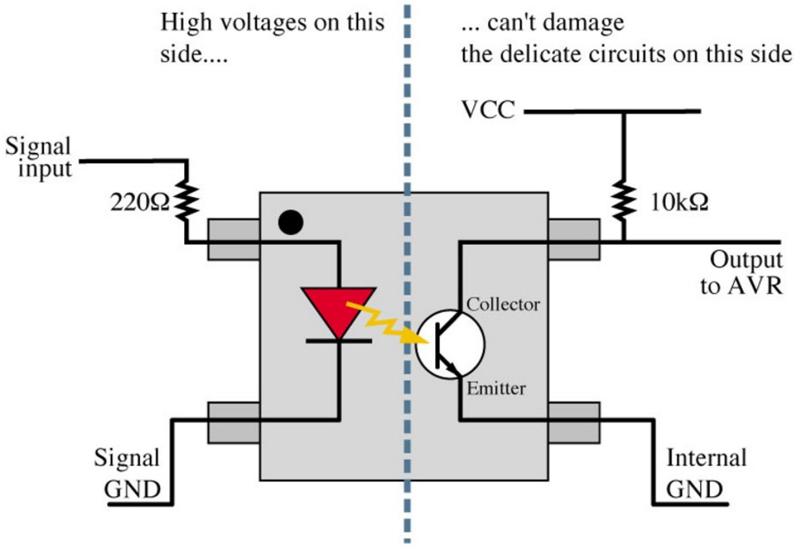


Figure 6.26 Circuit with opto-isolator

Colour Temperature Measurement

The *colour temperature* of a light source has already been mentioned in section 6.2.1.3 as the characteristic parameter of its colour. The colour temperature curve was also given in colour space in Fig. 6.14. An exact definition of colour temperature is the temperature of an ideal black-body radiator that radiates light of a colour comparable to that of the light source. Colour temperature is a characteristic of visible light that has important applications in lighting, photography, astrophysics, and other fields. In practice, colour temperature is meaningful only for light sources that do in fact correspond somewhat closely to the radiation of some black body, i.e., those on a line from reddish/orange via yellow and more or less white to blueish white; it does not make sense to speak of the colour temperature of, e.g., a green or a purple light. Colour temperature is conventionally expressed in kelvin, using the symbol K, a unit of measure for absolute temperature.

Colour temperatures over 5000 K are called 'cool colours' (bluish white), while lower colour temperatures (2700–3000 K) are called 'warm colours' (yellowish white through red). 'Warm' in this context is an analogy to radiated heat flux of traditional incandescent lighting rather than temperature. The spectral peak of warm-coloured light is closer to infrared, and most natural warm-coloured light sources emit significant infrared radiation. The fact that 'warm' lighting in this sense has a 'cooler' colour temperature often leads to confusion. The simple method of colour temperature measurement with a colour temperature meter is given in Fig. 6.27.

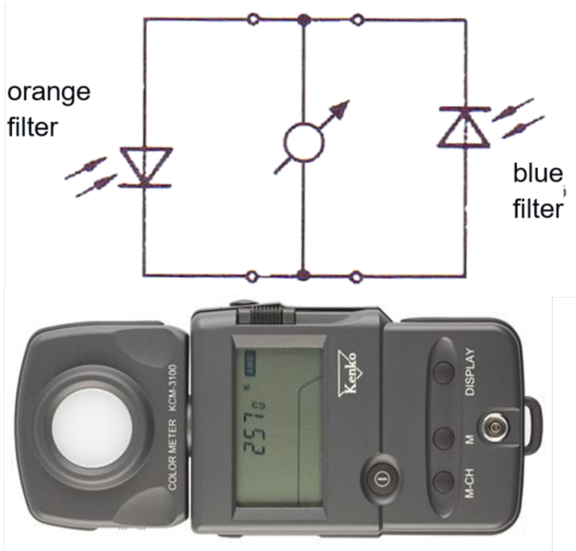


Figure 6.27 Device for colour temperature measurement

The basic measurement method is simple as two of photodetectors are connected anti-parallel and are equipped with blue and orange (red) colour filters. Measuring the mixed colour, depending of the ratio of blue-orange components, the indicator will display the colour temperature.

Digital Imaging Devices (CCD, CMOS)

Digital imaging or *digital image acquisition* is the creation of photographic images, such as of a physical scene or of the interior structure of an object. The term is often assumed to imply or include the processing, compression, storage, printing, and display of such images.

Digital imaging can be classified by the type of electromagnetic radiation or other waves whose variable attenuation, as they pass through or reflect off objects, conveys the information that constitutes the image. In all classes of digital imaging, the information is converted by *image sensors* into digital signals that are processed by a computer and create an output as a visible-light image. For example, the medium of visible light allows digital photography (including digital videography) with various kinds of digital cameras (including digital video cameras). X-rays allow digital X-ray imaging (digital radiography, fluoroscopy, and Computer Tomography, CT), and gamma rays allow digital gamma ray imaging (digital scintigraphy, etc.). Sound allows ultrasonography (such as medical ultrasonography) and sonar, and radio waves allow radar. Digital imaging lends itself well to image analysis by software, as well as to image editing (including image manipulation).

An image sensor is a sensor that detects and conveys the information that constitutes an image. It does so by converting the variable attenuation of light waves (as they pass through or reflect off objects) into signals that convey information.

Early analogue sensors for visible light were video camera tubes however, nowadays we use semiconductor Charge-Coupled Devices (CCD), active pixel sensors in Complementary Metal–Oxide–Semiconductor (CMOS), or N-type Metal-Oxide-Semiconductor (NMOS, Live MOS) technologies.

In both, CCD and CMOS technology the applied elementary light sensors at pixel points are broadband sensors having reasonable sensitivity in a full spectrum of visible light.

There are several main types of colour image sensors, differing by the type of colour-separation mechanism. One of the most commonly used methods is the Bayer pattern method, shown in Fig. 6.28.

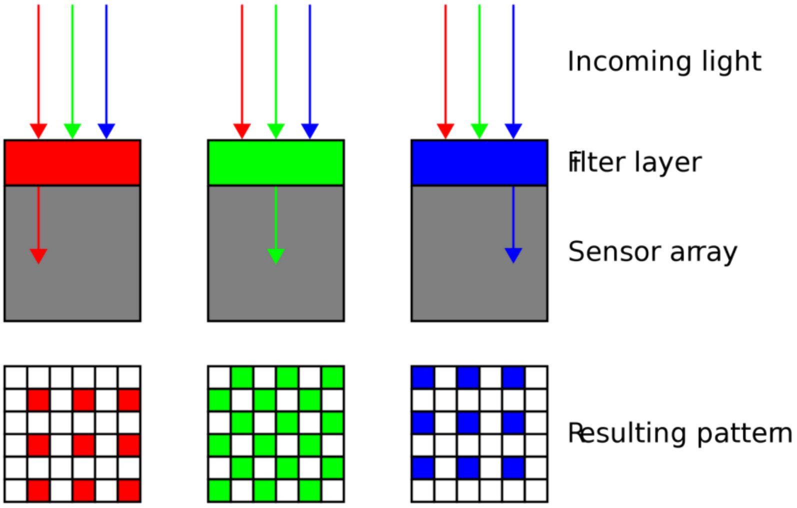


Figure 6.28 Bayer pattern sensor

The *Bayer filter sensor* applies a colour filter array that passes red, green, and blue light to selected pixel sensors. Each individual sensor element is made sensitive to red, green, and blue by means of a colour gel made of chemical dye placed over each individual element. Although inexpensive to manufacture, this technique lacks the colour purity of dichroic filters. Because the colour gel segment must be separated from the others by a frame (like stained glass windows), less of the areal density of a Bayer filter sensor is available to capture light, making the Bayer filter sensor less sensitive than other colour sensors of similar size. This loss can be negated by using a larger sensor size, albeit at greater cost. The most common Bayer filter matrix uses two green pixels, and one each for red and blue. This results in less resolution for red and blue colours, corresponding to the human eye's reduced sensitivity at the limits of the visual spectrum. The general operational diagram for both, CCD and CMOS equipped imaging sensors are shown in Fig. 6.29. CCD and CMOS arrays have theoretically almost the same matrix structure. It is a bit confusing that we call CCD sensor the simple CCD pixel array with analogue output only, while CMOS sensor is the *smart sensor* containing not only a CMOS pixel array but also the analogue digital conversion and digital control unit for the pixel array.

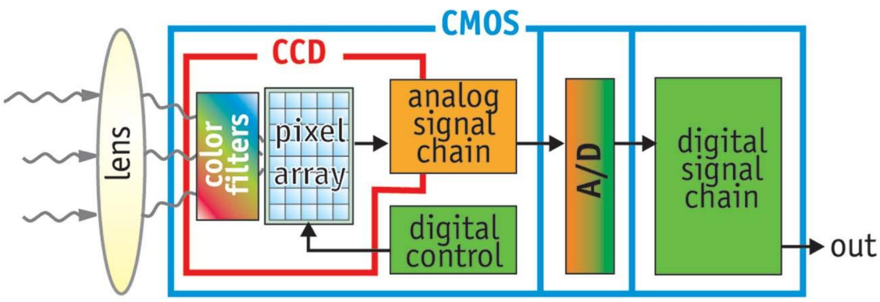


Figure 6.29 Digital imaging smart sensor

Image File Formats

The output *data* of an image sensor is a set of digital information that needs to be post-processed and stored as an image file. The file format of the sensor output is so called RAW, that is the bitmap of the pixel array, supplemented with metadata information about sensor parameters. Because the RAW file format is sensor dependent it must be post-processed to produce standard image format. There are number of image file formats but, due to the internet and web the widely used image file format is JPEG that is the acronym for Joint Photographic Experts Group. Many bitmap images are saved in .jpg format to compress the files. This makes it easier to transfer or download the files and use them on the Web. The JPG file format is based on a 24-bit colour palette; however, these files tend to drop information when they are decompressed. The higher the rate of compression was when the JPG file was created, the more the quality of the image will be affected when the file is decompressed.

The 24-bit colour palette means 16,777,216 different colours. The human eye can separate up to 10 million different colours, so the 24-bit colour range in a JPG file is more than enough.

These 24 bits means 8-bit luminance depth in each colour channel (RGB), so the luminance resolution is 256 different luminance levels. For a human eye's resolution it is also sufficient, but not in every case, as will be shown in the followings.

If we need to post-process the image file for any reason i.e. the exposition value must be corrected, the 8-bit dynamic range in luminance of the JPG file is insufficient. In addition, there is lossy compression in JPG files, which is another reason for the poor quality of JPG post-processing.

In practice the AD converters, applied in image sensors have 12 or 14-bit resolutions.

If we use a RAW file for external post-processing, even the 12-bit dynamic range means 4095 luminance levels, that is 16 times higher than the JPG dynamic range which should be adequate for exposure value correction in post-processing. A significant correction in the exposure value as image post-processing in a RAW file is demonstrated in Fig. 6.30. It is impossible to notice any degradation in the image quality while the details are recovered due to the existing dynamic range for correction.

Figure 6.30 Exposure value correction in RAW format

6.2.2 Infrared Measurements

Infrared radiation can be measured by an increase in temperature detected, which means that an infrared sensor is a thermal sensor complete with an absorber of infrared energy. For design of an appropriate absorber we must consider the following.

According to the rule of energy conservation, surface energy is divided into three parts as seen in Fig. 6.31.

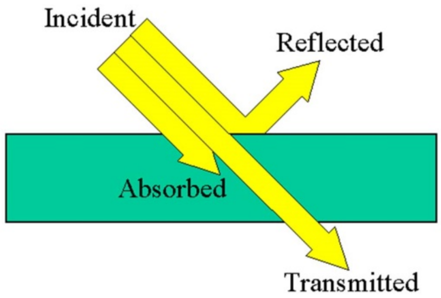


Figure 6.31 Three parts of surface energy

We can describe this rule as in (6.11) where *E* is the total (incident) energy and E_r, E_a and E_t are its reflected, absorbed and transmitted parts in order.

$$E = E_r + E_a + E_t \tag{6.11}$$

We can write (6.11) in the following form also.

$$1 = \frac{E_r}{E} + \frac{E_a}{E} + \frac{E_t}{E} \tag{6.12}$$

Introducing the coefficients of reflection, absorption and transmission we have the following equation.

$$1 = r + a + t \tag{6.13}$$

For an ideal absorber these material parameters are t = 0, r = 0 and a = 1. Soot is an example of such a material. Other non-reflective and 'infrared black' paints are also used on metal layers because the opacity of a metal layer is very low at the infrared range. The absorbed energy increases the temperature of the absorber as described in (6.14).

$$aE = C_v \Delta\vartheta \rightarrow \Delta\vartheta = \frac{a}{C_V}\Delta E \tag{6.14}$$

Thus, the measuring sensitivity can be expressed as in (6.15), where '*a*' is the absorption coefficient and C_V is the heat capacity of the material, applied.

$$S = \frac{\Delta\vartheta}{\Delta E} = \frac{a}{C_V} \tag{6.15}$$

Our goal is to have sensitivity as high as possible so, we need materials with high absorption and low heat capacity.

In practical applications we use so called pyrometers and bolometers described below.

Pyrometer

This infrared sensor consists of a thermocouple usually with an attached optical system. Substituting (6.14) into (6.1) we have (6.17). This equation shows the principal dependency of the output voltage on the incident infrared energy of this IR sensor.

$$V(\vartheta) = k_{AB} \cdot \Delta\vartheta = \frac{\alpha \cdot k_{AB}}{C_V} \Delta E \qquad (6.17)$$

Fig. 6.32 demonstrates a practical contactless thermometer, equipped with pyrometer detector.

Figure 6.32 Practical thermometer with pyrometer detector

Bolometer

While pyrometers usually use thermocouples, bolometers are equipped with a metal layer of resistors. By using metals, we have the benefit of no transmission and also the constant absorption in the wide spectrum from UV to an infrared region. Knowing the temperature dependency of metals, we can write (6.16) that there is a characteristic relationship between the infrared energy and resistance. A change in resistance can be transformed to a change in voltage (according to Ohm's law).

$$R(\vartheta) = R_0(1 + \alpha\Delta\vartheta) = R_0\left(1 + \frac{\alpha \cdot a}{C_V}\Delta E\right) \qquad (6.16)$$

A simple practical bolometer chip is shown in Fig. 6.33. The main benefit of a bolometer over a pyrometer is its simplicity and lower price.

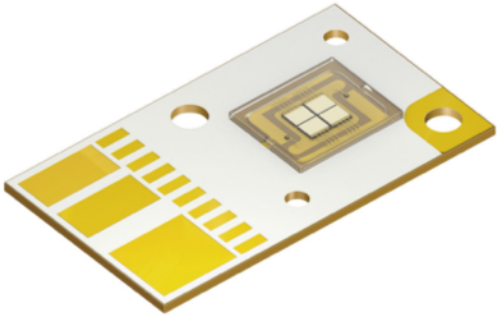


Figure 6.33 Bolometer chip

Applications

Infrared sensors are used in several fields of industrial and personal life i.e. in infrared imaging, medical diagnostics, building diagnostics, motion detection, etc.

For example, Fig. 6.34 shows a motion detection application in home security.

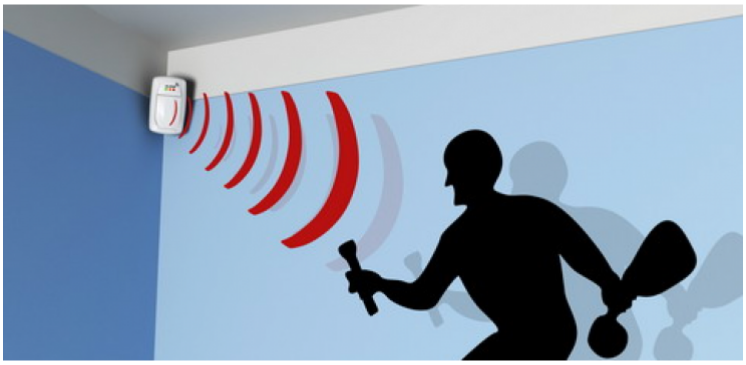


Figure 6.34 IR sensor as motion detector

Fig. 6.35 demonstrates a thermal mapping application example, using thermographic camera.

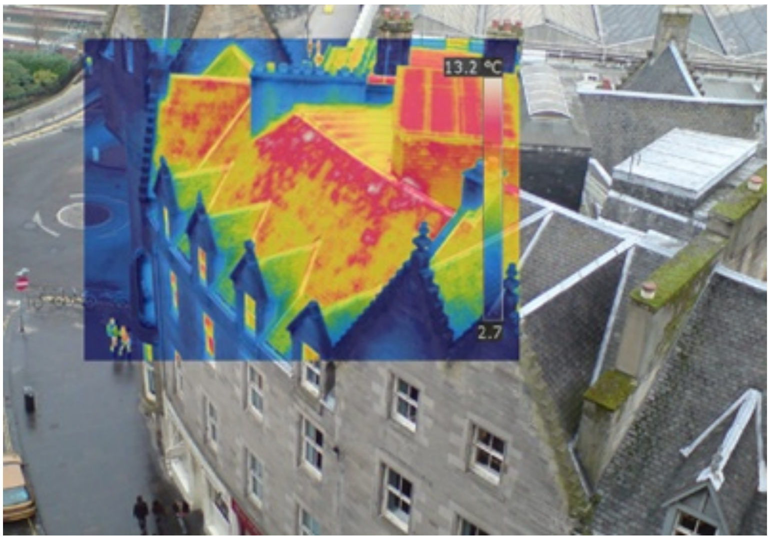


Figure 6.35 Thermal mapping

A practical thermographic camera is shown in Fig. 6.35.

Figure 6.36 Thermographic camera

6.2.3 Nuclear Measurements

Radioactive or ionizing radiation takes a few forms: alpha, beta, and neutron particles, and gamma and X-rays. All types are caused by unstable atoms, which have either an excess of energy or mass (or both). To reach a stable state, they must release that extra energy or mass in the form of radiation.

6.2.3.1 Alpha Radiation

Alpha radiation occurs when an atom undergoes radioactive decay, giving off a particle (called an alpha particle) consisting of two protons and two neutrons (essentially the nucleus of a helium-4 atom), changing the original atom to one an element with an atomic number 2 less and atomic weight 4 less than it started with. Typical alpha decay is shown in Fig. 6.37.

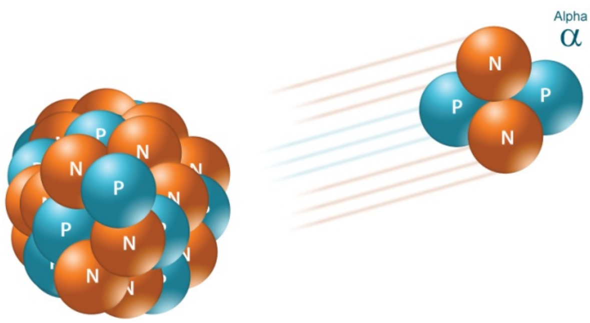
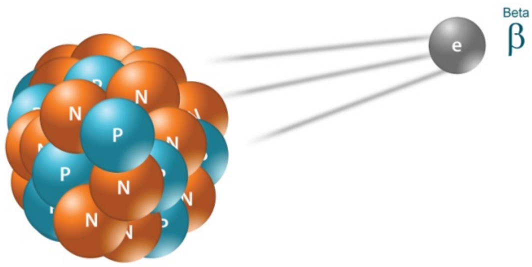


Figure 6.37 The emission of an alpha particle from the nucleus of an atom

Due to their charge and mass, alpha particles interact strongly with matter, and only travel a few centimetres in air. Alpha particles are unable to penetrate the outer layer of dead skin cells, but are capable, if an alpha emitting substance is ingested in food or inhaled, of causing serious cell damage. Alexander Litvinenko, who was poisoned with tea laced with polonium 210, is a famous example.

6.2.3.2 Beta Radiation

Beta radiation takes the form of either an electron or a positron (a particle with the size and mass of an electron, but with a positive charge) being emitted from an atom. Beta radiation is shown in Fig. 6.38.



6.38 The emission of a beta particle from the nucleus of an atom

Due to the smaller mass, it can travel further in the air, up to a few meters, but can be stopped by a thick piece of plastic, or even a stack of paper. It can penetrate skin a few centimetres, posing somewhat of an external health risk. However, the main threat is still primarily from internal emission from ingested material.

6.2.3.3 Gamma Radiation

Gamma radiation, unlike alpha or beta, does not consist of any particles, instead consisting of a photon of energy being emitted from an unstable nucleus as shown in Fig. 6.39.

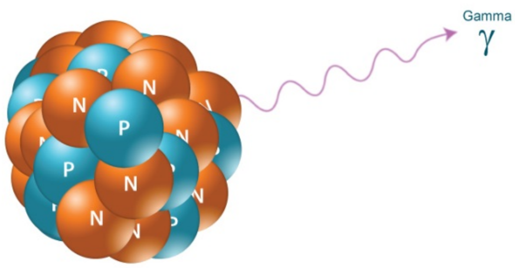


Figure 6.39 The emission of a high-energy wave from the nucleus of an atom

Having no mass or charge, gamma radiation can travel much farther through the air than alpha or beta, losing (on average) half its energy for every 500 feet. Gamma waves can be stopped by a thick or dense enough layer of material with a high atomic number, such as lead or depleted uranium, being the most effective form of shielding.

6.2.3.4 X-Rays

X-rays are like gamma radiation, with the primary difference being that they originate from the electron cloud. This is generally caused by energy changes in an electron, such as moving from a higher energy level to a lower one, causing the excess energy to be released. X-Rays are longer-wavelength and (usually) lower energy than gamma radiation, as well. An X-ray radiation is shown in Fig. 6.40.

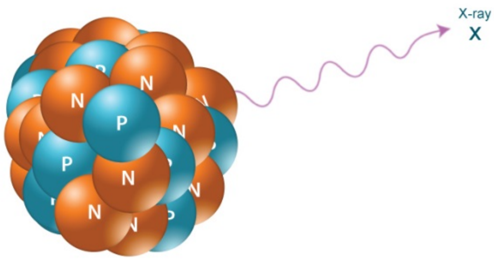


Figure 6.40 The emission of a high energy wave from the electron cloud of an atom

6.2.3.5 Neutron Radiation

Lastly, Neutron radiation consists of a free neutron, usually emitted because of spontaneous or induced nuclear fission. It is shown in Fig. 6.41.

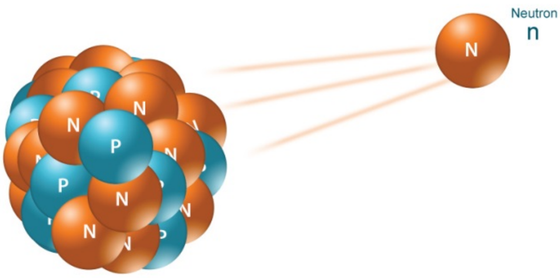


Figure 6.41 The emission of a neutron from the nucleus of an atom

Although they can travel hundreds or even thousands of meters in the air, they can be effectively stopped if blocked by a hydrogen-rich material, such as concrete or water. Typically, they are unable to ionize an atom directly due to their lack of a charge, so neutrons most commonly are indirectly ionizing, in that they are absorbed into a stable atom, thereby making it unstable and more likely to emit ionizing radiation of another type. Neutrons are, in fact, the only type of radiation that can turn other materials radioactive.

6.2.3.6 Penetration of Radioactive Radiation

The penetrating power of different types of radiation strongly depends on their type and ranges from alpha radiation, which has weak penetration, to neutrons which have the strongest penetration, as illustrated in Fig. 6.42. This ability of radiation to penetrate different materials brought about x-ray technology, which discovered in 1901, that has revolutionized the world of modern medicine.

In fact, German physicist Wilhelm Conrad Röntgen was even awarded a Nobel prize for his discovery of X-rays radiation. X-rays effortlessly pass through skin, bone and metal to produce images that the human eye would never be able to see.

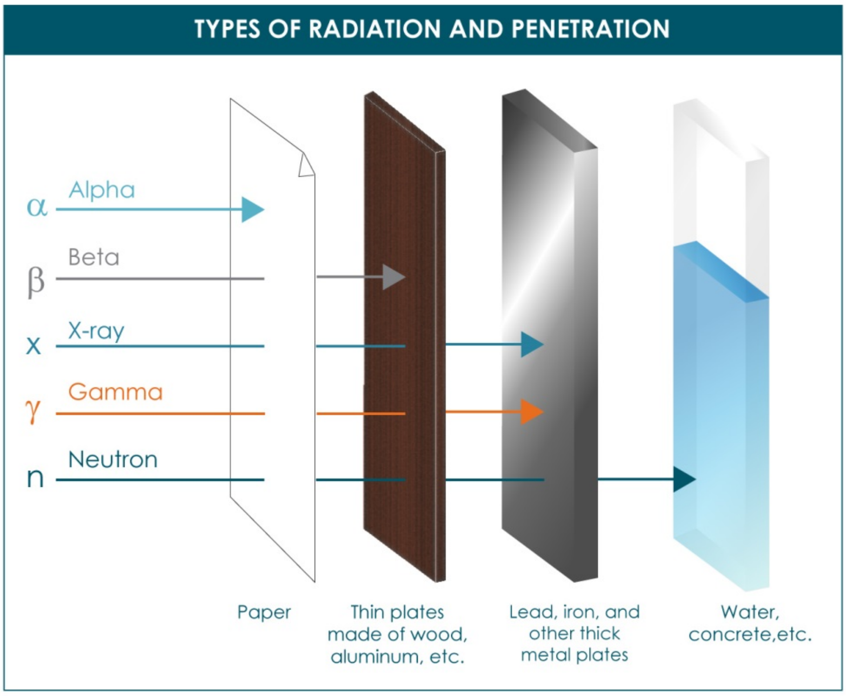


Figure 6.42 Penetration of different types of nuclear radiation

6.2.3.7 Quantum Detectors

There are three main types of radioactive sensors applied in nuclear measurement technology. In contrast to *ionization detectors* and - so called - indirect or *scintillation detectors* which have been used for a long time, *direct radiation detectors* are newly developed, semiconductor-based devices.

Geiger Counter

Geiger counters can detect gamma, alpha and beta radiation. A classic Geiger counter has a Geiger - Mueller tube, a visual readout, and an audio readout. The Geiger - Mueller tube is a metal tube filled with gas that detects radiation. This metal tube has a central wire electrode that is connected to a power supply. At one end of the tube, there is a thin window which is made from a material that can be penetrated by alpha, beta, and gamma rays. When ionizing radiation penetrates the thin window at that end of the metal tube, the gas inside the tube becomes ionized creating electrons and ions. This ionized gas becomes an electrical conductor because of the ions and the free electrons produced. The flow of electrons produces an electrical current flowing between the tube and the inside wire. Current flows every time the gas-filled metal tube is exposed to radiation. The bursts of current are detected by the Geiger counter which records the amount of radiation detected by these bursts. The Geiger counter beeps when it detects a burst of current and calculates the amount of radiation. The operation of the Geiger counter can be followed in Fig. 6.43.

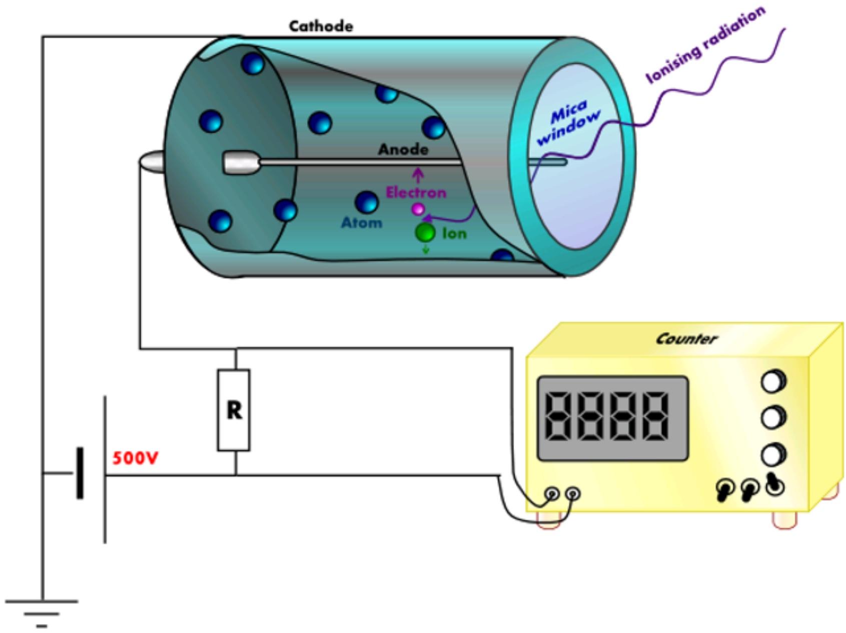


Figure 6.43 Geiger counter

A typical practical realization of the Geiger counter is shown in Fig. 6.44.

Figure 6.44 Practical GM tube with measuring device

Scintillation Counter

Detecting radiation can also be accomplished with the use of a scintillator. A scintillator is a substance which emits light when struck by an ionizing particle. Scintillation counters can be used to detect all types of ionizing radiation and are therefore the most useful. A scintillation counter detects radiation by using a transparent crystal coat, usually phosphor that emits bright flashes of light when ionizing radiation strikes the crystal. Electrons from the ionizing radiation effect are trapped in an excited state and emit a photon when they decay to the ground state. The emission of a photon causes a bright flash of light to appear called scintillations. The number of flashes and their energy are detected electronically with the scintillation counter. The information is then converted into electronic pulses, which are measured and recorded. The principle of operation of the scintillation detector is demonstrated in Fig. 6.45.

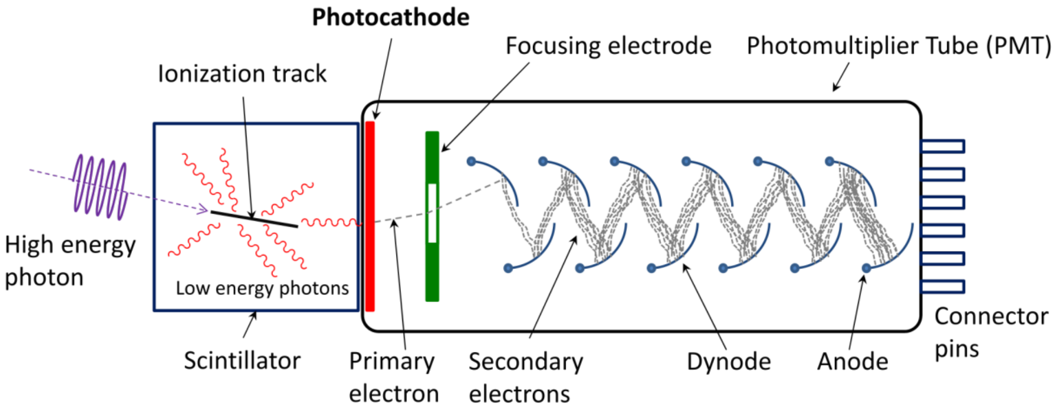


Figure 6.45 The scintillation detector

Semiconductor Detectors

In semiconductor detectors, ionizing radiation is measured by the number of charge carriers set free in the detector material, which is arranged between two electrodes, by the radiation. Ionizing radiation produces free electrons and holes. The number of electron-hole pairs is proportional to the energy of the radiation to the semiconductor. As a result, a number of electrons are transferred from the valence band to the conduction band, and an equal number of holes are created in the valence band. Under the influence of an electric field, electrons and holes travel to the electrodes, where they result in a pulse that can be measured in an outer circuit, as described by the Shockley-Ramo theorem. The holes travel in the opposite direction and can also be measured. As the amount of energy required to create an electron-hole pair is known and is independent of the energy of the incident radiation, measuring the number of electron-hole pairs allows the intensity of the incident radiation to be determined.

The energy required to produce electron-hole-pairs is very low compared to the energy required to produce paired ions in a gas detector. Consequently, in semiconductor detectors the statistical variation of the pulse height is smaller, and the energy resolution is higher. As the electrons travel fast, the time resolution is also very good, and is dependent upon rise time. Compared with gaseous ionization detectors, the density of a semiconductor detector is very high, and charged particles of high energy can give off their energy in a semiconductor of relatively small dimensions.

Semiconductor-based devices are built on intrinsic semiconductors i.e. 'clear' silicone or *High Purity Germanium* (HPGe). The main drawback of these detectors is their reasonable dark current. The negative effect of the dark current can to be compensated by applying a low operation temperature, for example, between -50 and -80 °C. These detectors are usually equipped with a liquid nitrogen cooler as shown in Fig. 6.46.

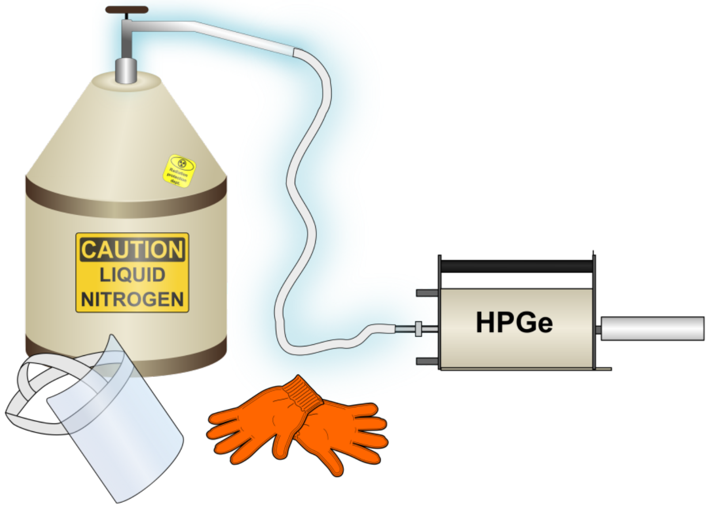


Figure 6.46 HPGe detector with reasonable dark current

6.3 Mechanical Sensors

Mechanical measurement is an umbrella term that covers a wide range of tests, which can be divided broadly into several types. According to very extensive technological development in this field of testing, there are number of new methods and applications introduced in practical measurements. In the followings we focus on the most important measuring methods and applied sensor devices for testing force, torque, pressure (or tension) strain (or deformation) and position.

6.3.1 Strain Gauges

Strain is the amount of deformation of a body due to an applied force. More specifically, strain (ε) is defined as the fractional change in length, as described in (6.17).

$$\varepsilon = \frac{\Delta l}{l} \tag{6.17}$$

Strain can be positive (tensile) or negative (compressive).

The principle of operation of electro-mechanical sensors, which are used for force, torque and pressure measurements, is based on the strain (deformation) dependent electrical resistance i.e. deformation causes a change in electrical resistance. Resistance for a metal thread can be described as in (6.18).

$$R = \rho \frac{l}{A} \tag{6.18}$$

Thus, the fractional change in resistance is

$$R = \rho\frac{l}{A} \rightarrow \frac{\Delta R}{R} = \frac{\Delta \rho}{\rho} + \frac{\Delta l}{l} - \frac{\Delta A}{A} \tag{6.19}$$

When a thread is strained by a uniaxial force, a phenomenon known as Poisson Strain causes a contract in the transverse, or perpendicular, direction. The magnitude of this transverse contraction is a material property indicated by its Poisson's Ratio (μ). Poisson's Ratio for steel, for example, ranges from 0.25 to 0.3. Thus, a fractional change in the cross section caused by strain is given in (6.20)

$$\frac{\Delta A}{A} = -2\mu\varepsilon \tag{6.20}$$

While there are several methods of measuring strain, the most common is with a strain gauge, a device whose electrical resistance varies in proportion to the amount of strain it encounters.

In practical applications either metal alloys or silicon alloys (semiconductors) are used in strain sensors. The most commonly used metal alloy is constantan (Cu:58%, Ni:42%), even though semiconductor-based detectors are more sensitive. The main disadvantage of semiconductor-based sensors is their non-linear characteristics.

The main disadvantage of semiconductor-based sensors is their non-linear characteristics.

A metallic strain gauge consists of a very fine wire or, more commonly, metallic foil arranged in a grid pattern. The grid pattern maximizes the amount of metallic wire or foil subject to strain in the parallel direction as shown in Fig. 6.47.

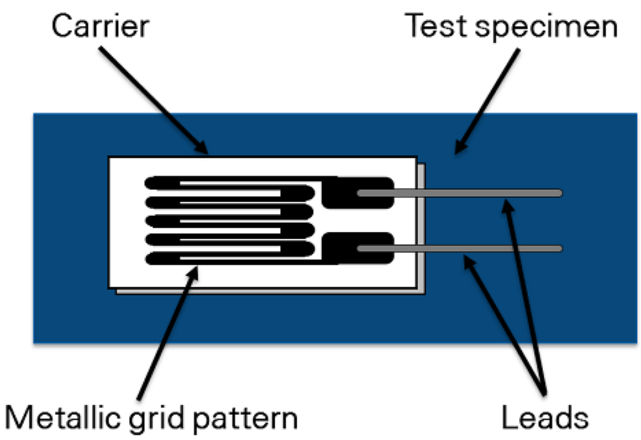


Figure 6.47 Strain gauge

The cross-sectional area of the grid is minimized to reduce the effect of shear strain and Poisson strain. The grid is bonded to a thin backing, called the carrier, which is attached directly to the test specimen. Therefore, the strain experienced by the test specimen is transferred directly to the strain gauge, which responds with a linear change in electrical resistance. Strain gauges are available commercially with nominal resistance values from 30 to 3000 Ω, with 120, 350, and 1000 Ω being the most common values.

A fundamental parameter of the strain gauge is its sensitivity to strain, expressed quantitatively as the gauge factor (g). The gauge factor is defined as the ratio of fractional change in electrical resistance to the fractional change in length (strain).

$$g = \frac{\Delta R / R}{\varepsilon} \tag{6.21}$$

Substituting (6.20) into (6.21) the gauge factor can be written also as in (6.22).

$$g = \frac{\Delta\rho / \rho}{\varepsilon} + 1 + 2\mu \tag{6.22}$$

Although, semiconductor-based strain sensors have better sensitivity, metal sensors are also widely used because of their ('close to') linear characteristics. The Gauge factor for metallic strain gauges is typically around 2. Strain sensitivity for different materials is demonstrated in Fig. 6.48.

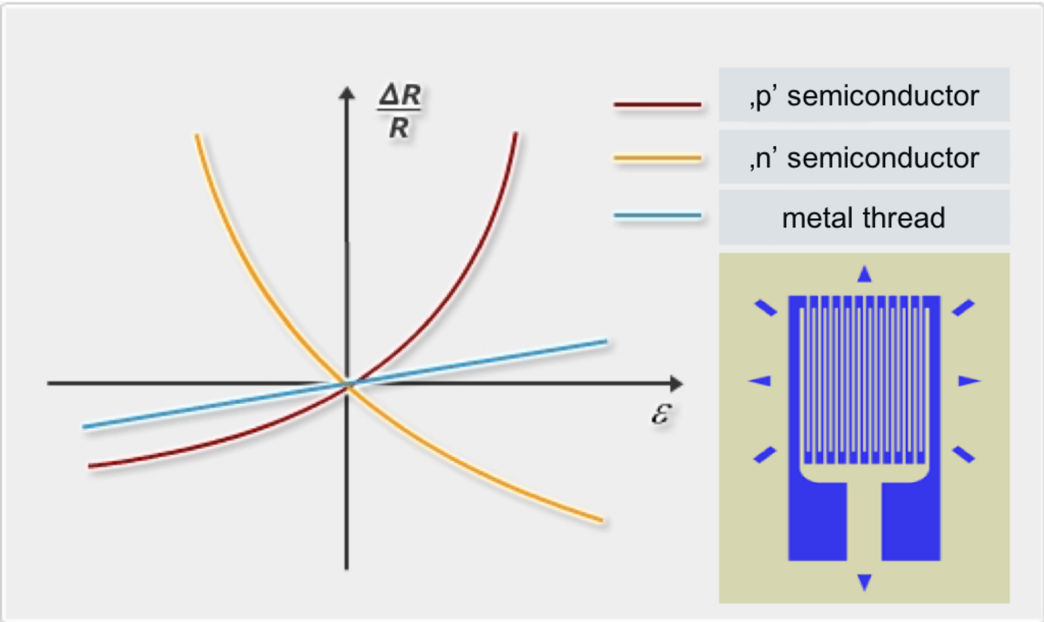


Figure 6.48 Strain sensitivity for different materials

A practical application of strain gauge for force measurement is shown in Fig. 6.49, for example.

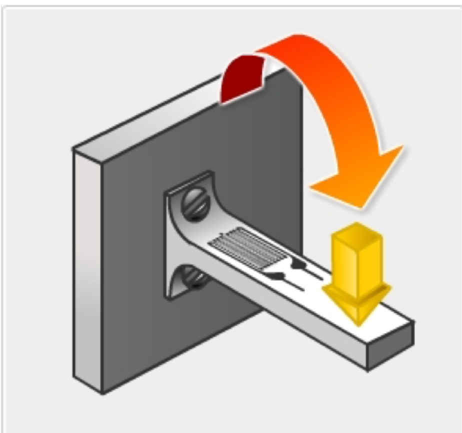


Figure 6.49 Force measurement using strain gauge

Ideally, we would like the resistance of the strain gauge to change only in response to applied strain. However, the strain gauge material, as well as the specimen material to which the gauge is applied, will also respond to changes in temperature. Strain gauge manufacturers attempt to minimize sensitivity to temperature by processing the gauge material to compensate for the thermal expansion of the specimen material for which the gauge is intended. While compensated gauges reduce most thermal sensitivity, they do not remove it completely. Therefore, additional temperature compensation is important.

6.3.1.1 Strain Gauge Cell

To measure small changes in resistance and compensate for the temperature sensitivity, strain gauges are almost always used in a bridge configuration with a voltage or current excitation source. The general Wheatstone bridge, illustrated in Fig. 6.50, consists of four resistive arms with an excitation voltage, V_{EX}, that is applied across the bridge.

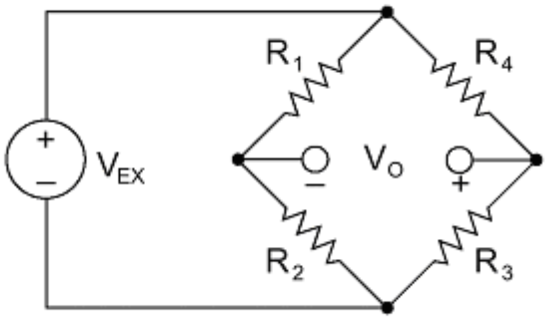


Figure 6.50 Wheatstone bridge

The output voltage of the bridge is given in (6.23).

$$V_0 = V_{EX}\left[\frac{R_1}{R_3 + R_4} - \frac{R_2}{R_1 + R_2}\right] \tag{6.23}$$

From this equation, it is apparent that when $R_1/R_2 = R_4/R_3$, the voltage output V_O will be zero. Under these conditions, the bridge is said to be *balanced*. Any change in resistance in any arm of the bridge will result in a non-zero output voltage. Therefore, if we replace R_4 with an active strain gauge, any changes in the strain gauge resistance will unbalance the bridge and produce a non-zero output voltage. If the nominal resistance of the strain gauge is designated as R_G, then the strain-induced change in resistance, ΔR, can be expressed as the following.

$$\Delta R = gR_G\varepsilon \tag{6.24}$$

Assuming that $R_1 = R_2$ and $R_3 = R_G$, the bridge equation above can be rewritten to express V_O/V_{EX} as a function of strain

$$\frac{V_0}{V_{EX}} = \frac{g\varepsilon}{4}\left(\frac{1}{1 + g\frac{\varepsilon}{2}}\right) \tag{6.25}$$

Note the presence of the bracketed part in (6.25) indicates the non-linearity of the quarter-bridge output with respect to strain.

More details and other aspects of the bridge circuits in measurement are described in Chapter 2, Classic Electrical Measurements.

6.3.2 Piezoelectric Accelerometers

Piezoelectric accelerometers are accelerometers that employ the piezoelectric effect of certain materials to measure dynamic changes in mechanical variables (e.g. acceleration, vibration, and mechanical shock). The principle of piezoelectric behaviour is shown in Fig. 6.51.

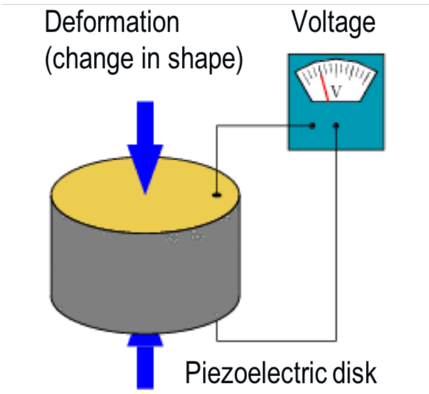


Figure 6.51 Piezoelectric behaviour

Note that strain gauges can also be used in accelerometers as sensor devices, but it is more common in practice to apply piezo detectors.

Piezoelectric accelerometers convert mechanical energy into electrical energy and provide an electrical signal in response to a quantity, property, or condition that is being measured. Using the general sensing method upon which all accelerometers are based, acceleration acts upon a seismic mass that is restrained by a spring or suspended on a cantilever beam and converts a physical force into an electrical signal. Before the acceleration can be converted into an electrical quantity it must first be converted into either a force or displacement. This conversion is done via the mass spring system shown in Fig. 6.52.

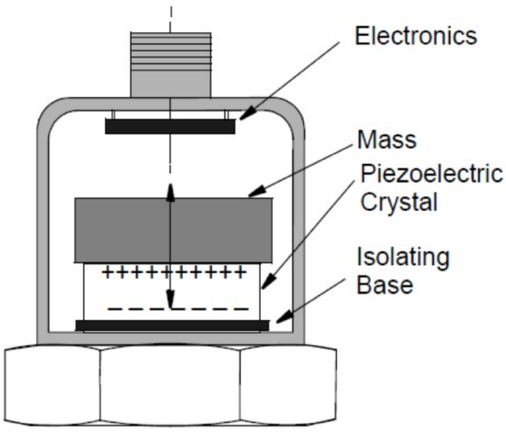


Figure 6.52 Piezoelectric accelerometer

The internal seismic mass is mounted by a pre-loaded (tuned) spring that can also be used for calibration and validation purposes. Piezoelectric accelerometers are also equipped with internal electronics, so called *Integrated Circuit Piezoelectric* (ICP) or *Integrated Electronics, PiezoElectric* (IEPE) amplifiers. (ICP is trademark of PCB Piezotronics Inc.) These inseparable amplifiers have the role of impedance matching because piezoelectric crystals have extremely high internal impedance.

6.3.3 Position and Displacement Sensors

The simplest displacement sensors that have long been used in many fields of mechatronics are potentiometers. They can be used as position transducers, operated by a mechanism.

A potentiometer is a three-terminal resistor with a sliding or rotating contact that forms an adjustable voltage divider.

Besides simple voltage division, used by potentiometers, several other principles can be used for position and displacement detection, like inductive, capacitive, magneto strictive, magneto resistive, galvanomagnetic and radiative principles.

A very basic and widely used transducer, called the *Linear Variable Displacement Transducer* (LVDT), applies the inductive principle, which is discussed in more detail below.

6.3.3.1 Linear Variable Displacement Transducer

A LVDT works under the principle of mutual induction, and displacement which is a non-electrical energy is converted into electrical energy. The construction of a LVDT is shown in Fig. 6.53. It consists of a cylindrical former which is surrounded by one primary winding in the centre of the former and the two secondary windings at the sides. The number of turns in both the secondary windings are equal, but they are opposite to each other. If the left secondary winding is in the clockwise direction, the right secondary windings will be in the anti-clockwise direction, hence the net output voltage will be the difference of the voltages between the two, secondary coils. The two secondary coils are represented as S1 and S2. An iron core is placed in the centre of the cylindrical former which can move in a to and fro motion as shown in the Fig. 6.53.

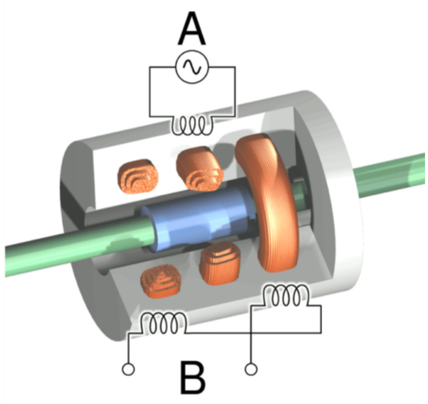


Figure 6.53 Construction of LVDT

Studying the working of LVDT we split the cases into 3 movements, based on the iron core position inside the insulated casing.

Case 1 - On applying an external force which is the displacement, if the core remains in the null position without any movement then the voltage induced in both the secondary windings are equal which results in a net output equal to zero as in (6.26).

$$E_{OUT} = E_{s1} - E_{s2} = 0 \tag{6.26}$$

Case 2 - When an external force is applied and if the steel iron core tends to move in the left-hand direction then the voltage induced in the secondary coil 1 is greater than the voltage induced in the secondary coil 2. Therefore, the output voltage is given in (6.27).

$$E_{OUT} = E_{s1} - E_{s2} > 0 \qquad (6.27)$$

Case 3 - When an external force is applied and if the steel iron core moves in the right-hand direction then the voltage induced in the secondary coil 2 is greater than the voltage induced in the secondary coil 1. Therefore, the output voltage is given in (6.28).

$$E_{OUT} = E_{s2} - E_{s1} > 0 \qquad (6.28)$$

The output voltage i.e. the magnitude of differential AC output of the LVDT versus displacement is shown in Fig. 6.54. When displacement is small, the core iron is still 'close' to the balanced position (case 1), and the output characteristic is a linear function. But, according to the AC voltage, we cannot distinguish the direction of displacement. The magnitude of the AC output is positive in both, case 2 and case 3 displacements.

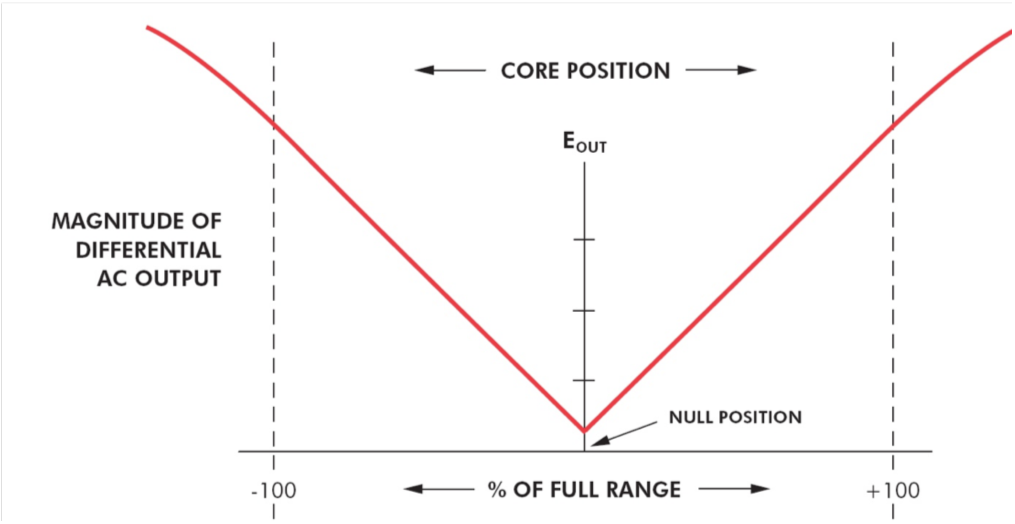


Figure 6.54 Output voltage of LVDT

According to (6.27) and (6.28) there is 180° phase step in the output voltage in case 2 and case 3 that can be detected with applied electronics. Thus, the DC output voltage from electronics is a linear function of displacement in full operation range. The DC output and the phase angle between the primary and secondary voltages are shown in Fig. 6.55.

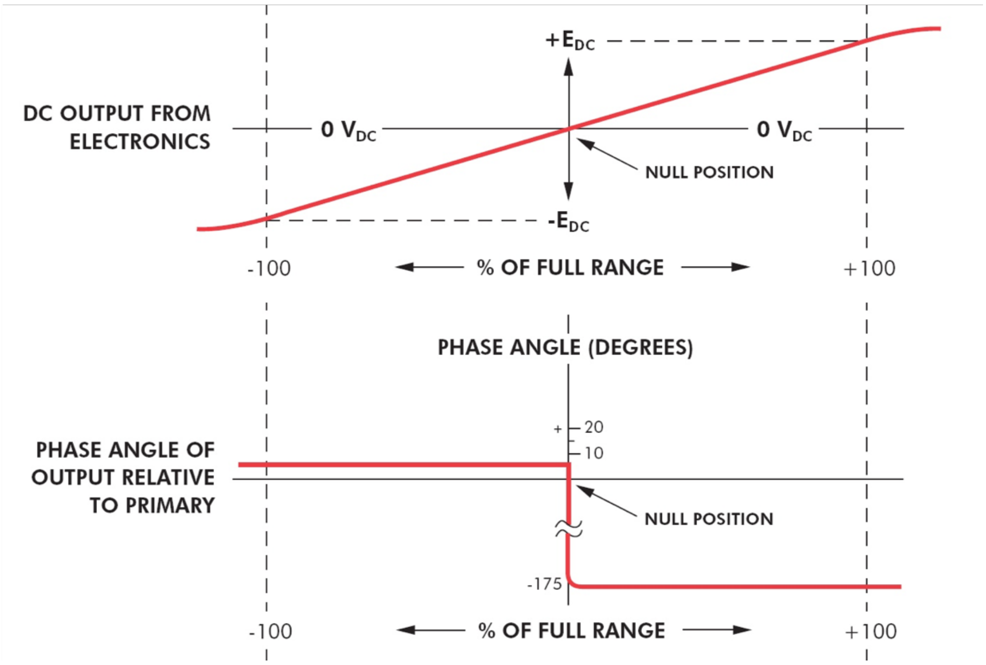


Figure 6.55 DC output and phase shift of LVDT

The main advantages of LVDT are its high sensitivity, good linearity, low power consumption and low hysteresis. There are some drawbacks, such as shielding is required (since it is sensitive to the magnetic field), the performance of the transducer gets affected by vibrations and it is greatly affected by temperature change.

6.3.3.2 Selsyn (or Synchro)

Finally, we briefly cover an electro-mechanical device that can be used for angular data transmission between remote points. Selsyn (short for self-synchronous), but also known as synchro, is a transformer whose primary-to-secondary coupling may be varied by physically changing the relative orientation of the two windings. Selsyns are often used for measuring the angle of a rotating machine such as an antenna platform. In its general physical construction, it is much like an electric motor. The primary winding of the transformer, fixed to the rotor, is excited by an alternating current, which by electromagnetic induction, causes currents to flow in three Y-connected secondary windings fixed at 120 degrees to each other on the stator. The relative magnitudes of secondary currents are measured and used to determine the angle of the rotor relative to the stator, or the currents can be used to directly drive a receiver synchro that will rotate in unison with the synchro transmitter. In the latter case, the whole device may be called a selsyn (portmanteau of *self* and *synchronizing*). The principle of operation is shown in Fig. 6.56.

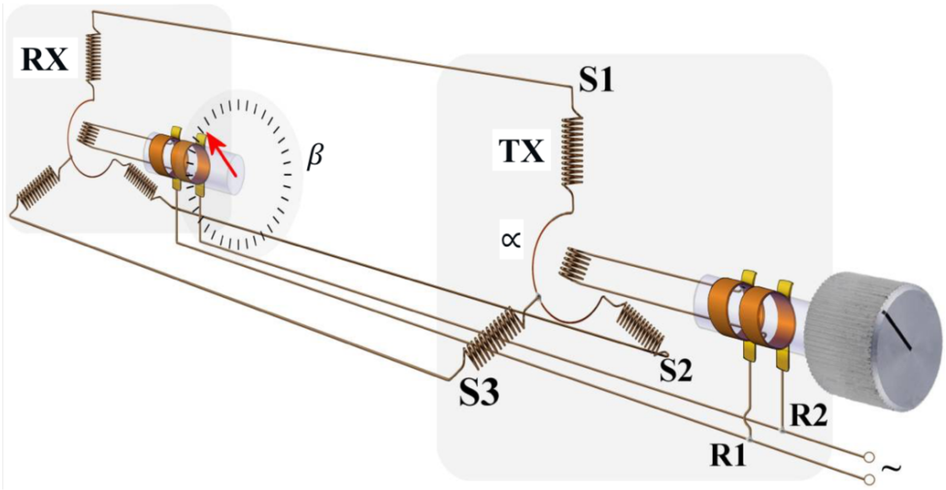


Figure 6.56 Principle of selsyn

The primary coils of the transformers in both, the transmitter (TX) and receiver (RX) side are supplied by an AC source i.e. by 230-V, 50-Hz network. The angle position dependent inducted voltages in the secondary coils at the transmitter and the receiver side are the same if the coil positions are also the same $(\alpha = \beta)$. Thus, in this resting position there is no electric current in the three phase line system. When the primary coil is rotated on the transmitter side balancing currents will flow in the three-phase lines causing a synchronous torque that is proportional to the phase difference described in (6.29).

$$M_{TX} = -M_{RX} = M_m \sin \varepsilon \leftarrow \varepsilon = \alpha - \beta \qquad (6.29)$$

This synchronous torque rotates the secondary coil of the receiver until the balanced condition $(\alpha = \beta)$ occurs again.

A selsyn can be used in fire-control systems, for example to transmit angular information from a target tracking system to guns. Smaller selsyns are used to remotely drive indicator gauges and as rotary position sensors for aircraft control surfaces, where the reliability of these rugged devices is essential. Selsyn motors were also widely used in motion picture equipment to synchronize movie cameras and sound recording equipment.

7. Reference Sources and Recommended Materials

7.1 In English

1. S. Tumanski: Principles of electrical measurement, CRC Press, 2006
   ISBN 0-7503-1038-3
2. T. Briggs: Introduction to Oscilloscopes
   Dept. of Computer Science & Engineering Shippensburg University of Pennsylvania
3. Rhode & Schwarz: Oscilloscope Overview, Primer, Rich Markley, 06.2015 – 1.0
   (www.rohde-schwarz-usa.com/Scope-Resources.html)
4. J. Fraden: Handbook of Modern Sensors, Springer NY. 2010
   ISBN 978-1-4419-6465-6
5. Bharathidasan, V. A. Sai Ponduru: Sensor Networks: An Overview,
   University of California, Davis, CA 95616
6. National Instruments: Strain Gauge Measurement – Tutorial, NI Co, 1998
   Application Note 078, 341023C-01
7. BIPM (Bureau International des Poids et Measures): International Vocabulary of
   Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition)
   JCGM 200:2012
8. I. Gyurcsek: Innovative Scanning Method in Non-Destructive Roof Moisture Diagnostics
   35th International Scientific Conference (Science in Practice 2017) Pécs, Hungary
   Published by University of Pécs (https://sip.mik.pte.hu) ISBN: 978-963-429-130-5
9. Gy. Elmer, I. Gyurcsek: Lessons from a Loop Impedance Test,
   34th International Scientific Conference (Science in Practice 2016), Subotica, Serbia
   Published by Subotica Tech – College of Applied Sciences
   (http://www.vts.su.ac.rs/en/page/sip-2016)
   ISBN: 978-86-918815-1-1
10. I. Gyurcsek, T. Pósa: AyCare: a sensory presence for the elderly and disabled (paper id: 76)
    Regional Conference of Embedded and Ambient Systems (RCEAS) , Budapest, 2007

7.2 In Hungarian

9. Lambert M.: Szenzorok – elmélet és gyakorlat, Budapest, 2009
   ISBN 978-963-874001-1-3
10. Máté J.: Méréstechnika 1, PTE PMMIK, ERFP-DD2001-HU-B-01
11. Dr. Petróczky K.: Bevezetés a nyúlásmérő bélyeges méréstechnikába
    SZIE, Gödöllő (kézirat)
12. Elmer Gy., Gyurcsek I.: Az érzékeny elektronika esete a hurokimpedancia-méréssel
    Elektrotechnika MEE 109. évf. 2016/9 12-14. old.
13. Gyurcsek I., Ring G.: Lapostetők és falak nedvességtartalmának roncsolásmentes mérése
    Mérés és Automatika 1987./2.sz. 64-67.old
14. Gyurcsek I., Ring G.: Hazai mérőműszer tetőfödémek nedvesség diagnosztikai
    vizsgálatához, Mérés és Automatika 1984./6. 236-239. old.