EE 306 – ELECTRICAL ENGINEERING TECHNOLOGIES

LECTURE NOTES

PREPARED BY

Prof. Dr. Bahattin Karagözoğlu

INTRODUCTION

FUNDAMENTAL ELECTRICAL ENGINEERING COMPONENTS

MEASUREMENT AND ERROR

MEASUREMENT OF ELECTRICAL QUANTITIES

OSCILLOGRAPHIC MEASUREMENTS AND PICTURE DISPLAYS

SOURCES OF ELECTRICAL ENERGY

TEMPERATURE MEASUREMENT

MEASUREMENT OF DISPLACEMENT AND MECHANICAL STRAIN

PRACTICAL AND REPORTING

KING ABDULAZIZ UNIVERSITY
FACULTY OF ENGINEERING
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
JEDDAH – SAUDI ARABIA

Shawwal 1432H – September 2011G

Table of Contents

Detailed Table of Contents

INTRODUCTION

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Define electrical and electronics engineering.
2. State the responsibilities of and career opportunities for graduates of electronics and communications, computer and biomedical engineering groups.
3. Express novel and emerging application fields of electronics engineering such as mechatronics, avionics.
4. Recognize the applications of electronics engineering in automotive Industry, e-health, biomechanics and rehabilitation, cognitive radio and fiber optics communication systems.
5. Define basic and derived units in engineering.
6. Identifies engineering standards and standard units for a given application.
7. Use engineering prefixes in expressing numerical values.

ELECTRICAL AND COMPUTER ENGINEERING SPECIALTIES

Definition of Electrical and Electronic Engineering

Electrical engineering is an engineering discipline that deals with the study and application of electricity and electromagnetism. Its practitioners are called electrical engineers. Electrical engineering is a broad field that encompasses many subfields and after 1980 it is generally referred to the engineering discipline that deals with electrical energy and its utilization. It has two major branches:

Power engineering: generation, distribution and utilization of electrical energy

Machines engineering: conversion of electrical energy into mechanical action and work

Electronics Engineering is a specialized branch of Electrical Engineering which deals with components such as semiconductor diodes, triodes, transistors, computer and similar microcircuit chips, printed circuit boards, etc. Depending on where they are to be used (the applications), electronic circuits can be built to handle a very wide range of power. Electronics is the study and use of electrical devices that operate by controlling the flow of electrons or other electrically charged particles in devices such as thermionic valves and semiconductors. The pure study of such devices is considered as a branch of physics, while the design and construction of electronic circuits to solve practical problems is part of the fields of electrical, electronic and computer engineering. Figure 1.1 illustrates a functional diagram of electronics engineering.

Electronics Engineering (also referred to as electronic engineering) is an engineering discipline which uses the scientific knowledge of the behavior and effects of electrons to develop components, devices, systems, or equipment (as in electron tubes, transistors, integrated circuits, and printed circuit boards) that uses electricity as part of its driving force. Both terms denote a broad engineering field that encompasses many subfields including those that deal with power, instrumentation engineering, telecommunications, semiconductor circuit design, and many others.

The electronics engineering deals with communicating an information from one place into another place and developing tools and techniques to achieve it. It takes a physical process that is in form of mechanical and chemical in nature and converts them into electrical variables in form of voltage and current or other derived electrical variables. A device that converts a type of energy into another type is called the transducer. It is called the sensor if the converted energy is electrical. The information flow is in form of flow of electrons in electrical circuits. Several electronic utilities are used to process the signal including amplifiers, filters, analog to digital and digital to analog converters and digital computers.

The computer is a programmable machine that receives input, stores and manipulates data, and provides output in a useful format. Computer Engineering is a branch of engineering that deals with the machine (hardware) and programs (software) that are used to operate the machines (system and applications). Computer engineering has two major branches as computer hardware and software (system and applications). The software part is called as the computer science. Computer hardware and electronics have many components in common and they are almost remerging. It deals with computer networks, interfacing computers with other electronic and non-electronic devices, embedded systems, robotics, vision and control systems, and computer graphics.

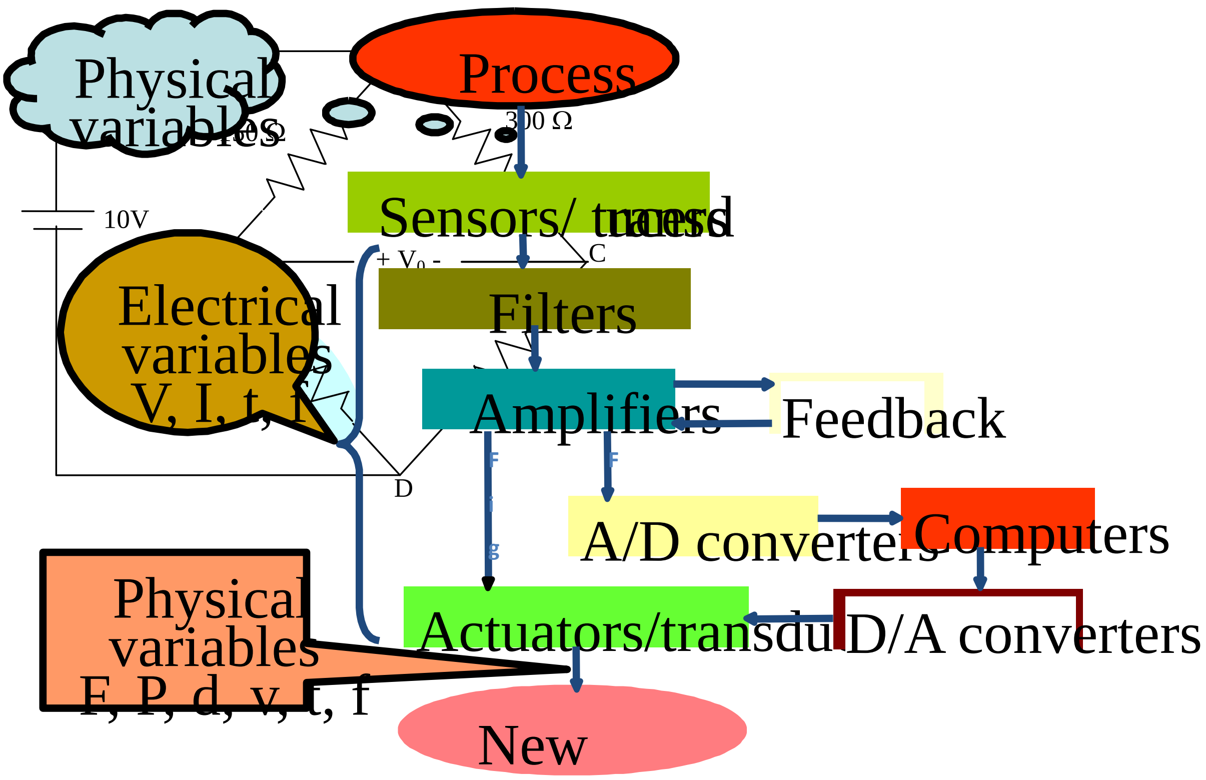


Figure 1. 1 A schematic diagram illustrating various activities of electronics engineering

The information perceived by the user must be in form of mechanical and chemical processes. The electrical information is converted into this convertible form using another type of sensor that is called the actuator. Electronic engineering principles and devices are used in many other engineering disciplines such as telecommunications engineering, biomedical engineering, mechatronics and avionics.

The activities of electronics engineering are handled by three distinct groups in the Electronics and Computer Engineering in Faculty of Engineering at King Abdulaziz University.

Electronics and Communications Group

The Group is concerned with :

- The Electronics Engineering that covers electronic devices, circuits, systems, and measurement and measuring instruments,
- The Communications Engineering that deals with signals, signal processing, signal transmission and transmission mediums, noise and signal detection, and applications of electronic devices, systems and circuits in various areas of communication.

The Electronics and Communications Specialization has a very wide application area. Graduates from the specialty work in

- Installation, management and maintenance of variety of communication systems such as microwave and radar systems, optical and laser communication systems, and mobile communication systems etc.
- Design, construction, operation and maintenance of
- Electronic instrumentation in various industrial installations,
- Control systems, data logging stations and related instruments,
- Information technology and local area networks,
- Building management systems, and
- Electronic entertainment devices

Computer Engineering Group

The computer Engineering Group deals with computer hardware and software (systems and applications), computer networks, interfacing computers with other electronic and non-electronic devices, embedded systems, robotics, vision and control systems, and computer graphics.

Graduates from the specialty work in government and private organizations. Their responsibilities cover

- Design, construction, operation and maintenance of
- Computer networks,
- Information technology departments,
- Graphic workstations and electronic publishing utilities,
- Specialized computer labs,
- Interfacing computers in measurement and control applications, control systems and data logging applications,
- Computerized automotive systems,
- Computer Aided Design (CAD) and Computer Aided manufacturing (CAM) systems,
- Building management systems
- Development of operating systems for special applications,

- Database system design, operation and maintenance.

Biomedical Engineering Group

The biomedical engineering deals with applications of engineering principles and know-how in medicine and biology. The specialty areas are:

- bioinstrumentation,
- biomaterials;
- biomechanics;
- cellular, tissue and genetic engineering;
- clinical engineering;
- medical imaging;
- orthopedic surgery;
- rehabilitation engineering; and
- systems physiology

The program in our Department is concentrated around medical electronics that deals with measurement and processing of medical signals, and medical instrumentation for diagnostic, monitoring and therapeutic purposes.

- **Bioinstrumentation**: application of electronics, computers and measurement techniques to develop devices used in diagnosis and treatment of disease.
- **Medical Imaging**: combines knowledge of a unique physical phenomenon (sound, radiation, magnetism, etc.) with high speed electronic data processing, analysis and display to generate an image.
- **Clinical Engineering**: application of technology to health care in hospitals.

The clinical engineer is an engineer who is able to perform certain engineering tasks in a health care facility and who has the knowledge and experience to work as a partner with health professionals to plan and implement appropriate programs for improving the health care delivery. He is generally an in-house engineer working in the hospital to fulfill some of the following responsibilities:

- Supervision on proper operation and safety of instruments. Ensuring electrical safety in medical environment, preparation and follow-up of the preventive (operational) and corrective maintenance procedures for medical equipment;
- Specification and purchase of new equipment, and training of staff on its proper use;

- Working with physicians to adapt instrumentation to the specific needs of the physician and the hospital. This often involves modification of medical equipment to meet local needs; and the interface of instruments with computer systems and customized software for instrument control and data acquisition and analysis;
- Coordination of medical information flow between different departments in the hospital and introduction of industrial or management engineering techniques to optimize information handling; developing and maintaining computer databases of medical instrumentation and equipment records and for the purchase and use of sophisticated medical instruments.

A biomedical engineer in the medical instrumentation track is an engineer competent in medical electronics and computer applications in medicine. He may work in the biomedical engineering department of a hospital or in a private company that provides services to health care facilities. His major responsibilities include:

- Installation, planning and handling of maintenance procedures and repair of medical equipment under his responsibility;
- Designing of engineering systems and components of systems that are not commercially available;
- Preparation of bidding for maintenance contracts;
- Pursuing technological developments in the medical instrumentation field and enlightening medical personnel about them.

An electrical engineer specialized in the general field of instrumentation, measurement and control is an engineer who deals with signal detection, transduction, processing and information presentation techniques used in biomedical engineering that are also widely utilized in industrial applications. Hence, biomedical engineering graduates can easily adapt themselves into such applications.

MISCELLANEOUS ELECTRICAL ENGINEERING FIELDS OF ACTIVITIES

There are important application fields that are not currently covered in the Department of Electrical and Computer Engineering: mechatronics, avionics, biomechanics, rehabilitation engineering, e-health and telemedicine, cognitive radio and fiber optic communication systems.

Mechatronics

Mechatronics is the synergistic combination of precision mechanical engineering, electronic control and systems thinking in the design of products and manufacturing processes. It relates to the design of systems, devices and products aimed at achieving an optimal balance between basic mechanical structure and its overall control. It has extensions as the robotics, microelectromechanical systems (MEMS) and applications in automotive industry. The logo of mechatronics is shown in Figure 1.2 and the domains of its activities are illustrated in Figure 1.3.

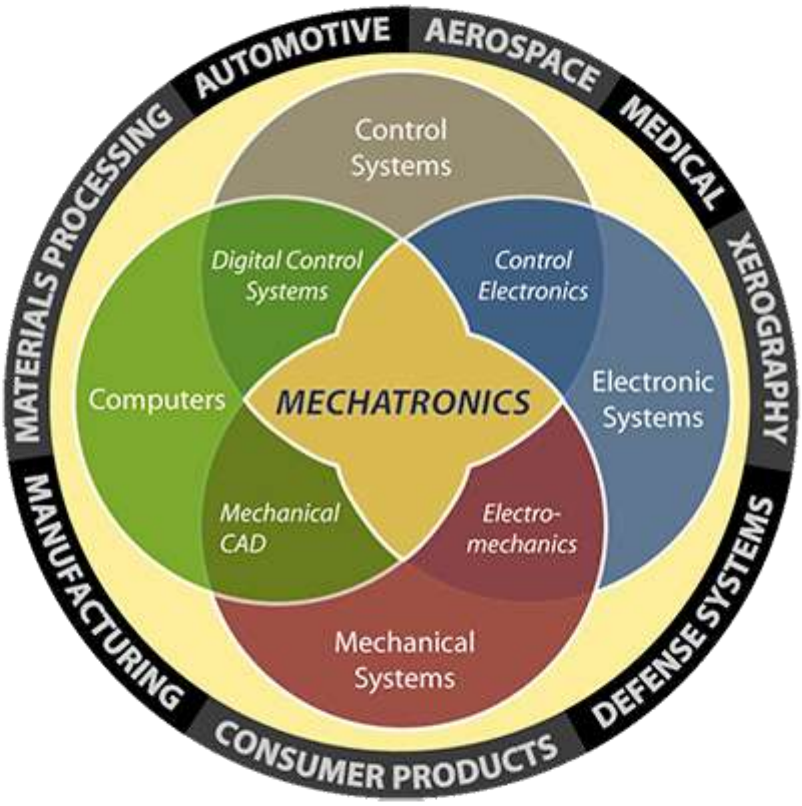


Figure 1. 2 Logo of mechatronics (source: http://www.edn.com/article/511901-PLM_and_mechatronics.php)

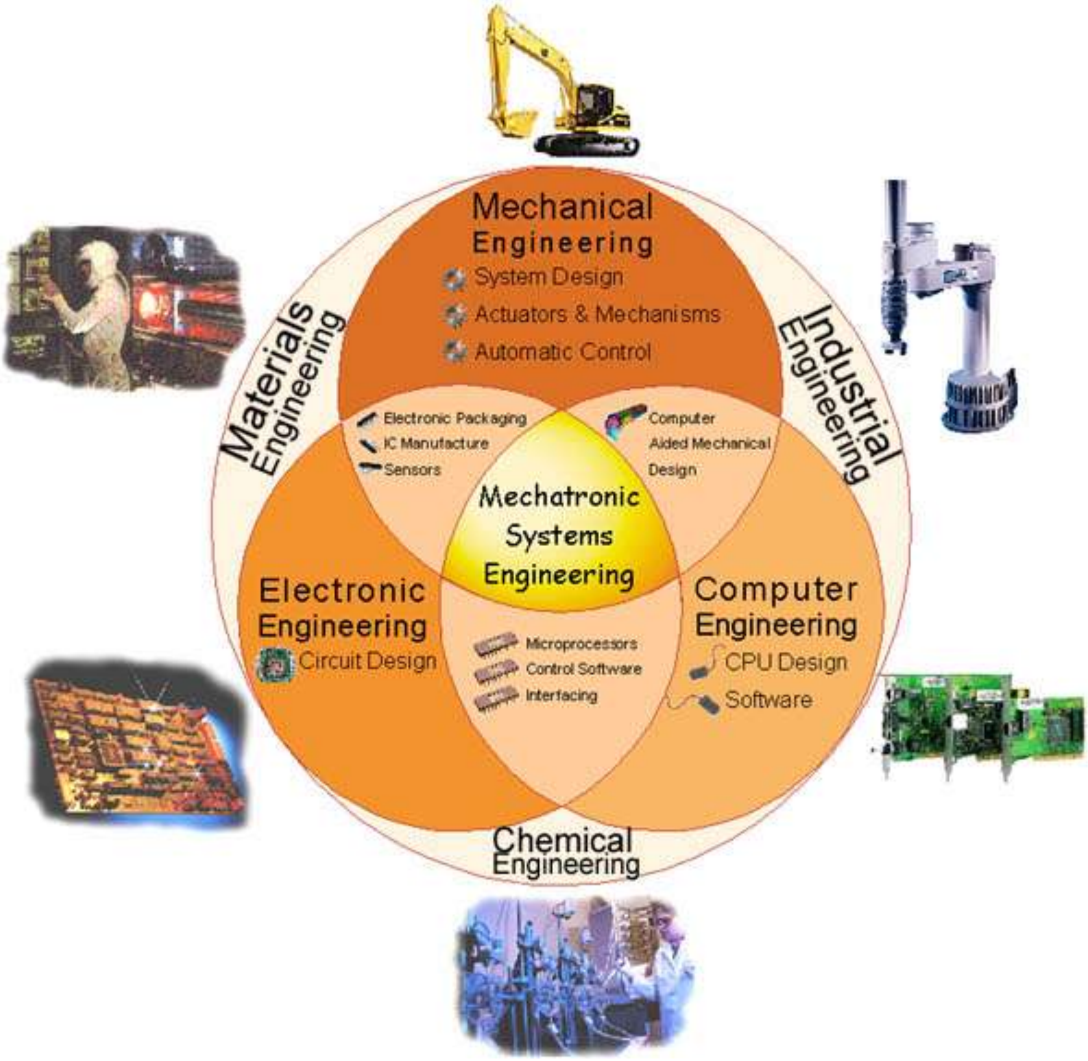


Figure 1. 3 Domain of activities of mechatronics (source: http://www.uomcoe.org/ar/index.php?option=com_content&view=article&id=580:2011-08-08-21-15-49&catid=10:2010-01-01-20-55-22&Itemid=140)

Robotics: a robot's design, manufacture, application, and structural disposition. It is related to electronics, mechanics, and software. Figure 1.4 shows a gripper (mechanical hand) which is a very challenging application.

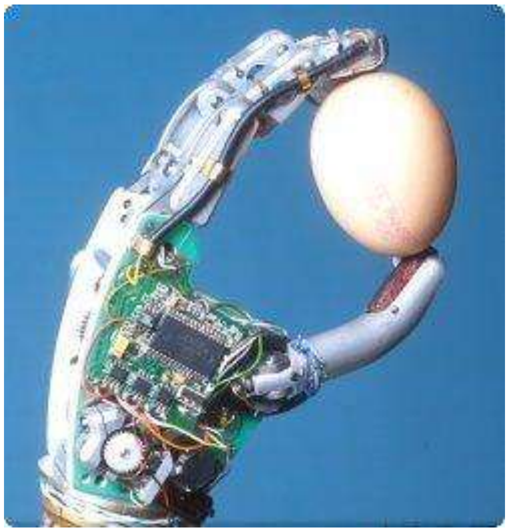


Figure 1. 4 A robot hand (gripper) (source: http://www.amazon.co.uk/Photographic-artificial-Science-Photo-Library/dp/B001NJ9DLY)

MicroElectroMechanical Systems (MEMS): technology of very small mechanical devices driven by electricity; it merges at the nano-scale into nanoelectromechanical systems (NEMS) and nanotechnology. MEMS are also referred to as micromachines (in Japan), or Micro Systems Technology - MST (in Europe). Figure 1.5

shows an assembly drawing for a safety lock and its interface using an optical fiber.

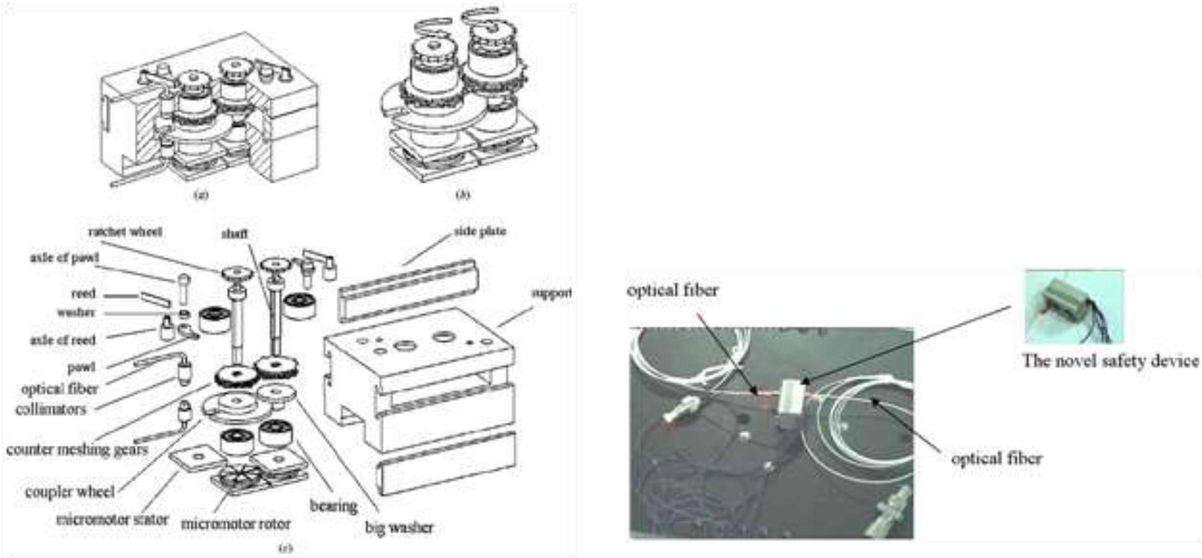


Figure 1. 5 A safety lock using MEMS technology and its interface (source: http://spie.org/x35991.xml?ArticleID=x35991)

Automotive Industry

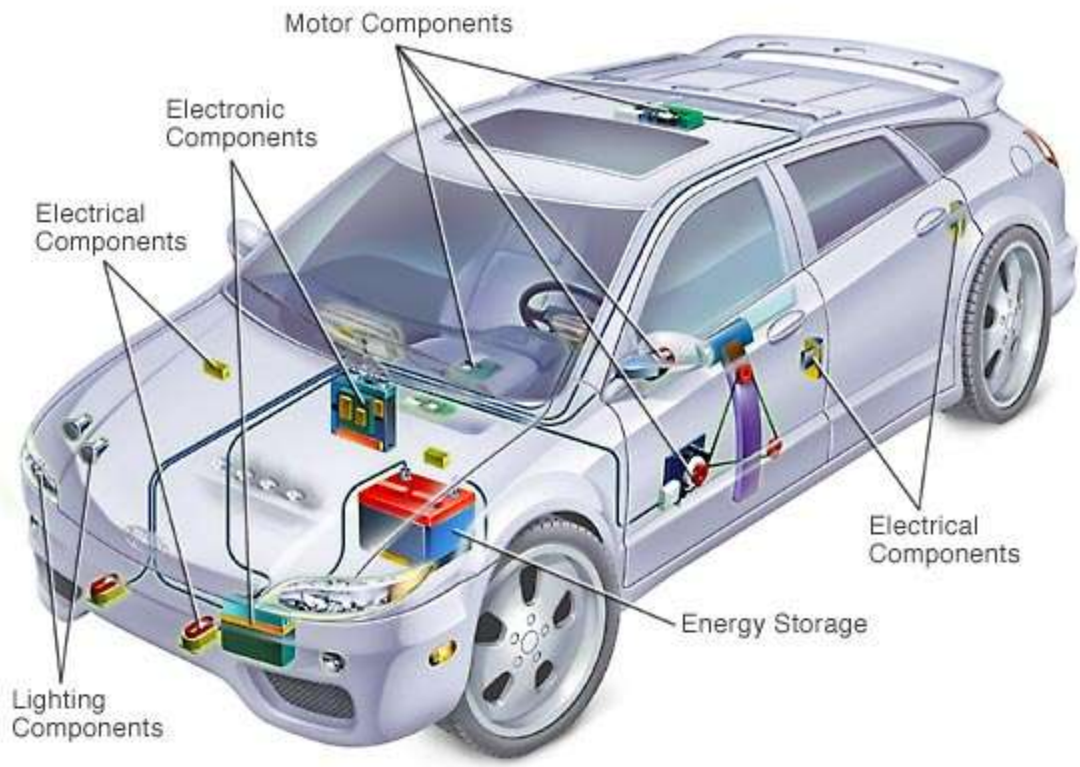


Figure 1. 6 Common electrical components in a car

The automotive industry contains many applications such as software design tools, electronic gadgets and controls, break by wire, GPRS, etc. Figure 1.6 shows the common electrical components in a car.

Figure 1.7 illustrates automotive electronics that ranges from entertainment and navigation systems into lighting and control systems.

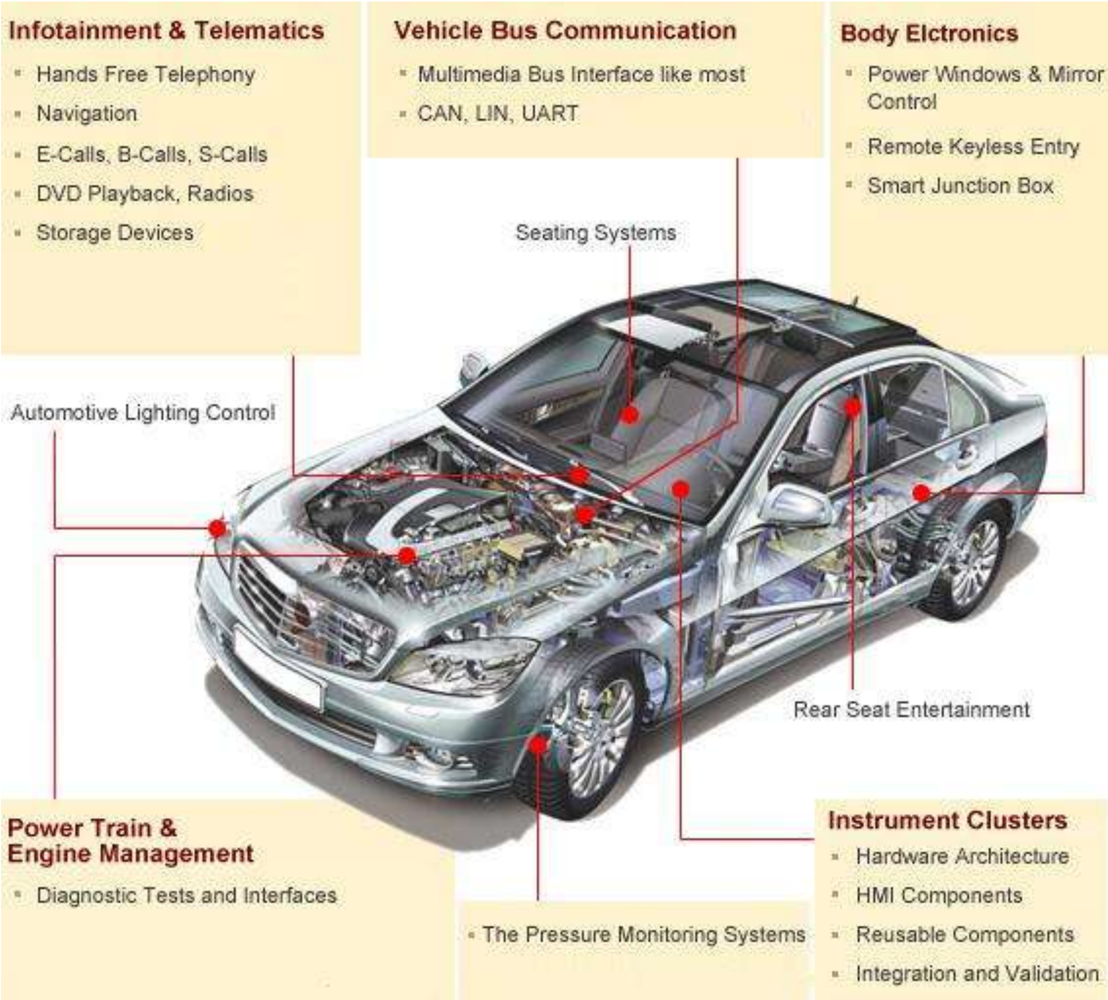


Figure 1. 7 Automotive electronic systems

Avionics

Avionics: combination of "aviation" and "electronics". It comprises electronic systems for use on aircraft, artificial satellites and spacecraft, comprising communications, navigation and the display and management of multiple systems. It also includes the hundreds of systems that are fitted to aircraft to meet individual roles, these can be as simple as a search light for a police helicopter or as complicated as the tactical system for an Airborne Early Warning platform. Figure 1.8 shows the control panel in the cockpit of an airplane.

Figure 1. 8 Cockpit of an airplane

Biomedical Engineering Extensions

e-health: relatively recent term for healthcare practice supported by electronic processes and communication, dating back to at least 1999 as illustrated in Figure 1.9.

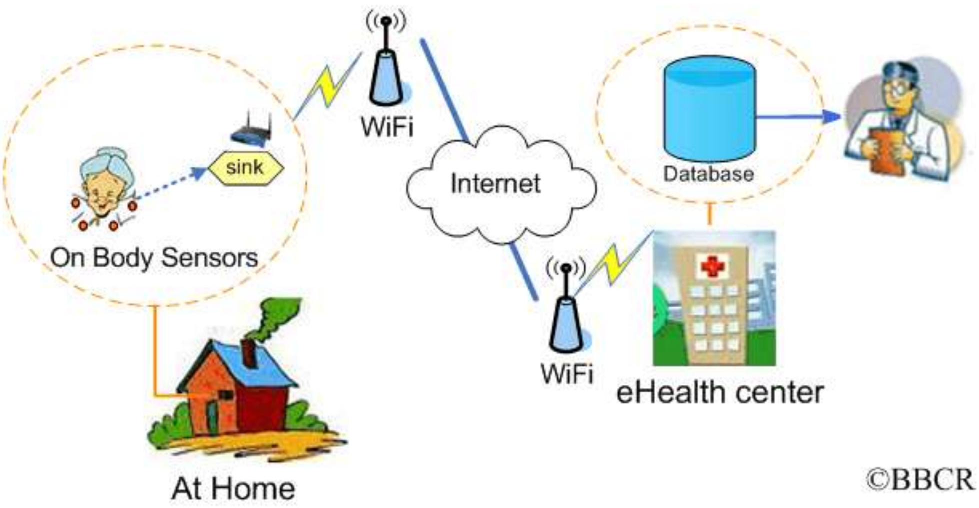


Figure 1. 9 Illustration of e-health technology

Rehabilitation: the process of helping an individual achieve the highest level of independence and quality of life possible - physically, emotionally, socially, and spiritually. Rehabilitation engineering is to develop tools and facilities for the disabled people to help them in recovery and gain independence in their activities. Figure 1.10 shows an instrumented wheelchair that provides mobility for the disabled.

Figure 1. 10 An instrumented wheelchair

Biomechanics: the application of mechanical principles to biological systems, such as humans, animals, plants, organs, and cells. Perhaps one of the best definitions was provided by Herbert Hatze in 1974: "Biomechanics is the study of the structure and function of biological systems by means of the methods of mechanics". The word biomechanics developed during the early 1970s, describing the application of engineering mechanics to biological and medical systems.

Biomechanics is close related to engineering, because it often uses traditional engineering sciences to analyze biological systems. Some simple applications of Newtonian mechanics and/or materials sciences can supply correct approximations to the mechanics of many biological systems. Applied mechanics, most notably mechanical engineering disciplines such as continuum mechanics, mechanism analysis, structural analysis, kinematics and dynamics play prominent roles in the study of biomechanics.

Usually biological system are more complex than man-built systems. Numerical methods are hence applied in almost every biomechanical study. Research is done in a iterative process of hypothesis and verification, including several steps of modeling, computer simulation and experimental measurements. Figure 1.11 shows a microprocessor controlled leg prosthesis.

Figure 1. 11 A microprocessor controlled prosthetic leg

Cognitive Radio

Cognitive radio is a paradigm for wireless communication in which either a network or a wireless node changes its transmission or reception parameters to communicate efficiently avoiding interference with licensed or unlicensed users. This alteration of parameters is based on the active monitoring of several factors in the external and internal radio environment, such as radio frequency spectrum, user behavior and network state. Figure 1.12 illustrates the operation of the cognitive radio.

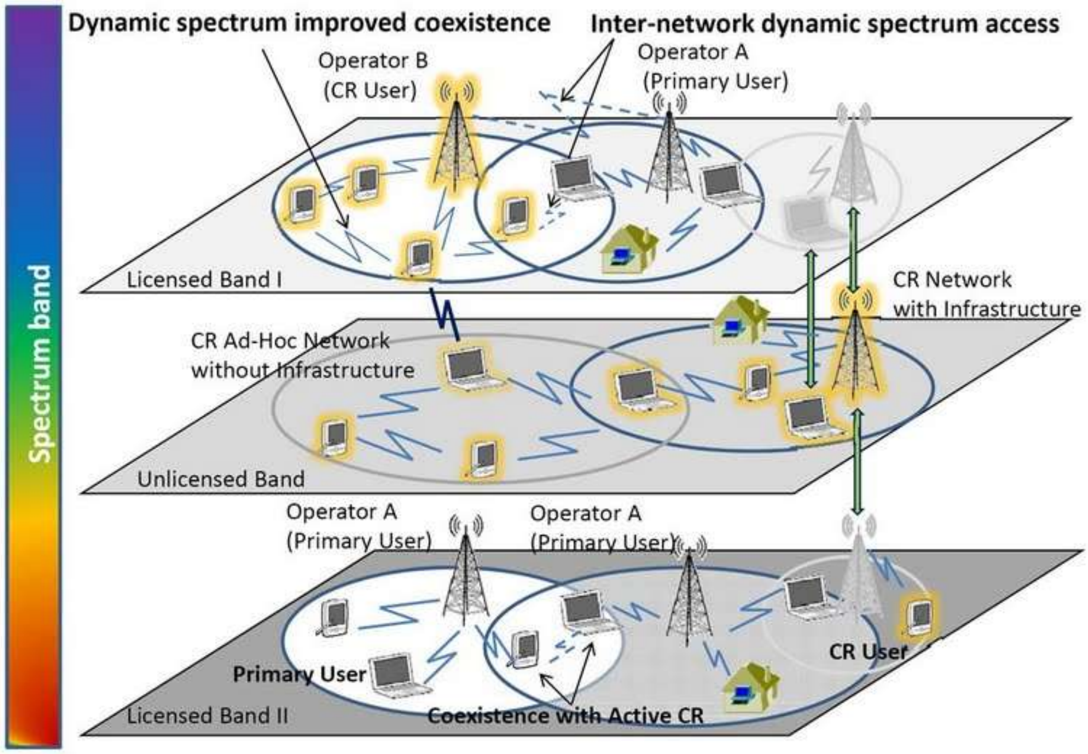


Figure 1. 12 The principles of operation of the cognitive radio

Fiber Optics Communication Systems

Optical communication is as old as the humanity. Optical communication systems in the past consisted of techniques such as fire signals, smoke signals, flash lanterns, reflected sunlight and signal flags. Such systems had limited bandwidth and were not competitive with electronic communications (like radio). The invention of the laser however provided a coherent optical source capable of transmitting information at extremely high data rates. Yet, limitations on transmission of light through the atmosphere (such as turbulence, haze, fog, absorption and rain) limited the usefulness of lasers for transmission of information through the atmosphere. Modern optical communication systems use semiconductor lasers that transmit light through optical fibers. Such systems have become widely used for telecommunications. Laser communication systems are used to transfer information from one point to a distant point. The information may be an audio conversation, a stream of data from one computer to another, or several simultaneous television broadcasts. The distance may range from a few feet to thousands of miles.

Industrial revolution of 19th century gave way to information revolution during the 1990s. Table 1 illustrates the milestones of developments in electrical and optical era. The optical era started in late 70's but experienced a speedy development after 90's. Emergence of internet caused a new age in which the world is reshaping and the Fiber-Optic Revolution is a natural consequence of the Internet growth. The information flow is managed at a much economical rates yet with a very high throughput via the optical communication systems.

Table 1. Milestones of developments in electrical and optical era

Electrical Era	Optical Era
• Telegraph; 1836	• Optical Fibers; 1978
• Telephone; 1876	• Optical Amplifiers; 1990
• Coaxial Cables; 1840	• WDM Technology; 1996
• Microwaves; 1948	• Multiple bands; 2002
Microwaves and coaxial cables limited to B 100 Mb/s.	Optical systems can operate at bit rate >10 Tb/s. Improvement in system capacity is related to the high frequency of optical waves (200 THz at 1.5 μm).

Fiber optic is applied in parts of our life now from connecting peripheral devices up to advanced telecommunication systems as illustrated in Figure 1.13. The bandwidth extends from a few Hz up to 10 GHz and the length covered ranges from a few meters up to thousands of kilometers.

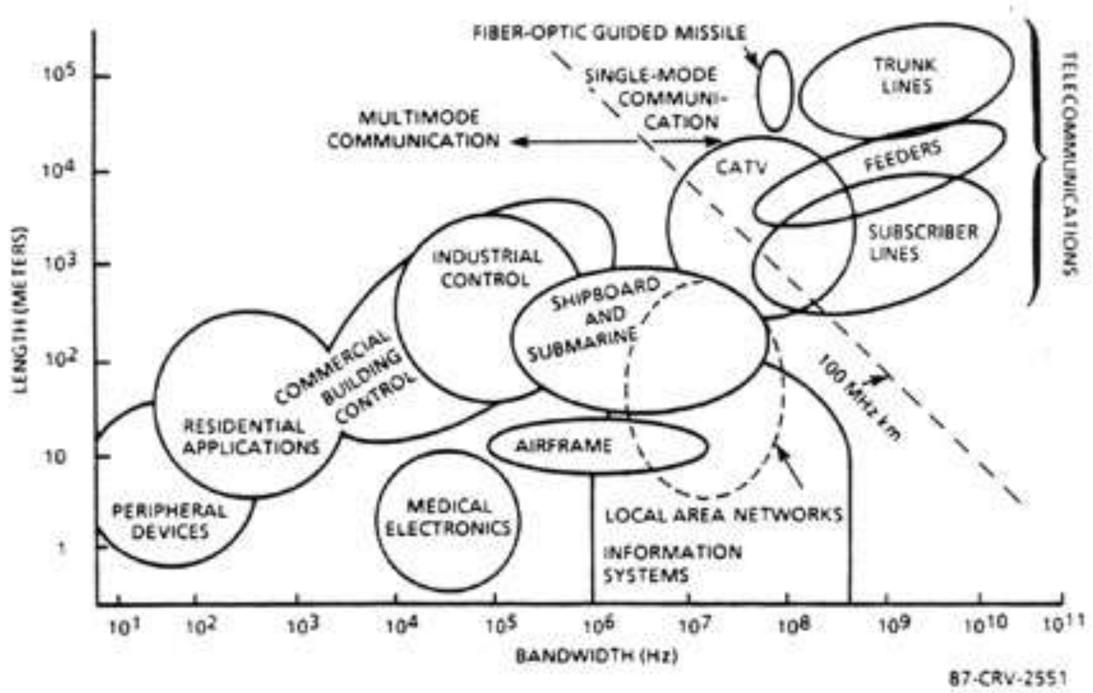


Figure 1. 13 Typical fiber optic applications

From: www.master-photonics.org/uploads/media/Govind_Agrawal1.pdf

Components of a light wave system is illustrated in Figure 1.14. A generic system receives electrical inputs that drive the optical transmitter. A communication channel carries the optical signals into an optical receiver that converts them back to electrical signals. The optical transmitter has an optical sources whose output is modulated by the incoming electrical signals. The optical receiver is photodetector whose output is demodulated to obtain the original electrical signal. The communication channel contains optical fibers that carry the light pulses. The intensity of light drops as it progresses along the fiber. Hence, optical amplifiers are used to boost up the light intensity and eventually to regenerate the transmitted pulses.

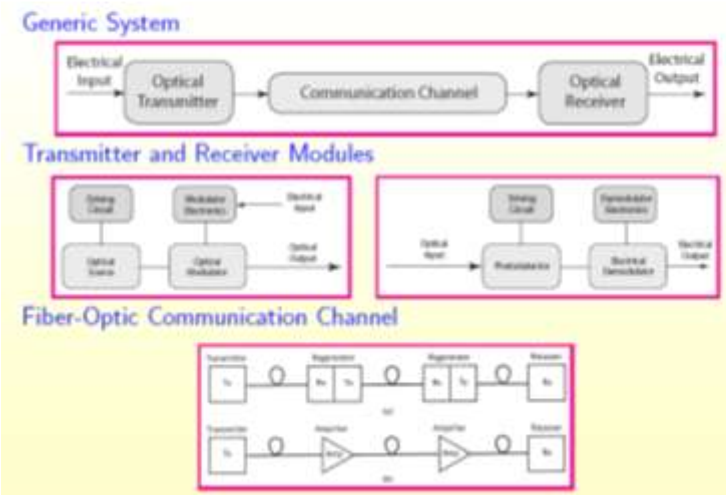
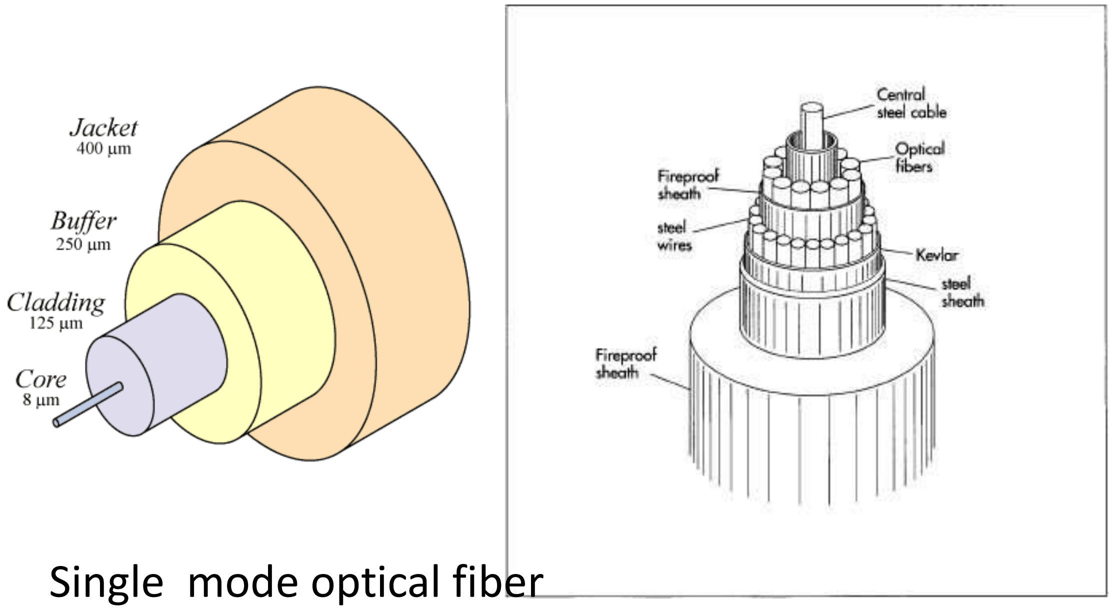


Figure 1. 14 Components of a light wave system

An optical fiber is basically a thin glass rod as shown in Figure 1.15. The single mode fiber has a cladding covered by a buffer material that is further covered by a fire-proof jacket. A multi core fiber contains many optical fibers. The structure is mechanically strengthened using steel core and sheath. Again, the overall structure is covered with a fire-proof jacket.



Single mode optical fiber

Multi core optical fiber

Figure 1. 15 Examples of fiber optic fibers

QUANTITIES, UNITS AND STANDARDS

Definitions

A quantity is a quantifiable or assignable property ascribed to phenomena, bodies, or substances. Examples are speed of a car and mass of an electron. A physical quantity is a quantity that can be used in the mathematical equations of science and technology. A unit is a particular physical quantity, defined and adopted by convention, with which other particular quantities of the same kind are compared to express their value. The value of a physical quantity is the quantitative expression of a particular physical quantity as the product of a number and a unit, the number being its numerical value. Thus, the numerical value of a particular physical quantity depends on the unit in which it is expressed. For example, the value of the height *h* of a light pole is *h* = 16 m. Here *h* is the physical quantity, its value expressed in the unit "meter," unit symbol m, is 16 m, and its numerical value when expressed in meters is 16.

Basic Units and Derived Units

In all conversations, the physical quantities are presented with their proper values compared to the standard, the units. The general unit of a physical quantity is defined as its dimension. A unit system can be developed by choosing, for each basic dimension of the system, a specific unit. For example, the internationally established (SI) units are the meter for length, the kilogram for mass, and the second for time, abbreviated as the mks system of units. Such a unit is called a basic unit. The corresponding physical quantity is called a basic quantity. All units that are not basic are called derived units. In the mks system the derived units for force and energy are a convenient size in an engineering sense, and all the practical units fit in as the natural units to form a comprehensive unit system.

If we define the dimensions of length, mass, and time as [L], [M], and [T], respectively, then physical quantities may be expressed as $[L]^x[M]^y[T]^z$. For instance, the dimension of acceleration is $[L][T]^{-2}$ and that of force is $[L][M][T]^{-2}$. In the mks system of units, the systematic unit of acceleration is therefore 1 m/s^2 and that of force is 1 kgm/s^2.

Systems of units in which the mass is taken as a basic unit are called absolute systems of units, whereas those in which the force rather than the mass is taken as a basic unit are called gravitational systems of units. The metric engineering system of units is a gravitational system of units and is based on the meter, kilogram-force, and second as basic units.

Standards

The international system of units (SI) is the internationally agreed on system of units for expressing the values of physical quantities. In this system four basic units are added to the customary three basic units (meter, kilogram, second) of the mks absolute system of units. The four added basic units are ampere as the electric current, the Kelvin as the unit of thermodynamic temperature, the candela as the unit of luminous intensity, and the mole as the unit of amount of substance. Thus in SI units the meter, kilogram, second, ampere, Kelvin, candela, and mole constitute the seven basic units. There are two auxiliary units in the SI units: the radian, which is the unit of a plane angle, and the steradian, which is the unit of a solid angle.

Many countries established standardization institutions and standard laboratories where they keep the standard units that are calibrated against the world standards and kept as national standards. All other standards in the country are calibrated against these national standards and used as secondary standards.

In this courses we will use notations in accordance with the current International Standards. Units for engineering quantities are printed in upright roman characters, with a space between the numerical value and the unit, but no space between the decimal prefix and the unit, e.g. 275 kV. Compound units have a space, dot or / between the unit elements as appropriate, e.g. 1.5 N m, 300 m/s , or 9.81 m.s-2. Variable symbols are printed in italic typeface, e.g. V. For ac quantities, the instantaneous value is printed in lower case italic, peak value in lower case italic with caret (^), and rms value in upper case, e.g. i, î, I. Symbols for the important electrical quantities with their units are given in Table 1.

Table 1 Symbols for standard quantities and units

Symbol	Quantity	Unit	Unit symbol
A	geometric area	square meter	m^2
B	magnetic flux density	tesla	T
C	Capacitance	farad	F
E	electric field strength	volt per meter	V/m
F	mechanical force	Newton	N
Fm	magnetomotive force (mmf)	Ampere	A or A.t
G	conductance	Siemens	S
H	magnetic field strength	ampere per metre	A/m
I	electric current	ampere	A
J	electric current density	ampere per square metre	A/m^2
J	moment of inertia	kilogram metre squared	$kg.m^2$
L	self-inductance	henry	H
M	mutual inductance	henry	H
N	number of turns		

Symbol	Quantity	Unit	Unit symbol
P	active or real power	watt	W
Q	electric charge	coulomb	C
Q	reactive power	volt ampere reactive	VAR
R	electrical resistance	ohm	Ω
Rm	Reluctance	ampere per weber	A/Wb
S	apparent power	volt ampere	V.A
T	mechanical torque	newton meter	N.m
V	electric potential or voltage	volt	V
W	energy or work	joule	J
X	Reactance	ohm	Ω
Y	Admittance	Siemens	S
Z	Impedance	ohm	Ω
f	Frequency	hertz	Hz
i or j	square root of -1		
l	Length	Meter	m
m	Mass	Kilogram	kg
n	rotational speed	revolution per minute	rpm
p	Number of machine poles		
t	Time	Second	s
v	linear velocity	meter per second	m/s
ε	Permittivity	farad per meter	F/m
η	Efficiency		
θ	Angle	radian or degree	rad or °
λ	power factor		
Λ	Permeance	weber per ampere	Wb/A
μ	Permeability	henry per meter	H/m
ρ	Resistivity	ohm meter	Ω.m
σ	Conductivity	siemens per meter	S/m
ϕ	phase angle	radian	rad
Φ	magnetic flux	weber	Wb
Ψ	magnetic flux linkage	weber or weber-turn	Wb or Wb.t
ω	angular velocity or angular frequency	radian per second	rad/s

Prefixes

The SI prefixes used to form decimal multiples and submultiples of SI units are given in Table 2. The kilogram is the only SI unit with a prefix as part of its name and symbol. Because multiple prefixes may not be used, in the case of the kilogram the prefix names of Table 2 are used with the unit name "gram" and the prefix symbols are used with the unit symbol "g." With this exception, any SI prefix may be used with any SI unit, including the degree Celsius and its symbol °C.

Because the SI prefixes strictly represent powers of 10, they should not be used to represent powers of 2. Thus, one kilobit, or 1 kbit, is 1000 bit and **not** 2^{10} bit = 1024 bit. To alleviate this ambiguity, prefixes for binary multiples have been adopted by the International Electrotechnical Commission (IEC) for use in information technology. This is beyond the context of this textbook. Listing and further descriptions of basic and derived units and standards are given in Appendix-A.

Table 2 Standard prefixes for the SI units of measure

Multiples			Fractions		
Name	Symbol	Factor	Name	Symbol	Factor
		10^0			10^0
deca	Da	10^1	deci	d	10^{-1}
hecto	H	10^2	centi	c	10^{-2}
kilo	K	10^3	milli	m	10^{-3}
mega	M	10^6	micro	µ	10^{-6}
giga	G	10^9	nano	n	10^{-9}
tera	T	10^{12}	pico	p	10^{-12}
peta	P	10^{15}	femto	f	10^{-15}
exa	E	10^{18}	atto	a	10^{-18}
zetta	Z	10^{21}	zepto	z	10^{-21}
yotta	Y	10^{24}	yocto	y	10^{-24}

PROBLEMS

Review Questions

1. What is engineering and who is engineer?
2. What are the similarities and differences between electrical and electronics engineering?
3. Briefly describe the fields of activities of electronics engineering.

4. Define the computer science and computer engineering.
5. What are the similarities and differences between computer science and computer engineering?
6. State the responsibilities of and career opportunities for graduates of your specialization.
7. Interpret the logo of mechatronics that was illustrated in Figure 1.2.
8. Discuss the importance of electronics in design of mechanical systems.
9. List important electrical/electronic components in your car. What do you understand from the term "brake by a wire"?
10. Define avionics and list critical applications of electronics and communications engineering related to the operation and safety in airplanes.
11. Discuss the applications of electronics and communications engineering in the geriatric medicine (care for elderly).
12. State a few examples in which the electrical/electronic engineering contribute positively to the welfare of disabled people.
13. Compare the cognitive radio communication to conventional radio and discuss its advantages.
14. Make a web search and identify the salient features of optical communication.
15. State seven basic internationally recognized (SI) units and specify quantities that they identify.
16. Please circle the best choice in the following questions:

1. 1 pF (pico farad) is			
a. 10^{-3} F	b. 10^{-6} μF	c. 10^{-9} A	d. 10 V/s
2. 1 Farad is			
a. 1 Coulomb/V	b. 1 A*s	c. 1 Coulomb	d. 1 ohm/s
3. 1 Coulomb is			
a. 1 V/s	b. 1 Wb*s	c. 1 F	d. 1 A*s
4. 1 Hertz (Hz) is			
a. 1 radian	b. 1 radian/(2π)	c. 1 cycles/s	d. 1 radian/s (rps)
5. 1 Watt is			
a. 1 A*s	b. 1 Joule/s	c. 1 A/s	d. 1 Joule*s
6. 1 Tesla is			
a. 1 Weber/m^2	b. 1 Coulomb*s	c. 1 Volt/m	d. 1 V*A
7. 1 ohm is			
a. 1 V*A	b. 1 Joule/s	c. 1 V/A	d. 1 Farad/s
8. The velocity is			
a. Distance*s	b. Integral of acceleration	c. Distance/s^2	d. Force/area
9. 1 Siemens (mho) is			
a. 1 Ohm*m	b. 1 Farad/s	c. 1/ohm	d. 1 A/s
10. 1 Newton is			
a. 1 kg*m	b. 1 Watt*s	c. 1 Ampere*s	d. 1 Pascal*m^2

BIBLIOGRAPHY

Further Reading

Useful Websites

FUNDAMENTAL ELECTRICAL ENGINEERING COMPONENTS

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Identify the subatomic particles and their contributions to the electrical activities within an atom.
2. Define in precise terms electricity, magnetism, electrical charge, electrical field, magnetic field, electrical conduction and electromagnetism, and express the relationship between them.
3. Describe various forms of generation, transmission and distribution of electrical energy.
4. Define in precise terms conductors, semiconductors and insulators.
5. Classify wire conductors, cables and transmission lines, recognize their international standards.
6. Explain properties of wire conductors in terms of ampacity, resistance and effects of temperature and frequency.
7. Define electrical resistors and their functionalities.
8. Classify fixed resistors according to their compositions and areas of applications.
9. Describe adjustable resistors, their use and limitations.
10. Identify resistors according to their color code marking and tell the preferred values.
11. Determine the power rating requirements of resistors and choose the proper ones for a given applications.
12. Explain the behavior of resistors at high frequencies and be familiar with noise in resistors.
13. Be familiar with the reasons for the failures of resistors and failure modes.
14. Define the capacitance and capacitors, their use in electrical circuits.
15. Describe the non-ideal behavior of capacitors such as the breakdown voltage, ripple current and instability.
16. Identify various capacitor types that are used in practice using capacitor markings.
17. Select the proper capacitor for a given application.
18. Discuss the principles and applications of capacitive sensing.
19. Identify hazards related to capacitors and required safety measures.
20. Describe principles and applications of supercapacitors (electric double-layer capacitors).
21. Define the inductance and inductors, their use in electrical circuits.
22. Discuss the inductor types and non-ideal behavior of inductors with their effects in performance of inductive circuits.
23. Discuss the transformer as a circuit elements and effects of its the non-ideal behavior.

ENERGY SOURCES

The Atom and Subatomic Particles

The earth is made of elements each of which has distinct characteristics. The smallest part of an elements that carries its characteristics is called the atom. The atom is also made up of subatomic particles. Among them we have protons that are located in the center (nucleus) of the atom and they are loaded with positive electrical charge. We have negatively loaded particles that spin around their own axes and also travel around selected orbits around the nucleus as depicted in Figure 2.1. The magnitude of the charge of an electron is the same as that of the proton. The number of electrons are equivalent to the number of protons for a given atom and eventually there is a charge neutrality. Each orbit for the electrons has a specific energy level. The electrons are loosely connected to the atom and they can jump into a higher energy orbit if they receive a suitable external energy. However, they don't stay in the new orbit and they return back to their original orbit by ejecting the additional energy as an electromagnetic wave.

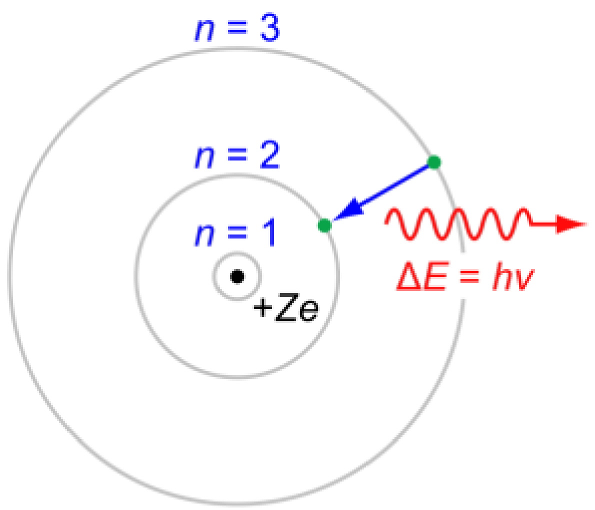


Figure 2. 1 Atom and its charged particles

Electrons moving around the nucleus establish a cloud of negative charges as illustrated in Figure 2.2 for the helium atom depicting the nucleus (pink) and the electron cloud distribution (black). The nucleus (upper right) in helium-4 is in reality spherically symmetric and closely resembles the electron cloud, although for more complicated nuclei this is not always the case. The black bar is one angstrom, equal to 10^{-10} m or 100,000 fm.

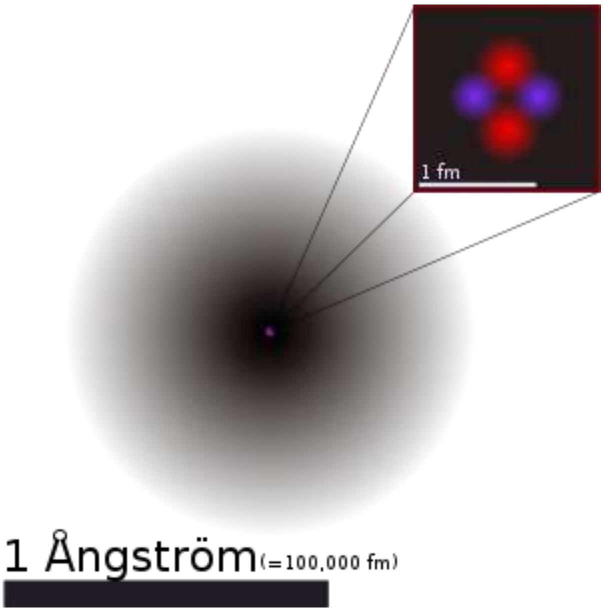


Figure 2. 2 Electron cloud around the nucleus of the helium atom

Electricity

In general usage, the word "electricity" adequately refers to a number of physical effects. In scientific usage, however, the term is vague, and these related, but distinct, concepts are better identified by more precise terms.

Electric Charge

The electric charge is a property of some subatomic particles, which determines their electromagnetic interactions. Electrically charged matter is influenced by, and produces, electromagnetic fields. The charge on electrons and protons is opposite in sign as mentioned above, hence an amount of charge may be expressed as being either negative or positive. By convention, the charge carried by electrons is deemed negative, and that by protons positive. The amount of charge is usually given the symbol *Q* and expressed in coulombs; each electron carries the same charge of approximately -1.6022×10^{-19} coulomb. The proton has a charge that is equal and opposite, and thus $+1.6022\times10^{-19}$ coulomb.

Electric Field

The electric field is an influence produced by an electric charge on other charges in its vicinity. An electric field is created by a charged body in the space that surrounds it, and results in a force exerted on any other charges placed within the field. Figure 2.3 shows the electrical field lines for a positive electrical charge.

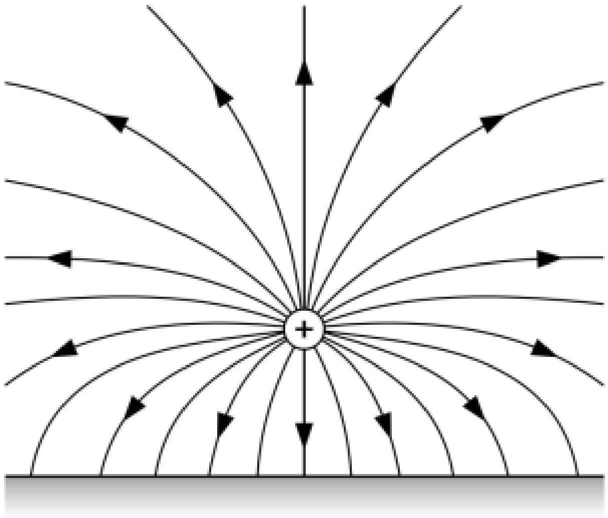


Figure 2. 3 Field lines emanating from a positive charge above a plane conductor

Electric Potential

The electric potential is the capacity of an electric field to do work on an electric charge. The concept of electric potential is closely linked to that of the electric field. A small charge placed within an electric field experiences a force, and to have brought that charge to that point against the force requires work. The electric potential at any point is defined as the energy required to bring a unit test charge from an infinite distance slowly to that point. It is usually measured in volts, and one volt is the potential for which one joule of work must be expended to bring a charge of one coulomb from infinity.

Electrical Conduction

The electrical conduction is the movement of electrically charged particles through a transmission medium (electrical conductor). Its nature varies with that of the charged particles and the material through which they are travelling. This charge transport may reflect a potential difference due to an electric field, or a concentration gradient in carrier density. The latter reflects diffusion of the charge carriers. The physical parameters governing this transport depend upon the material. Examples of electric currents include metallic conduction, where electrons flow through a conductor such as metal, and electrolysis, where ions (charged atoms) flow through liquids.

The movement of electric charge is known as an **electric current**, the intensity of which is usually measured in amperes. Current can consist of any moving charged particles; most commonly these are electrons, but any charge in motion constitutes a current. By historical convention, a **positive current** is defined as having the same direction of flow as any positive charge it contains, or to flow from the most positive part of a circuit to the most negative part. Current defined in this manner is called conventional current. The motion of negatively charged electrons around an electric circuit, one of the most familiar forms of current, is thus deemed positive in the *opposite* direction to that of the electrons. However, depending on the conditions, an electric current can consist of a flow of charged particles in either direction, or even in both directions at once. The positive-to-negative convention is widely used to simplify this situation.

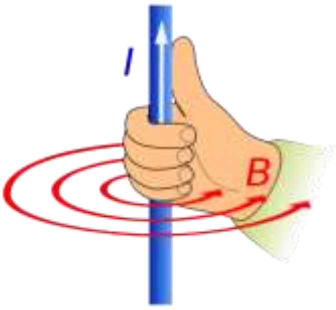


Figure 2. 4 The magnetic field around a current carrying conductor

Magnetic Field

A **magnetic field** is a field of force produced by moving electric charges, by electric fields that vary in time, and by the 'intrinsic' magnetic field of elementary particles associated with the spin of the particle. The magnetic field strength **B** is a vector quantity that has both magnitude and direction. A current flowing in a conductor produces a rotational magnetic field as depicted in Figure 2.4. The direction is identified with the right-hand grip rule. The unit of **B** is Tesla or Gauss (1 Tesla = 10,000 Gauss)

The current in a solenoid coil generates a translational magnetic field through the coil as shown in Figure 2.5.

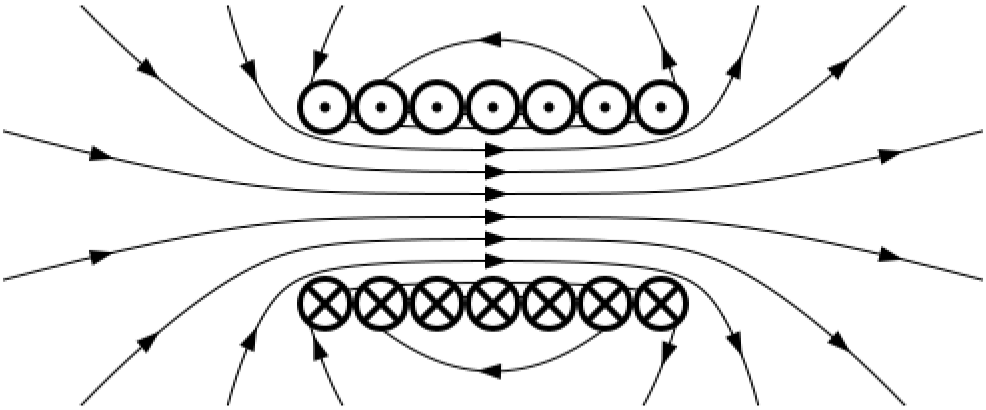


Figure 2. 5 The magnetic field lines for a solenoid coil

A current carrying conductor in an external magnetic field experiences a mechanical force due to interaction of the field lines as illustrated in Figure 2.6.

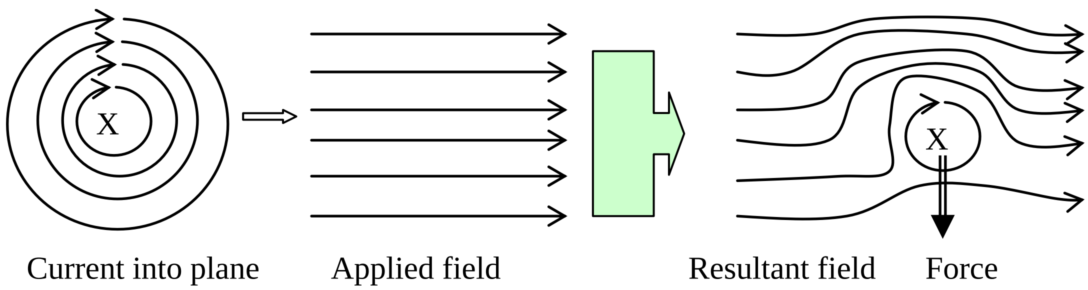


Figure 2. 6 A current carrying conductor in an external magnetic field

A current bearing coil inserted in an external magnetic field experiences a torque as illustrated in Figure 2.7. This is the fundamental principle of electric motors. Equivalently, a loop of conductor moving in an external magnetic field will have an electrical current induced in it. This is the principle of generators.

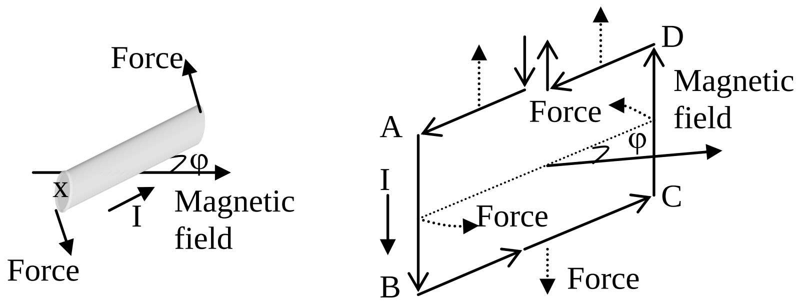


Figure 2. 7 Torque experienced by a current carrying coil as it is exposed to an external magnetic field

Electromagnetism

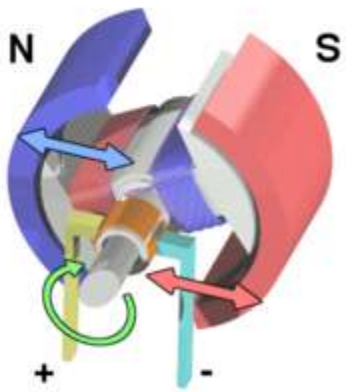


Figure 2. 8 Principle of an electric motor

Electromagnetism is a fundamental interaction between the magnetic field and the presence and motion of an electric charge. The relationship between the magnetic and electric fields, and the currents and charges that create them, is described by the set of Maxwell's equations that are covered in EE 302 – Electromagnetic Fields. The electric motor shown in Figure 2.8 exploits an important effect of electromagnetism: a current through a magnetic field experiences a force at right angles to both the field and current.

Electrostatics

The study of electric fields created by stationary charges is called **electrostatics**. The principles of electrostatics are important when designing items of high-voltage equipment. There is a finite limit to the electric field strength that may be withstood by any medium. Beyond this point, electrical breakdown occurs and an **electric arc** causes flashover between the charged parts as illustrated in Figure 2.9. Air, for example, tends to arc across small gaps at electric field strengths which exceed 30 kV per centimeter. Over larger gaps, its breakdown strength is weaker, perhaps 1 kV per centimeter. The most visible natural occurrence of this is lightning, caused when charge becomes separated in the clouds by rising columns of air, and raises the electric field in the air to greater than it can withstand. The voltage of a large lightning cloud may be as high as 100 MV and have discharge energies as great as 250 kWh.

Figure 2. 9 An electric arc (from http://en.wikipedia.org/wiki/File:Lichtbogen_3000_Volt.jpg)

Generation of Electrical Energy

Electrical energy is not generally referred to as electrical energy for the layperson, and is most commonly known as electricity. Electrical energy is the scientific form of electricity, and refers to the flow of power or the flow of charges along a conductor to create energy. Electrical energy doesn't exist in nature in large quantities to the a work. It is known to be a secondary source of energy, which means that we obtain electrical energy through the conversion of other forms of energy. These other forms of energy are known as the primary sources of energy and can be used from coal, nuclear energy, natural gas, or oil as illustrated in Figure 2.10. These are called the non-renewable sources of energy.

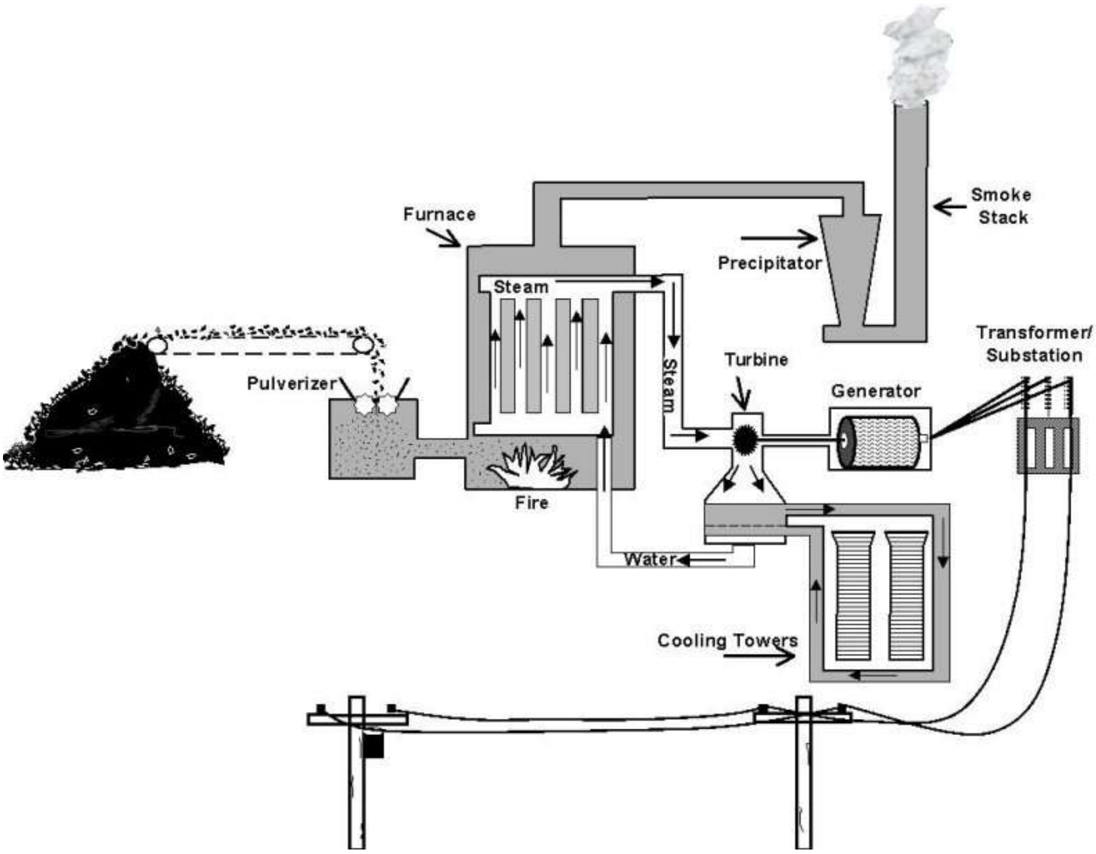


Figure 2. 10 Generation of electrical energy from fossil fuels

Electrical energy is a standard part of nature, and today it is our most widely used form of energy. The primary sources from which we obtain electrical energy can be renewable forms of energy as well. Electrical energy however is neither non-renewable or renewable. Many towns and cities were developed beside waterfalls which are known to be primary sources of mechanical energy. Wheels would be built in the waterfalls and the falls would turn the wheels in order to create energy that fueled the cities and towns. Figure 2.11 illustrates four different forms of obtaining electrical energy from renewable sources. The upper left corners shows a wind farm and the upper right corner shows the solar cells for generating electricity. There is a hydroelectric power station at the lower left and nuclear power station at the lower right.

Figure 2. 11 Generation of electrical energy from renewable sources

Transmission and Distribution of Electrical Energy

The beauty of electrical energy is its cleanliness and efficiency in use as well as the speed of transmission. While the particles themselves can move quite slowly, sometimes with an average drift velocity only fractions of a millimeter per second, the electric field that drives them itself propagates at close to the speed of light (c = 300,000 km/s), enabling electrical signals to pass rapidly along wires. With the discovery of Alternating Current (AC) energy, electrical energy could be transmitted over much larger distances. With this discovery, electrical energy could then be used to light homes and to power machines that would be more effective at heating homes as well.

In order for electrical energy to transfer at all, it must have a conductor or a circuit that will enable the transfer of the energy. Electrical energy will occur when electric charges are moving or changing position from one element or object to another. Storing the electrical energy at large quantities is also not possible. Hence, the energy must be used as it is produced. It is frequently stored in small quantities today as batteries or energy cells.

Figure 2.12 illustrates the generation, transmission and utilization of electrical energy. It is important to understand that electrical energy is not a kind of energy in and of itself, but it is rather a form of transferring energy from one object or element to another. The energy that is being transferred is the electrical energy.

Electrical energy is produced from fossil fuels or renewable sources in the generating plant. The energy in joules is time integral of the electrical power in watts. The instantaneous electrical power is the product of the voltage and current. The conductors that are used in transmitting the electrical energy have certain resistances that dissipate a portion of the energy. Hence, it is preferable to use higher voltages to transmit the energy in order to reduce the transmission losses. The generator produces 14 kV that is increased to 230 kV for the transmission. This high voltage is reduced to 72 kV or 130 kV at transformer switching stations before the industrial installations. It is further reduced to 25 kV for commercial, business and residential districts. Finally, it is reduced to 127/220 V for domestic and business customers. The voltage levels used may vary but the voltage supplied to the customer is fixed. The frequency of the voltage is 60 Hz in Saudi Arabia. There is new voltage standard of 230/400 V that will be enforced in all over the Kingdom in the next 10 years.

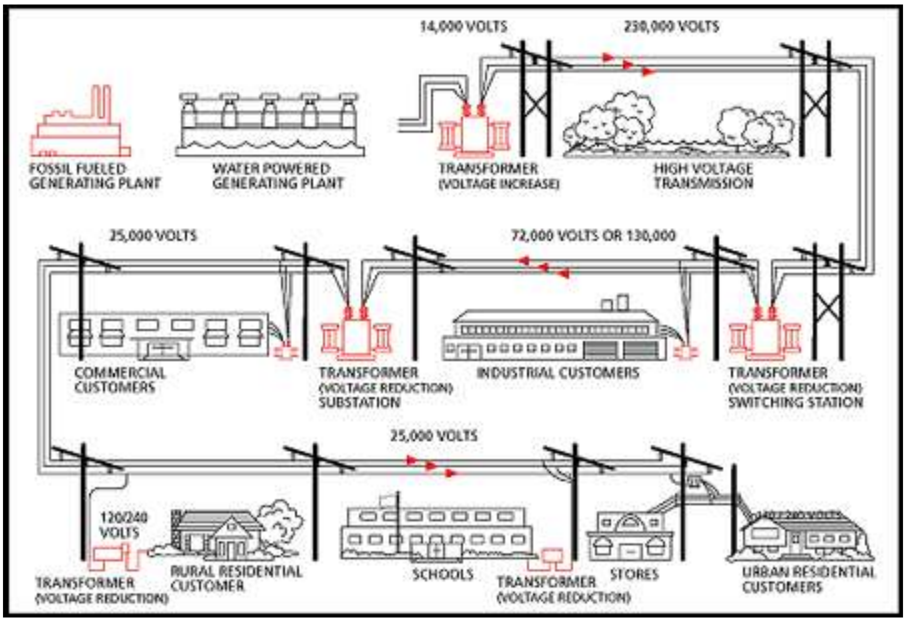


Figure 2. 12 A symbolic illustration of generation, transmission and distribution of electrical energy

CONDUCTORS AND INSULATORS

Definitions

Conductors

An electrical conductor is any material through which electrical current flows easily. Most metals are good electrical conductors, with silver the best and copper second. Their atomic structure allows free movement of the outer most orbital electrons. Copper wire is generally used for practical conductors because it costs much less than the silver. The purpose of using a conductor is to carry electric current with minimal opposition.

Semiconductors

Carbon is considered a semiconductor, conducting less than metal conductors but more than insulators. In the same group are germanium and silicon, which are commonly used for transistors and other semiconductor components. The degree of *doping* in semiconductors makes a large difference in conductivity. To a point, more doping leads to higher conductivity. Practically all transistors are made of silicon.

Superconductors

Superconductivity is a property of certain materials for which the electrical resistance of becomes exactly zero below a characteristic temperature. The electrical resistivity of a metallic conductor decreases gradually as the temperature is lowered. However, in ordinary conductors such as copper and silver, this decrease is limited by impurities and other defects. Even near absolute zero (0 K = -273 °C), a real sample of copper shows some resistance. Despite these imperfections, in a superconductor the resistance drops abruptly to zero when the material is cooled below its critical temperature. An electric current flowing in a loop of superconducting wire can persist indefinitely with no power source.

In 1986, it was discovered that some ceramic materials have critical temperatures above 90 K (−183 °C). These high-temperature superconductors renewed interest in the topic because of the prospects for improvement and potential room-temperature superconductivity. From a practical perspective, even 90 kelvins is relatively easy to reach with readily available liquid nitrogen (which has a boiling point of 77 kelvins), resulting in more experiments and applications.

Insulators

An insulator is any material that resists or prevents the flow of electric charge, such as electrons. The resistance of an insulator is very high, typically hundreds of mega ohms or more. An insulator provides the equivalent of an open circuit with practically infinite resistance and almost zero current. It is from a material with atoms in which the electrons tend to stay in their own orbits and hence cannot conduct electricity easily. Insulators can be useful when it is necessary to prevent the current flow. In addition, for applications requiring the storage of electric charge, as in capacitors, a dielectric material must be used because a good conductor cannot store any charge. An insulating material, such as glass, plastic, rubber, paper, air, or mica, is also called *dielectric*, meaning it can store electric charge.

Atomic structures that effect the properties of conductors and insulators are illustrated in Figure 2.13.

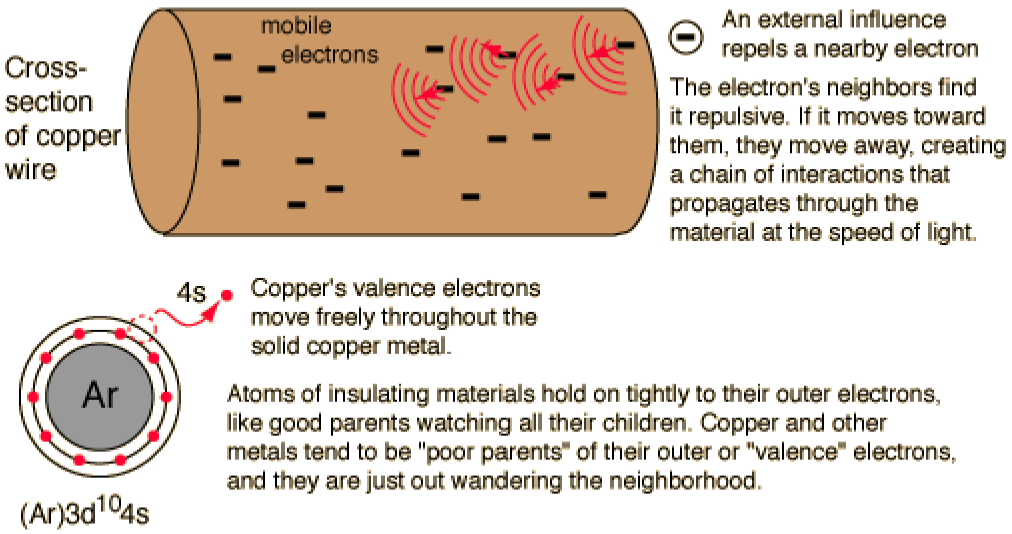


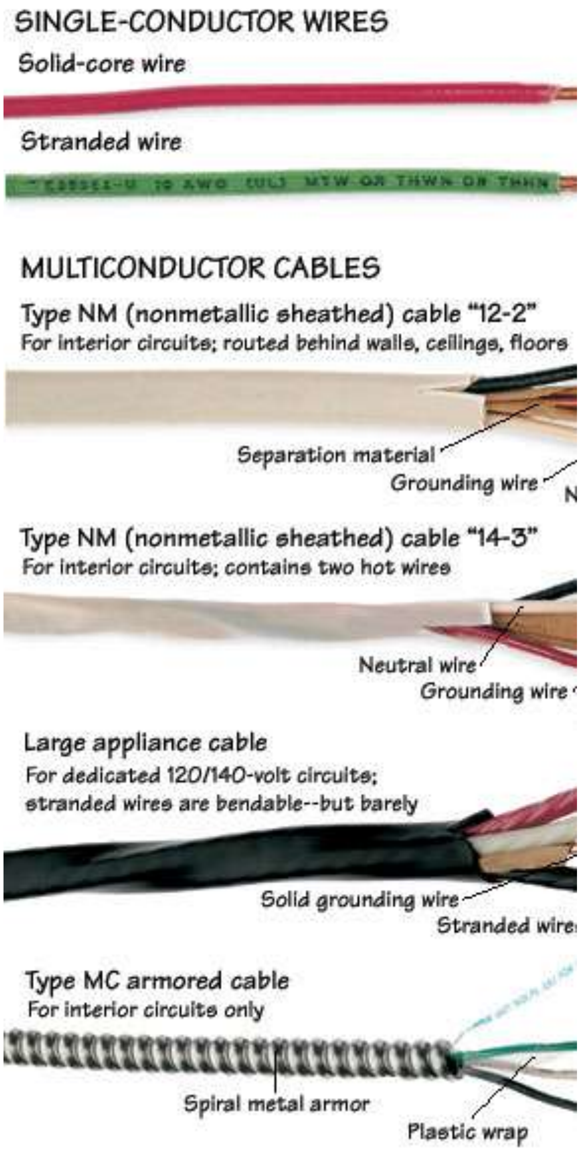
Figure 2. 13 Atomic structure of conducting and insulating materials

Wire Conductors

Types of Wire Conductors

Most wire conductors are copper due to its low cost, although aluminum and silver are also used sometimes. The copper may be tinned with a thin coating of solder, which gives a silvery appearance. Tinned wire is easier to solder for connections. The wire can be solid or stranded. Solid wire is made up of only one conductor. If it bent or flexed repeatedly, solid wire may break. Therefore solid wire is used in places where bending and flexing is not encountered. House wiring is a good example of the use of solid wire. Stranded wire is made up of several individual strands put together in a braid. Some uses for stranded wire include telephone cords, extension cords and speaker wire, to name a few. Figures 14 and 15 show wire conductors for variety of applications.

Figure 2. 14 Wires and cables used for various applications



Stranded wire is flexible, easier to handle, and less likely to develop an open break. Sizes for stranded wire are equivalent to sum of areas for the individual strands. For instance, two strands of No. 30 wire corresponds to solid No. 27 wire. Very thin wire, such as No. 30, often has an insulating coating of enamel or shellac. It may look like copper, but the coating must be scrapped off the ends to make a good connection. This type of wire is used for small coils. Heavier wires generally are in an insulating sleeve, which may be rubber or one of many plastic materials. General purpose wire for connecting electronic components is generally plastic coated hookup wire of No. 20 gage. Hookup wire that is bare should be enclosed in an insulating sleeve called *spaghetti*.

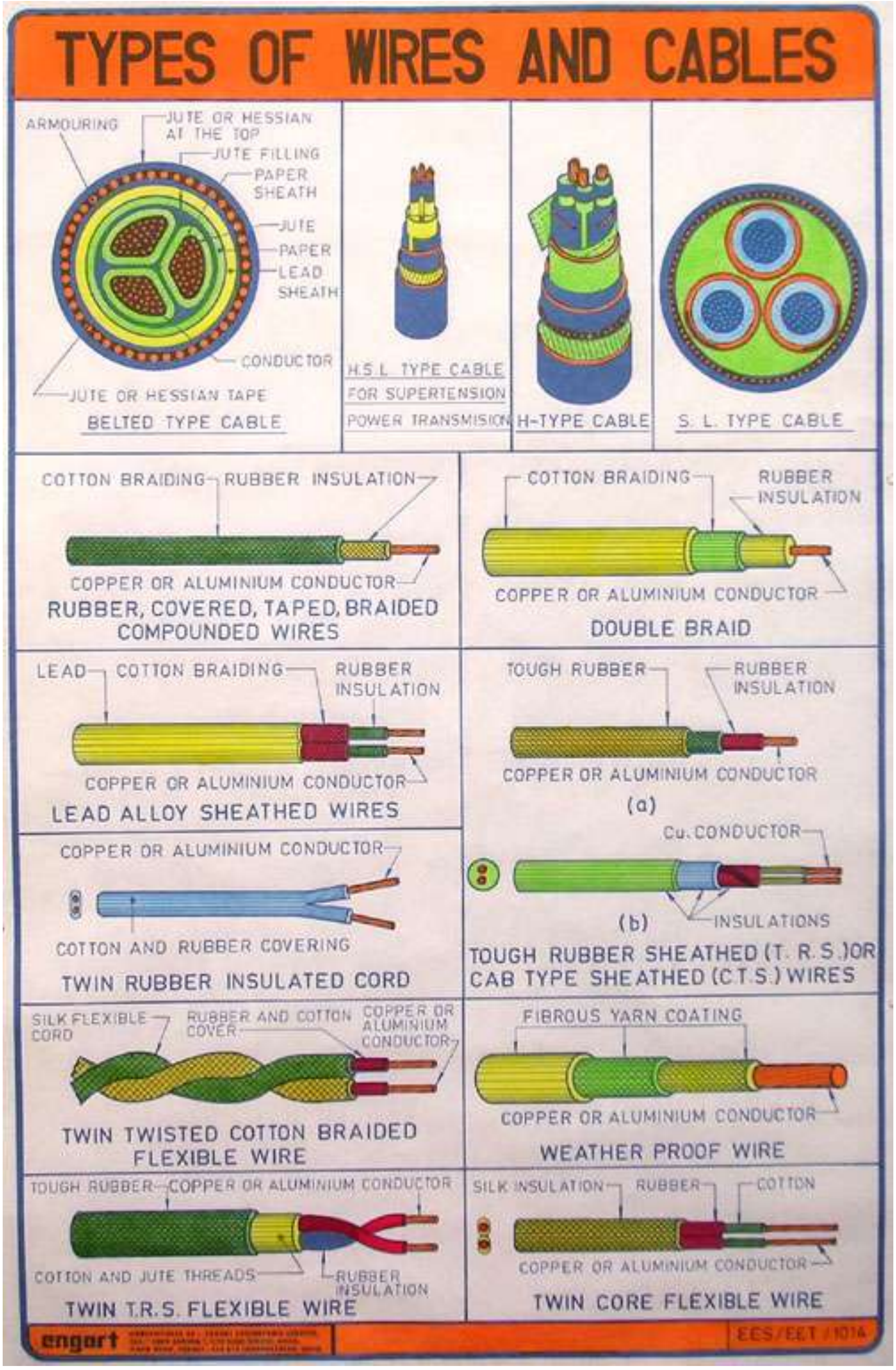


Figure 2. 15 Types of wires and cables

Twisted pairs are used for small signal applications in electronics. They may or may not be shielded as illustrated in Figure 2.16. They are good in preventing magnetic field pickups. The shielded ones are used especially in low noise applications.

Figure 2. 16 Shielded and unshielded twisted pairs

The braided conductor shown in Figure 2.17 is used for very low resistance. It is wide for low *R* and thin for flexibility, and braiding provides many strands. A common application is a grounding connection, which must have very low *R*.

Figure 2. 17 Braided conductors

Wire Cable

Two or more conductors in a common covering form a cable. Each wire is insulated from the others. Cables often consist of two, three, ten, or many more pairs of conductors, usually color coded to help to identify the conductor at both ends of a cable.

Transmission Lines

A transmission line is a cable setup used to carry electrical signals in various applications. Constant spacing between two conductors through the entire length provides a transmission line. Common examples are the coaxial cable, the twin lead and ribbon cable. The coaxial cable with outside diameter of 1/4 inch is generally used for the signals in cable television. In construction, there is an inner solid wire, insulated from metallic braid that serves as the other conductor. The entire assembly is covered by an outside plastic jacket. In operation, the inner conductor has the desired signal voltage with respect to ground, and metallic braid is connected to ground to shield the inner conductor against interference. Coaxial cable, therefore, is a shielded type of transmission line. Single core and dual core coaxial cables are shown in Figure 2.18.

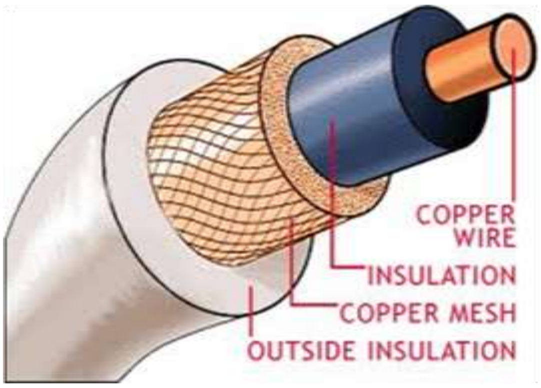


Figure 2. 18 Single and dual core caoxial cables

With twin-lead wire, two conductors are embedded in plastic to provide constant spacing (Figure 2.19). This type of line is commonly used in television for connecting the antenna to the receiver. In this application, the spacing is 5/8 inch between wires of No. 20 gage size, approximately. This line is not shielded.

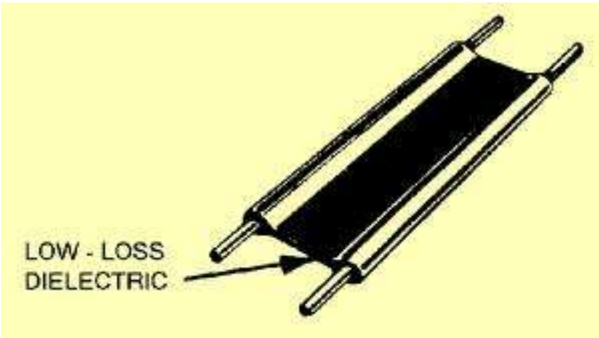


Figure 2. 19 Twin-lead TV antenna wire

The ribbon cable in Figure 2.20, has multiple conductors but not in pairs. This cable is used for multiple connections to a computer and associated equipment.

Figure 2. 20 The ribbon cable for connecting computer peripherals

Standard Wire Gage Sizes

Table 2.1 lists the standard wire sizes in the system knows as the American Wire Gage (AWG) expressed in metric system. The gage numbers specify the size of a round wire in terms of its diameter and cross-sectional area. Note the following three points:

As the gage number increase from 1 to 40, the diameter and the circular area decrease. Higher gage numbers indicate thinner wire sizes.

Table 2. 1 American Wire Gage (AWG) table in metric

A.W. G.	Diameter	Area	Weight		Length		Resistance	
B. & S.	Mm.	Sq. Mm.	Kg. per M.	Kg. per Ohm	M. per Kg.	M. per Ohm	Ohms per Kg.	Ohms per M.
0000	11.7	107.2	.953	5940	1.05	6,230	.000168	.000161
000	10.4	85.0	.756	3730	1.32	4,940	.000268	.000202
00	9.27	67.4	.599	2350	1.67	3,920	.000426	.000255
0	8.25	53.5	.475	1480	2.10	3,110	.000677	.000322
1	7.35	42.4	.377	929	2.65	2,460	.00108	.000406
2	6.54	33.6	.299	584	3.35	1,950	.00171	.000512
3	5.83	26.7	.237	367	4.22	1,550	.00272	.000645
4	5.19	21.2	.188	231	5.32	1,230	.00433	.000814
5	4.62	16.8	.149	145	6.71	975	.00688	.00103
6	4.11	13.3	.118	91.4	8.46	773	.0109	.00129
7	3.67	10.6	.0938	57.5	10.7	613	.0174	.00163
8	3.26	8.37	.0744	36.2	13.5	486	.0277	.00206
9	2.91	6.63	.0590	22.7	17.0	386	.0440	.00259
10	2.59	5.26	.0468	14.3	21.4	306	.0699	.00327
11	2.31	4.17	.0371	8.99	27.0	242	.111	.00413
12	2.05	3.31	.0294	5.66	34.0	192	.177	.00520
13	1.83	2.62	.0234	3.56	42.9	153	.281	.00656
14	1.63	2.08	.0185	2.24	54.1	121	.447	.00827
15	1.45	1.65	.0147	1.41	68.2	95.9	.711	.0104
16	1.29	1.31	.0116	.885	86.0	76.0	1.13	.0132
17	1.15	1.04	.00922	.556	108	60 3	1.80	.0166
18	1.02	.823	.00732	.350	136	47.8	2.86	.0209
19	.912	.653	.00580	.220	172	37.9	4.54	.0264
20	.812	.518	.00460	.138	217	30.1	7.23	.0333
21	.723	.410	.00365	.0871	274	23.9	11.5	.0419
22	.644	.326	.00289	.0548	346	18.9	18.3	.0529
23	.573	.258	.00229	.0344	436	15 0	29.1	.0667
24	.511	.205	.00182	.0217	550	11.9	46.2	.0841
25	.455	.162	.00144	.0136	693	9.43	73.4	.106
26	.405	.129	.00114	.00856	874	7.48	117	.134
27	.361	.102	.000908	.00538	1,100	5.93	186	.169
28	.321	.081	.000720	.00339	1,390	4.70	295	.213
29	.286	.0642	.000571	.00213	1,750	3.73	470	.268
30	.255	.0510	.000453	.00134	2,210	2.96	747	.338
31	.227	.0404	.000359	.000842	2,790	2.35	1,190	.426
32	.202	.0320	.000285	.000530	3,510	1.86	1,890	.537
33	.180	.0254	.000226	.000333	4,430	1.48	3,000	.678
34	.160	.0201	.000179	.000210	5,590	1.17	4,770	.855
35	.143	.0160	.000142	.000132	7,040	.928	7,590	1.08
36	.127	.0127	.000113	.0000829	8,880	.736	12,100	1.36
37	.113	.0101	.0000893	.0000521	11,200	.584	19,200	1.71
38	.101	.00797	.0000708	.0000327	14,100	.463	30,600	2.16
39	.0897	.00632	.0000562	.0000206	17,800	.367	48,500	2 73
40	.0799	.00501	.0000445	.0000130	22,500	.291	77,100	3 44

The circular area doubles for every three gage sizes. For example, No. 10 wire has approximately twice the area of No. 13 wire. The higher the gage number and thinner the wire, the greater the resistance of the wire for any given length. In typical applications, hookup wire for electronic circuits with current of the order of milliamperes in generally about No. 22 gage. For this size, 0.5 to 1 A is the maximum current the wire can carry without excessive heat. House wiring for circuits where the current is 5 to 15 A is usually No. 14 gage. Minimum sizes for house wiring are set by local electricity codes which are usually guided by the National Electrical Code published by the National Fire Protection Association.

Properties of Wire Conductors

Conductor Ampacity

The ampacity of a conductor, that is, the amount of current it can carry, is related to its electrical resistance: a lower-resistance conductor can carry more current. The resistance, in turn, is determined by the material the conductor is made from (as described above) and the conductor's size. For a given material, conductors with a larger cross-sectional area have less resistance than conductors with a smaller cross-sectional area. The economical factor plays an important role in selecting conductors in industrial (large) scale applications. Aluminum is lighter and cheaper that copper and it carries almost the same current as of copper for a given weight of the material. Hence, aluminum is mostly preferred in high voltage transmission lines as the electrical conductor.

For bare conductors, the ultimate limit is the point at which power lost to resistance causes the conductor to melt. Aside from fuses, most conductors in the real world are operated far below this limit, however. For example, household wiring is usually insulated with PVC insulation that is only rated to operate to about 60 °C, therefore, the current flowing in such wires must be limited so that it never heats the copper conductor above 60 °C, causing a risk of fire. Other, more expensive insulations such as Teflon or fiberglass may allow operation at much higher temperatures.

Wire Resistance

Electrical resistivity (also known as resistivity, specific electrical resistance, or volume resistivity) is a measure of how strongly a material opposes the flow of *electric current*. A low resistivity indicates a material that readily allows the movement of electric charge. The SI unit of electrical resistivity is the ohm meter [Ωm]. It is commonly represented by the Greek letter ρ (rho).

Electrical conductivity or specific conductance is the reciprocal quantity, and measures a material's ability to conduct an electric current. It is commonly represented by the Greek letter σ (sigma). Its SI unit is Siemens per meter ($S \cdot m^{-1}$). Many resistors and conductors have a uniform cross section with a uniform flow of electric current and are made of one material.

In a *hydraulic analogy*, increasing the diameter of a pipe reduces its resistance to flow, and increasing the length increases resistance to flow (and pressure drop for a given flow).

- A conductor such as a metal has high conductivity and a low resistivity.

- An insulator like glass has low conductivity and a high resistivity.

The conductivity of a semiconductor is generally intermediate, but varies widely under different conditions, such as exposure of the material to electric fields or specific frequencies of light, and, most important, with temperature and composition of the semiconductor material.

The resistance of a wire conductor is directly proportional to its length and inversely proportional to its cross sectional area. Hence, the longer a wire, the higher its resistance. More work must be done to make electron drift from one end to the other. However, the greater the diameter of the wire, the less the resistance, since there are more free electrons in the cross sectional area. As a formula,

$$R = \rho \frac{l}{A}$$

Where R (Ω) is the total resistance, l (m) the length, A (m^2) the cross-sectional area, and ρ (Ω.m) the specific resistance or *resistivity* of the conductor. The factor ρ then enables the resistance of different materials to be compared according to their nature without regard to different lengths or areas. Higher values of ρ means more resistance. Resistivity of metals that are most commonly used in electrical engineering applications is given in Table 2.2 for two temperatures.

Table 2. 2 Resistivity and temperature coefficient of metals of general interest in electrical engineering

Element	Symbol	ρ at 293 K (20° C)	ρ at 500 K (227° C)	Temperature coefficient α (/°C)
Graphite (carbon)	C	1375x10^{-8} Ωm		-0.0003
Aluminum	Al	26.5 nΩm	49.9 nΩm	0.0043
Vanadium	V	197 nΩm	348 nΩm	
Chromium	Cr	125 nΩm	201 nΩm	
Iron	Fe	96.1 nΩm	237 nΩm	0.0060
Nickel	Ni	69.3 nΩm	177 nΩm	0.0059
Copper	Cu	16.78 nΩm	30.9 nΩm	0.0040
Zinc	Zn	59 nΩm	108.2 nΩm	0.0038
Silver	Ag	15.87 nΩm	28.7 nΩm	0.0038
Tungsten	W	52.8 nΩm	103 nΩm	0.0044
Platinum	Pt	105 nΩm	183 nΩm	0.0038
Gold	Au	22.14 nΩm	39.7 nΩm	0.0037

Source: http://en.wikipedia.org/wiki/Electrical_resistivities_of_the_elements_(data_page)

Resistance Changes with Temperature

Resistance changes with temperature and how it does is indicated by a temperature coefficient with symbol alpha (α). Although α is not exactly constant, the resistance R_m at a temperature T_m is indicated by

$$R_m = R_0(1 + \alpha(T_m - T_0))$$

Where R_0 is the resistance at T_0.

All metals in their pure form, such as copper and tungsten, have positive temperature coefficients. In practical terms, a positive α indicates that heat increases R in a wire thereby the current I through the wire is reduced for a specified applied voltage. Carbon and all semiconductors, including germanium and silicon, have negative temperature coefficients. Some metal alloys, such as constantan and manganin have a value zero for α. The temperature coefficient for metals of general interest is given in the last column of Table 2.2.

Example: Let's take a look at an example circuit given in Figure 2.21 to see how temperature can affect wire resistance, and consequently circuit performance:

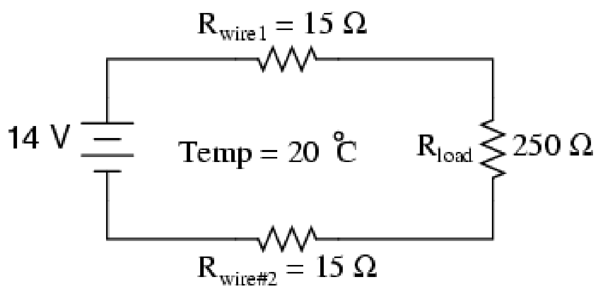


Figure 2. 21 Illustration of the effect of temperature on wire resistance

This circuit has a total wire resistance (wire 1 + wire 2) of 30 Ω at standard temperature. Setting up a table (Table 2.3) of voltage, current, and resistance values we get:

Table 2. 3 Voltage, current and resistances in Figure 2.21 at 20°C

	Wire-1	Wire-2	Load	Total	Unit
E	0.75	0.75	12.5	14	Volts
I	50 m	50 m	50 m	50 m	Amps
R	15	15	250	250	Ohms

At 20° Celsius, we get 12.5 volts across the load and a total of 1.5 volts (0.75 + 0.75) dropped across the wire resistance. If the temperature were to rise to 35° Celsius, we could easily determine the change of resistance for each piece of wire. Assuming the use of copper wire (α = 0.004041) we get:

$$R = R_{ref}\left[1 + \alpha\left(T - T_{ref}\right)\right]$$

Substituting the values; $R = (15\Omega)[1 + 0.004041(35°C - 20°C)]$ yields R = 15.909 Ω

Recalculating our circuit values, we see what changes this increase in temperature will bring the values displayed in Table 2.4:

Table 2. 4 Voltage, current and resistances in Figure 2.21 at 35°C

	Wire-1	Wire-2	Load	Total	Unit
E	0.79	0.79	12.42	14	Volts
I	49.677 m	49.677 m	49.677 m	49.677 m	Amps
R	15.909	15.909	250	281.82	Ohms

As you can see, voltage across the load went down (from 12.5 volts to 12.42 volts) and voltage drop across the wires went up (from 0.75 volts to 0.79 volts) as a result of the temperature increasing. Though the changes may seem small, they can be significant for power lines stretching miles between power plants and substations, substations and loads. In fact, power utility companies often have to take line resistance changes resulting from seasonal temperature variations into account when calculating allowable system loading.

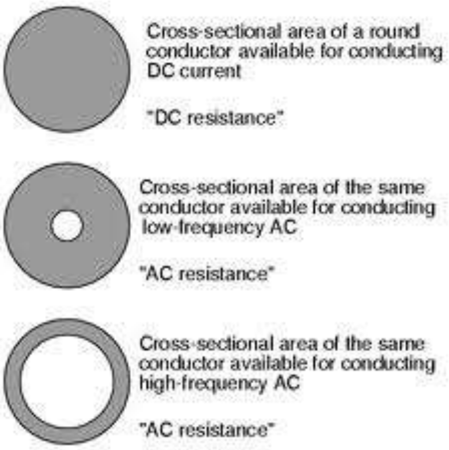


Figure 2. 22 An illustartion of the skin effect

Skin Effect

Skin effect is the tendency of an alternating electric current (AC) to distribute itself within a conductor with the current density being largest near the surface of the conductor, decreasing at greater depths. In other words, the electric current flows mainly at the "skin" of the conductor, at an average depth called the *skin depth*. The skin effect causes the effective resistance of the conductor to increase at higher frequencies where the skin depth is smaller, thus reducing the effective cross-section of the conductor. Figure 2.22 illustrates the distribution of electrical current through the cross-section of a current carrying conductor in DC, AC and high frequency applications.

Table 2. 5 Skin depth versus frequency

Frequency	Skin depth (μm)
60 Hz	8470
10 kHz	660
100 kHz	210
1 MHz	66
10 MHz	21
100 MHz	6.6

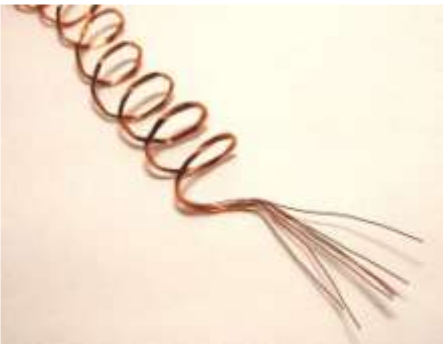


Figure 2. 23 The Litz wire

The skin effect is due to opposing eddy currents induced by the changing magnetic field resulting from the alternating current. At 60 Hz in copper, the skin depth is about 8.5 mm. At high frequencies the skin depth may be much smaller. Increased AC resistance due to the skin effect can be mitigated by using specially woven Litz wire (Figure 2.23). Because the interior of a large conductor carries so little of the current, tubular conductors such as pipe can be used to save weight and cost. In copper, the skin depth can be seen to fall according to the square root of frequency as given in Table 2.5.

RESISTORS

Definition and Use


The resistor is a two terminal electrical component that opposes the flow of either direct or alternating current, employed to protect, operate, or control the circuit. It is used in electrical circuits to maintain a constant relation between current flow and voltage. When a voltage *V* is applied across the terminals of a resistor, a current *I* will flow through the resistor in direct proportion to that voltage. The reciprocal of the constant of proportionality is known as the resistance *R*, since, with a given voltage V, a larger value of *R* further "resists" the flow of current *I* as given by Ohm's law: $I = \frac{V}{R}$. Voltages can be divided with the use of resistors, and in combination with other components resistors can be used to make electrical waves into shapes most suited for the electrical designer's requirements. Practical resistors can be made of various compounds and films, as well as resistance wire (wire made of a high-resistivity alloy, such as nickel-chrome). Resistors are also implemented within integrated circuits, particularly analog devices, and can also be integrated into hybrid and printed circuits.

The electrical functionality of a resistor is specified by its resistance: common commercial resistors are manufactured over a range of more than 9 orders of magnitude. When specifying that resistance in an electronic design, the required precision of the resistance may require attention to the manufacturing tolerance of the chosen resistor, according to its specific application. The temperature coefficient of the resistance may also be of concern in some precision applications. Practical resistors are also specified as having a maximum power rating which must exceed the anticipated power dissipation of that resistor in a particular circuit: this is mainly of concern in power electronics applications. Resistors with higher power ratings are physically larger and may require heat sinking. In a high voltage circuit, attention must sometimes be paid to the rated maximum working voltage of the resistor.

Resistors limit current. In a typical application, a resistor is connected in series with an LED as illustrated in Figure 2.24. Enough current flows to make the LED light up, but not so much that the LED is damaged. You are now ready to calculate a value for the resistor used in series with an LED. A typical LED requires a current of 10 mA and has a voltage of 2 V across it when it is working. The power supply for the circuit is 9 V. What is the voltage across the resistor? The answer is 9-2=7 V. You now have two bits of information about R1: the current flowing is 10 mA, and the voltage across R1 is 7 V. You can calculate the value of the resistor using Ohm's law;

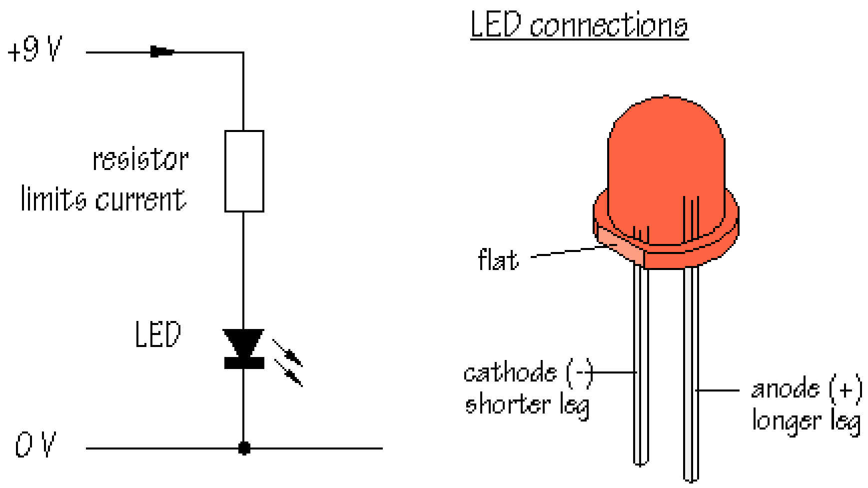


Figure 2. 24 A resistor that limits the current through a light emitting diode (led)

$$R = \frac{7}{0.01} = 700\Omega$$

The calculated value for the resistor is 700 Ω. As you will see below, resistors are manufactures at standard values and 680 Ω, 750 Ω and 820 Ω are available in E12/E24 series. 680 Ω is the obvious choice. This would allow a current slightly greater than 10 mA to flow. Most LEDs are undamaged by currents of up to 20 mA, so this is fine.

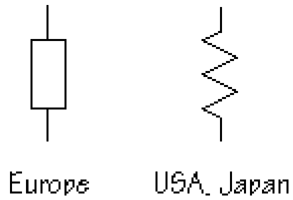
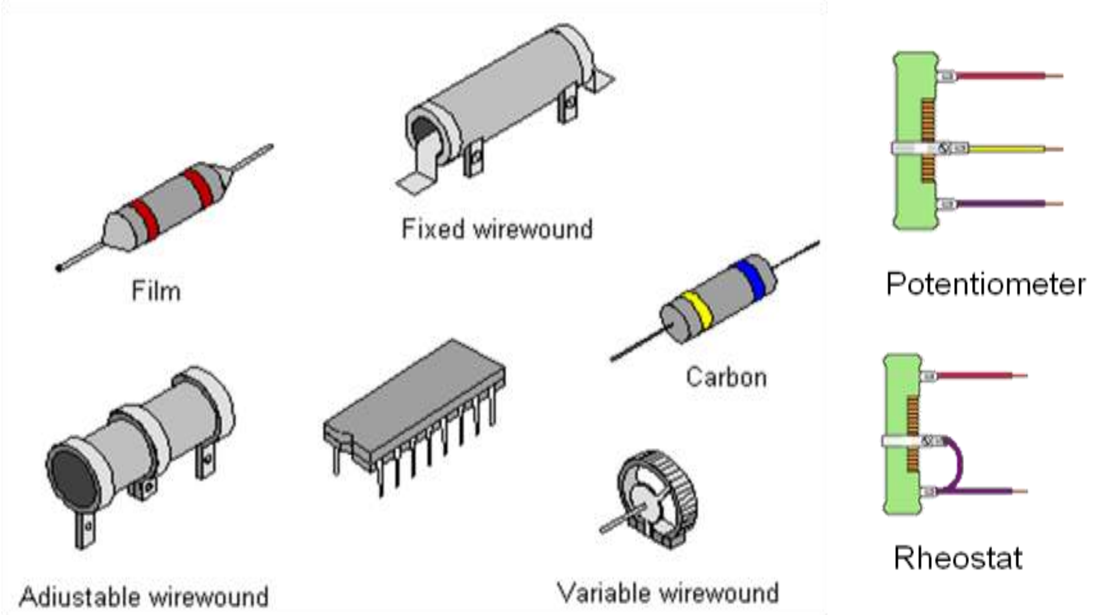


Figure 2. 25 Symbol of a resistor

Symbols of resistors are shown in Figure 2.25. The 'box' symbol for a fixed resistor is popular in the UK and Europe. A 'zig-zag' symbol is used in America and Japan.



Types of Fixed Resistors

The electrical behavior of a resistor obeys Ohm's law for a constant resistance; however, some resistors are sensitive to heat, light, or other variables. Resistors can have a fixed value of resistance, or they can be made variable or adjustable within a certain range, in which case they may be called rheostats, or potentiometers (Figure 2.26). The fixed resistor is an electrical component designed to introduce a known value of resistance into a circuit. Resistors are often made out of chunks of carbon or thin films of carbon or other resistive materials. They can also be made of wires wound around a cylinder.

The common resistor is a two-wire package with a fixed resistance measured in ohms; however, different types of resistors are adjustable by the circuit designer or the user. **Variable resistors** , or **rheostats**, have a resistance that may be varied across a certain range, usually by means of a mechanical device that alters the position of one terminal of the resistor along a strip of resistant

Figure 2. 26 Examples of fixed and variable resistors

material. The length of the intervening material determines the resistance. Mechanical variable resistors are also called *potentiometers,* and are used in the volume knobs of audio equipment and in many other devices.

Discrete resistors are individual packages as illustrated in Figure 2.27. On a circuit board, discrete axial resistors are commonly used with their two wires soldered into the holes of the board. Through-hole components typically have leads leaving the body axially. Others have leads coming off their body radially instead of parallel to the resistor axis. Other components may be SMT (surface mount technology) while high power resistors may have one of their leads

Figure 2. 27 Samples of axial resistors

designed into the heat sink. Generally smaller than axial resistors, discrete surface-mounted resistors are soldered on top of the board. In addition, resistors are built into microprocessors and other integrated circuits (ICs), but they use semiconductor structures for their fabrication in a manner similar to transistors and PN junctions.

A single in line (SIL) resistor package with 8 individual, 47 Ω resistors is shown in Figure 2.28. One end of each resistor is connected to a separate pin and the other ends are all connected together to the remaining (common) pin - pin 1, at the end identified by the white dot.

Figure 2. 28 Resistors in an SIL package

Carbon Composition Resistors

Carbon composition resistors consist of a solid cylindrical resistive element with embedded wire leads or metal end caps to which the lead wires are attached. The body of the resistor is protected with paint or plastic. These resistors were the mainstay of the radio and television industries prior to World War II. The resistive element is made from a mixture of finely ground (powdered) carbon and an insulating material (usually ceramic). A resin holds the mixture together. The conductive path is from particle to particle, each of which touches another along the path. Early 20th-century carbon composition resistors had uninsulated bodies; the lead wires were wrapped around the ends of the resistance element rod and soldered. The completed resistor was painted for color coding of its value.

These resistors were commonly used in the 1960s and earlier, but are not so popular for general use now as other types have better specifications, such as tolerance, voltage dependence, and stress (carbon composition resistors will change value when stressed with over-voltages). Moreover, if internal moisture content (from exposure for some length of time to a humid environment) is significant, soldering heat will create a non-reversible change in resistance value. Carbon composition resistors have poor stability with time and were consequently factory sorted to, at best, only 5% tolerance. These resistors, however, if never subjected to overvoltage nor overheating were remarkably reliable considering the component's size.

Carbon composition resistors were eclipsed in the early 60's by discrete metal film resistors. It was not noise levels but the rising cost of carbon composition resistors compared to falling prices for metal film devices that was the leading factor in their decline. They are still available, but comparatively quite costly. Values ranged from fractions of an ohm to 22 megohms. Because of the high price, these resistors are no longer used in most applications. However, carbon resistors are used in power supplies and welding controls.

Carbon Film Resistors

A carbon film is deposited on an insulating substrate, and a helix cut in it to create a long, narrow resistive path. Varying shapes, coupled with the resistivity of carbon, (ranging from 90 to 400 nΩ m) can provide a variety of resistances. Carbon film resistors feature a power rating range of 0.125 W to 5 W at 70 °C. Resistances available range from 1 Ω to 10 MΩ. The carbon film resistor has an operating temperature range of -55 °C to 155 °C. It has 200 to 600 volts maximum working voltage range. Special carbon film resistors are used in applications requiring high pulse stability.

The diagram in Figure 2.29 shows the construction of a **carbon film** resistor:

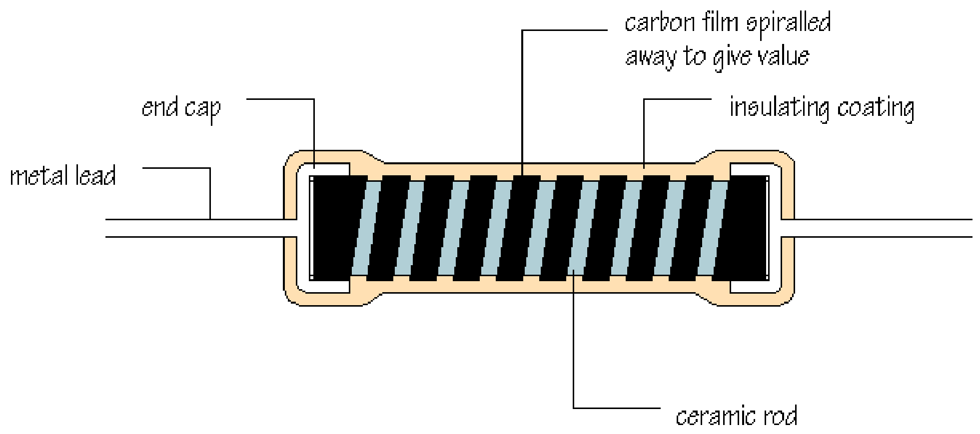


Figure 2. 29 Illustration of construction of a thin film resistor

During manufacture, a thin film of carbon is deposited onto a small ceramic rod. The resistive coating is spiraled away in an automatic machine until the resistance between the two ends of the rod is as close as possible to the correct value. Metal leads and end caps are added, the resistor is covered with an insulating coating and finally painted with colored bands to indicate the resistor value.

Metal Film Resistors

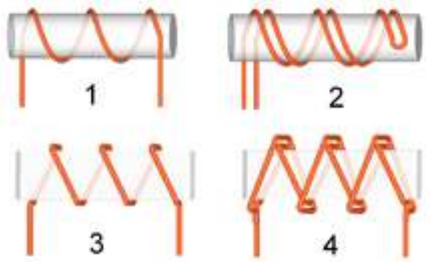
The introduction of metal film technologies brought significant reductions in both resistor size and noise. Metal film resistors are manufactured through the evaporation or sputtering of a layer of nickel chromium onto a ceramic substrate. The thickness of the layer is value-dependent and ranges from 10 Angstroms to 500 Angstroms thick. The thicker this layer is (the lower the value), the less noise is inserted. Higher values are noisier because the occlusions, surface imperfections, and non-uniform depositions are more significant to the production of noise when the nickel chromium layer is thin.

Grinding or laser adjusting techniques are used to generate the resistance grid. The first of these methods leaves a ragged edge and the second leaves a curled edge with eddy-current paths. Both are a source of noise, which is why metal film resistors have a noise range of -32 dB to -16 dB.

There are resistors that resemble metal film types, but are made of metal oxides such as tin oxide. This results in a higher operating temperature and greater stability/reliability than Metal film. They are used in applications with high endurance demands.

Wire-wound resistors

Wire-wound resistors are made of alloys similar to that used in foil resistors, described below. As a result, the only noise insertion caused by these devices comes from the tabs used to connect the fine wire to the coarse external leads. Because of the very high surface temperature these resistors can withstand temperatures of up to +450 °C. The aluminum-cased types are designed to be attached to a heat sink to dissipate the heat; the rated power is dependent on being used with a suitable heat sink, e.g., a 50 W power rated resistor will overheat at a fraction of the power dissipation if not used with a heat sink. Large wire-wound resistors may be rated for 1,000 watts or more.



Types of windings in wire resistors:
1 - common
2 - bifilar
3 - common on a thin former
4 - Ayrton-Perry

Figure 2. 30 Illustration of wire-wound resistors

Figure 2.30 shows four construction types of wire-wound resistors. Because wire-wound resistors are coils they have more undesirable inductance than other types of resistor, although winding the wire in sections with alternately reversed direction can minimize inductance. Other techniques employ bifilar winding, or a flat thin former (to reduce cross-section area of the coil). For most demanding circuits resistors with Ayrton-Perry winding are used.

Applications of wire-wound resistors are similar to those of composition resistors with the exception of the high frequency. A typical noise rating is -38 dB. The high frequency of wire-wound resistors is substantially worse than that of a composition resistor which is the major objection. Of serious concern instead is the inductance that chops the peaks and fails to replicate the higher frequencies of the second and third harmonics.

Foil Resistors

The primary resistance element of a foil resistor is a special alloy foil several micrometers thick. Since their introduction in the 1960s, foil resistors have had the best precision and stability of any resistor available. One of the important parameters influencing stability is the temperature coefficient of resistance (TCR). The TCR of foil resistors is extremely low, and has been further improved over the years. One range of ultra-precision foil resistors offers a TCR of 0.14 ppm/°C, tolerance ±0.005%, long-term stability (1 year) 25 ppm, (3 year) 50 ppm (further improved 5-fold by hermetic sealing), stability under load (2000 hours) 0.03%, thermal EMF 0.1 μV/°C, noise -42 dB, voltage coefficient 0.1 ppm/V, inductance 0.08 μH, capacitance 0.5 pF.

Carbon film resistors are cheap and easily available, with values within ±10% or ±5% of their marked, or 'nominal' value. **Metal film** and **metal oxide** resistors are made in a similar way, but can be made more accurately to within ±2% or ±1% of their nominal value. There are some differences in performance between these resistor types, but none which affect their use in simple circuits.

Wire-wound resistors are made by winding thin wire onto a ceramic rod. They can be made extremely accurately for use in multimeters, oscilloscopes and other measuring equipment. Some types of wire-wound resistors can pass large currents without overheating and are used in power supplies and other high current circuits.

Adjustable Resistors

A resistor may have one or more fixed tapping points so that the resistance can be changed by moving the connecting wires to different terminals. Some wire-wound power resistors have a tapping point that can slide along the resistance element, allowing a larger or smaller part of the resistance to be used. Where continuous adjustment of the resistance value during operation of equipment is required, the sliding resistance tap can be connected to a knob accessible to an operator. Such a device is called a rheostat and has two terminals.

Potentiometers

A common element in electronic devices is a three-terminal resistor with a continuously adjustable tapping point controlled by rotation of a shaft or knob. These variable resistors are known as potentiometers when all three terminals are present, since they act as a continuously adjustable voltage divider. A common example is a volume control for a radio receiver.

Accurate, high-resolution panel-mounted potentiometers (or "pots") have resistance elements typically wire-wound on a helical mandrel, although some include a conductive-plastic resistance coating over the wire to improve resolution. These typically offer ten turns of their shafts to cover their full range. They are usually set with dials that include a simple turns counter and a graduated dial. Electronic analog computers used them in quantity for setting coefficients, and delayed-sweep oscilloscopes of recent decades included one on their panels.

Resistance Decade Boxes

A resistance decade box or resistor substitution box is a unit containing resistors of many values, with one or more mechanical switches which allow any one of various discrete resistances offered by the box to be dialed in. Usually the resistance is accurate to high precision, ranging from laboratory/calibration grade accuracy of 20 parts per million, to field grade at 1%. Inexpensive boxes with lesser accuracy are also available. All types offer a convenient way of selecting and quickly changing a resistance in laboratory, experimental and development work without needing to attach resistors one by one, or even stock each value. The range of resistance provided, the maximum resolution, and the accuracy characterize the box. For example, one box offers resistances from 0 to 24 MΩ, maximum resolution 0.1 Ω, accuracy 0.1%.

Special Devices

There are various devices whose resistance changes with various quantities. The resistance of thermistors exhibit a strong negative temperature coefficient, making them useful for measuring temperatures. Since their resistance can be large until they are allowed to heat up due to the passage of current, they are also commonly used to prevent excessive current surges when equipment is powered on. Metal oxide varistors drop to a very low resistance when a high voltage is applied, making them useful for protecting electronic equipment by absorbing dangerous voltage surges. One sort of photodetector, the photoresistor, has a resistance which varies with illumination. The strain gauge is a type of resistor that changes value with applied strain. A single resistor may be used, or a pair (half bridge), or four resistors connected in a Wheatstone bridge configuration. The strain resistor is bonded with adhesive to an object that will be subjected to mechanical strain. With the strain gauge and a filter, amplifier, and analog/digital converter, the strain on an object can be measured. Some of these devices will be discussed later in detail with application examples.

Resistor Marking

Most axial resistors use a pattern of colored stripes to indicate resistance. Surface-mount resistors are marked numerically, if they are big enough to permit marking; more-recent small sizes are impractical to mark. Cases are usually tan, brown, blue, or green, though other colors are occasionally found such as dark red or dark gray. Early 20th century resistors, essentially uninsulated, were dipped in paint to cover their entire body for color coding. A second color of paint was applied to one end of the element, and a color dot (or band) in the middle provided the third digit. The rule was "body, tip, dot", providing two significant digits for value and the decimal multiplier, in that sequence. Default tolerance was ±20%. Closer-tolerance resistors had silver (±10%) or gold-colored (±5%) paint on the other end.

Four-Band Resistors

Four-band identification is the most commonly used color-coding scheme on resistors. It consists of four colored bands that are painted around the body of the resistor. The first two bands encode the first two significant digits of the resistance value, the third is a power-of-ten multiplier or number-of-zeroes, and the fourth is the tolerance accuracy, or acceptable error, of the value. The first three bands are equally spaced along the resistor; the spacing to the fourth band is wider. Sometimes a fifth band identifies the thermal coefficient, but this must be distinguished from the true 5-color system, with 3 significant digits. Each color corresponds to a certain digit, progressing from darker to lighter colors, as shown in the chart in Table 2.6.

Table 2. 6 Color codes for discrete resistors

Color	1st band	2nd band	3rd band (multiplier)	4th band (tolerance)	Temp. Coefficient
Black	0	0	$\times 10^0$		
Brown	1	1	$\times 10^1$	±1% (F)	100 ppm
Red	2	2	$\times 10^2$	±2% (G)	50 ppm
Orange	3	3	$\times 10^3$		15 ppm
Yellow	4	4	$\times 10^4$		25 ppm
Green	5	5	$\times 10^5$	±0.5% (D)	
Blue	6	6	$\times 10^6$	±0.25% (C)	
Violet	7	7	$\times 10^7$	±0.1% (B)	
Gray	8	8	$\times 10^8$	±0.05% (A)	
White	9	9	$\times 10^9$		
Gold			$\times 10^{-1}$	±5% (J)	
Silver			$\times 10^{-2}$	±10% (K)	
None				±20% (M)	

The tolerance of a resistor is shown by the **fourth band** of the color code. Tolerance is the **precision** of the resistor and it is given as a percentage. For example a 390 Ω resistor with a tolerance of ±10% will have a value within 10% of 390 Ω, between 390 - 39 = 351 Ω and 390 + 39 = 429 Ω (39 is 10% of 390).

An example of a four-band resistor is shown in Figure 2.31. When you want to read off a resistor value, look for the tolerance band, usually gold, and hold the resistor with the tolerance band at its right hand end. Reading resistor values quickly and accurately isn't difficult, but it does take practice!

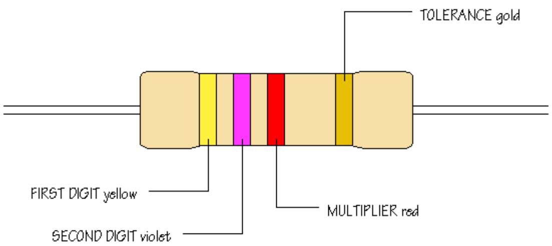


Figure 2. 31 Color codes for a four-band resistor

The first band on a resistor is interpreted as the FIRST DIGIT of the resistor value. For the resistor shown below, the first band is yellow, so the first digit is 4. The second band gives the SECOND DIGIT.

This is a violet band, making the second digit 7. The third band is called the MULTIPLIER and is not interpreted in quite the same way. The multiplier tells you how many naught you should write after the digits you already have. A red band tells you to add 2 naught. The value of this resistor is therefore 4 7 0 0 ohms, that is, 4 700 Ω, or 4.7 kΩ. Work through this example again to confirm that you understand how to apply the color code given by the first three bands. The remaining band is the TOLERANCE band. This indicates the percentage accuracy of the resistor value. Most carbon film resistors have a gold-colored tolerance band, indicating that the actual resistance value is with + or - 5% of the nominal value. Other tolerance colors are gold for 10%, red for 2% and for brown 1%. If no fourth band is shown the tolerance is ±20%. Tolerance may be ignored for almost all circuits because precise resistor values are rarely required.

For example, green-blue-yellow-red is 56×10^4 Ω = 560 kΩ ± 2%. An easier description can be as followed: the first band, green, has a value of 5 and the second band, blue, has a value of 6, and is counted as 56. The third band, yellow, has a value of 10^4, which adds four 0's to the end, creating 560,000 Ω at ±2% tolerance accuracy. 560,000 Ω changes to 560 kΩ ±2% (as a kilo- is 10^3).

Marking Low Valued Resistors

The color code as explained above allows you to interpret the values of any resistor from 100 Ω upwards. How does the code work for values less than 100 Ω? Here is the code for 12 Ω: brown, red, black

The multiplier color black represents the number 0 and tells you that *no* naught should be added to the first two digits, representing 1 and 2.

- What would be the color code for 47 Ω? The answer is: yellow, violet, black
- Using this method for indicating values between 10 Ω and 100 Ω means that all resistor values require the same number of bands.

The standard color code cannot show values of less than 10 Ω. To show these small values two special colors are used for the **third band**: **gold** which means × 0.1 and **silver** which means × 0.01. The first and second bands represent the digits as normal.

For example:

- **brown, black, gold** bands represent 10 × 0.1 = 1 Ω
- **red, red, gold** bands represent 22 × 0.1 = 2.2 Ω
- **red**, **violet**, **gold** bands represent 27 × 0.1 = 2.7 Ω
- **green**, **blue**, **silver** bands represent 56 × 0.01 = 0.56 Ω

Five-Band Axial Resistors

5-band identification is used for higher precision (lower tolerance) resistors (1%, 0.5%, 0.25%, 0.1%), to specify a third significant digit. The first three bands represent the significant digits, the fourth is the multiplier, and the fifth is the tolerance. Five-band resistors with a gold or silver 4th band are sometimes encountered, generally on older or specialized resistors. The 4th band is the tolerance and the 5th the temperature coefficient.

Metal film resistors, manufactured to 1 or 2% tolerance, often use a code consisting of *four* colored bands instead of three. The code works in the same way, with the first three bands interpreted as digits and the fourth band as the multiplier. For example, a 1 kΩ metal film resistor has the bands: brown, black, black, brown (+brown or red for tolerance), while a 56 kΩ metal film resistor has the bands: green, blue, black, red. It is worth pointing out that the multiplier for metal film resistors with values from 1 kΩ upwards is brown (rather than red, as in the three color system), while the multiplier for 10 kΩ upwards is red (instead of orange).

Resistor Shorthand

Resistor values are often written on circuit diagrams using a code system which avoids using a decimal point because it is easy to miss the small dot. Instead the letters R, K and M are used in place of the decimal point. To read the code: replace the letter with a decimal point, then multiply the value by 1000 if the letter was K, or 1000000 if the letter was M. The letter R means multiply by 1.

For example:

- 560R means 560 Ω
- 2K7 means 2.7 kΩ = 2700 Ω
- 39K means 39 kΩ
- 1M0 means 1.0 MΩ = 1000 kΩ

SMD Resistors

The image in Figure 2.32 shows four surface-mount resistors (the component at the upper left is a capacitor) including two zero-ohm resistors. Zero-ohm links are often used instead of wire links, so that they can be inserted by a resistor-inserting machine. Of course, their resistance is non-zero, although quite low. *Zero* is simply a brief description of their function.

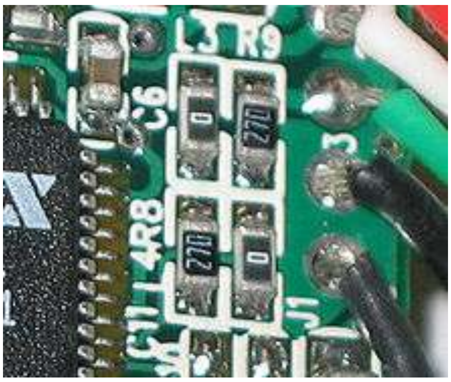


Figure 2. 32 SMD resistors in circuit

Surface mounted resistors are printed with numerical values in a code related to that used on axial resistors. Standard-tolerance surface-mount technology (SMT) resistors are marked with a three-digit code, in which the first two digits are the first two significant digits of the value and the third digit is the power of ten (the number of zeroes). For example:

334 = 33 × 10^4 ohms = 330 kΩ	*222* = 22 × 10^2 ohms = 2.2 kΩ
473 = 47 × 10^3 ohms = 47 kΩ	*105* = 10 × 10^5 ohms = 1.0 MΩ

Resistances less than 100 ohms are written: 100, 220, 470. The final zero represents ten to the power zero, which is 1. For example:

• *100* = 10 × 10^0 ohm = 10 Ω	• *220* = 22 × 10^0 ohms = 22 Ω

Sometimes these values are marked as *10* or *22* to prevent a mistake.

Resistances less than 10 ohms have 'R' to indicate the position of the decimal point (radix point). For example:

- *4R7* = 4.7 ohms
- *0R22* = 0.22 ohms
- *R300* = 0.30 ohms
- *0R01* = 0.01 ohms

Precision resistors are marked with a four-digit code, in which the first three digits are the significant figures and the fourth is the power of ten. For example:

- 1001 = 100 × 10^1 ohms = 1.00 kΩ
- 4992 = 499 × 10^2 ohms = 49.9 kΩ
- 1000 = 100 × 10^0 ohm = 100 Ω

000 and *0000* sometimes appear as values on surface-mount zero-ohm links, since these have (approximately) zero resistance.

More recent surface-mount resistors are too small, physically, to permit practical markings to be applied.

Preferred Values

Early resistors were made in more or less arbitrary round numbers; a series might have 100, 125, 150, 200, 300, etc. Resistors as manufactured are subject to a certain percentage tolerance, and it makes sense to manufacture values that correlate with the tolerance, so that the actual value of a resistor overlaps slightly with its neighbors. Wider spacing leaves gaps; narrower spacing increases manufacturing and inventory costs to provide resistors that are more or less interchangeable.

A logical scheme is to produce resistors in a range of values which increase in a geometrical progression, so that each value is greater than its predecessor by a fixed multiplier or percentage, chosen to match the tolerance of the range. For example, for a tolerance of ±20% it makes sense to have each resistor about 1.5 times its predecessor, covering a decade in 6 values. In practice the factor used is 1.4678, giving values of 1.47, 2.15, 3.16, 4.64, 6.81, 10 for the 1-10 decade (a decade is a range increasing by a factor of 10; 0.1-1 and 10-100 are other examples); these are rounded in practice to 1.5, 2.2, 3.3, 4.7, 6.8, 10; followed, of course by 15, 22, 33, … and preceded by … 0.47, 0.68, 1. This scheme has been adopted as the **E6** range of the International Electrotechnical Commission (IEC) 60063 preferred number series. There are also **E12**, **E24**, **E48**, **E96** and **E192** ranges for components of ever tighter tolerance, with 12, 24, 48, 96, and 192 different values within each decade. The actual values used are in the IEC 60063 lists of preferred numbers.

A resistor of 100 ohms ±20% would be expected to have a value between 80 and 120 ohms; its E6 neighbors are 68 (54-82) and 150 (120-180) ohms. A sensible spacing, E6 is used for ±20% components. E12 for ±10% and 12 values in one decade is used. Among other IEC 60063 series, E24 for ±5%; E48 for ±2%, E96 for ±1%; E192 for ±0.5% or better.

Consider 100 Ω and 120 Ω , adjacent values in the E12 range. 10% of 100 Ω is 10 Ω, while 10% of 120 Ω is 12 Ω. A resistor marked as 100 Ω could have any value from 90 Ω to 110 Ω, while a resistor marked as 120 Ω might have an actual resistance from 108 Ω to 132 Ω. The ranges of possible values overlap, but only slightly. Further up the E12 range, a resistor marked as 680 Ω might have and actual resistance of up to 680+68=748 Ω, while a resistor marked as 820 Ω might have a resistance as low as 820-82=738 Ω. Again, the ranges of possible values just overlap.

Resistors are manufactured in values from a few milliohms to about a gigaohm in IEC60063 ranges appropriate for their tolerance. Preferred values in one decade in E 12 and E24 series of resistors are given in Table 2.7.

Table 2. 7 Preferred values of resistors in one decade in E12 and E24 series

E12	10		12		15		18		22		27		33		39		47		56		68		82	
E24	10	11	12	13	15	16	18	20	22	24	27	30	33	36	39	43	47	51	56	62	68	75	82	91

Earlier power wire-wound resistors, such as brown vitreous-enameled types, however, were made with a different system of preferred values, such as some of those mentioned in the first sentence of this section.

Power Ratings of Resistors

When current flows through a resistance, electrical energy is converted into heat. The total amount of heat energy released over a period of time can be determined from the integral of the power over that period of time: $W = \int_{t_1}^{t_2} v(t)i(t)dt$

The power P dissipated by a resistor (or the equivalent resistance of a resistor network) is calculated as: $P = I^2R = IV = \frac{V^2}{R}$

Example:

What is the power output of a resistor when the voltage across it is 6 V, and the current flowing through it is 100 mA?

$P = IV =$6x100 mA=600 mW=0.6 W

0.6 W of heat are generated in this resistor. To prevent overheating, it must be possible for heat to be lost, or dissipated, to the surroundings at the same rate. The first form is a restatement of Joule's first law. Using Ohm's law, the two other forms can be derived.

Practical resistors are rated according to their maximum power dissipation. The vast majority of resistors used in electronic circuits absorb much less than a watt of electrical power and require no attention to their power rating. Such resistors in their discrete form, including most of the packages detailed below, are typically rated as 1/10, 1/8, or 1/4 watt. Resistors required to dissipate substantial amounts of power, particularly used in power supplies, power conversion circuits, and power amplifiers, are generally referred to as *power resistors*; this designation is loosely applied to resistors with power ratings of 1 watt or greater. Power resistors are physically larger and tend not to use the preferred values, color codes, and external packages described previously. Figure 2.33 shows power ratings of various resistors.

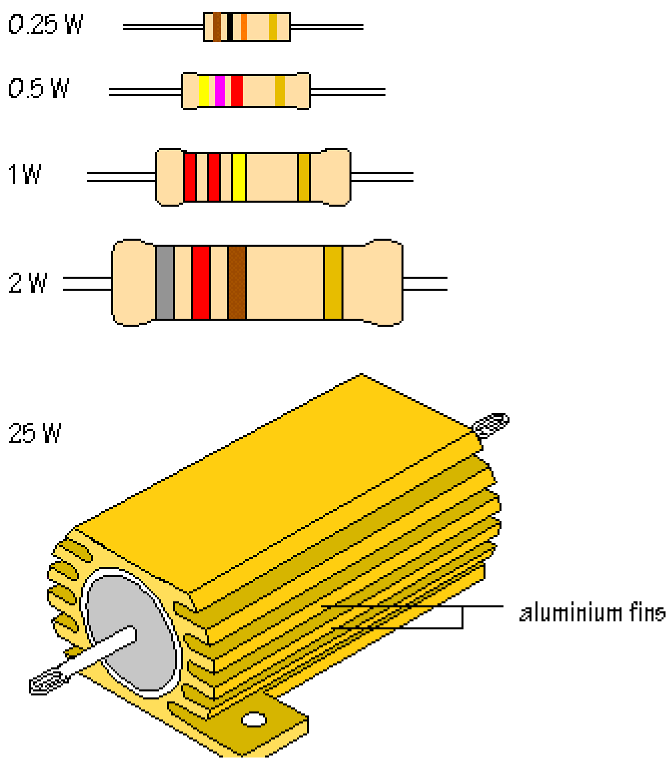


Figure 2. 33 Resistors for various power ratings

Power ratings of resistors are rarely quoted in parts lists because for most circuits the standard power ratings of 0.25W or 0.5W are suitable. For the rare cases where a higher power is required it should be clearly specified in the parts list, these will be circuits using **low value resistors** (less than about 300Ω) or **high voltages** (more than 15V).

Examples:

- A 470 Ω resistor with 10V across it, needs a power rating $P = V^2/R = 10^2/470 = 0.21W$. *In this case a standard 0.25W resistor would be suitable.*
- A 27 Ω resistor with 10V across it, needs a power rating $P = V^2/R = 10^2/27 = 3.7W$. *A high power resistor with a rating of 5W would be suitable.*

If the average power dissipated by a resistor is more than its power rating, damage to the resistor may occur, permanently altering its resistance; this is distinct from the reversible change in resistance due to its temperature coefficient when it warms. Excessive power dissipation may raise the temperature of the resistor to a point where it can burn the circuit board or adjacent components, or even cause a fire. There are flameproof resistors that fail (open circuit) before they overheat dangerously.

Note that the nominal power rating of a resistor is not the same as the power that it can safely dissipate in practical use. Air circulation and proximity to a circuit board, ambient temperature, and other factors can reduce acceptable dissipation significantly. Rated power dissipation may be given for an ambient temperature of 25 °C in free air. Inside an equipment case at 60 °C, rated dissipation will be significantly less; a resistor dissipating a bit less than the maximum figure given by the manufacturer may still be outside the safe operating area and may prematurely fail.

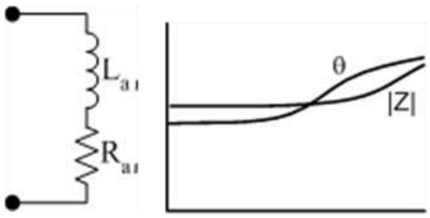


Figure 2. 34 Model of a low value resistor

Resistors at High Frequencies

The major problem with resistors at high frequencies is for wire-wound (power) resistors, that will act as inductors at high frequencies as illustrated in Figure 2.34. In addition, very small resistors, like chip resistors, can also exhibit capacitive effects. Special high frequency resistors are designed to offset these effect. The series inductance of a practical resistor causes its behavior to depart from ohms law; this specification can be important in some high-frequency applications for smaller values of resistance.

Noise in Resistors

In amplifying faint signals, it is often necessary to minimize electronic noise, particularly in the first stage of amplification. As dissipative elements, even an ideal resistor will naturally produce a randomly fluctuating voltage or "noise" across its terminals and eventually it is a fundamental noise source which depends only upon the temperature and resistance of the resistor. Using a larger resistor produces a larger voltage noise, whereas with a smaller value of resistance there will be more current noise, assuming a given temperature. The thermal noise of a practical resistor may also be somewhat larger than the theoretical prediction and that increase is typically frequency-dependent.

However, the "excess noise" of a practical resistor is an additional source of noise observed only when a current flows through it. This is specified in unit of μV/V/decade - μV of noise per volt applied across the resistor per decade of frequency. A noise index is expressed in decibels (dB), and the equation converting μV/V to dB is:

dB = 20 x log (noise voltage [in μV]/DC voltage [in V]).

For example, 0 dB equates to 1.0 μV/V, and 15 dB equates to 5.6 μV/V.

Hence, the μV/V/decade value of a resistor with a noise index of 0 dB will exhibit 1 μV (rms) of excess noise for each volt across the resistor in each frequency decade. Excess noise is thus an example of 1/*f* noise. Thick-film and carbon composition resistors generate more excess noise than other types at low frequencies; wire-wound and thin-film resistors, though much more expensive, are often utilized for their better noise characteristics. Carbon composition resistors can exhibit a noise index of 0 dB while bulk metal foil resistors may have a noise index of -40 dB, usually making the excess noise of metal foil resistors insignificant. Thin film surface mount resistors typically have lower noise and better thermal stability than thick film surface mount resistors. However, the design engineer must read the data sheets for the family of devices to weigh the various device tradeoffs.

Failure Modes

The failure rate of resistors in a properly designed circuit is low compared to other electronic components such as semiconductors and electrolytic capacitors. Damage to resistors most often occurs due to overheating when the average power delivered to it (as computed above) greatly exceeds its ability to dissipate heat (specified by the resistor's *power rating*). This may be due to a fault external to the circuit, but is frequently caused by the failure of another component (such as a transistor that shorts out) in the circuit connected to the resistor. Operating a resistor too close to its power rating can limit the resistor's lifespan or cause a change in its resistance over time which may or may not be noticeable. A safe design generally uses overrated resistors in power applications to avoid this danger.

When overheated, carbon-film resistors may decrease or increase in resistance. Carbon film and composition resistors can fail (open circuit) if running close to their maximum dissipation. This is also possible but less likely with metal film and wire-wound resistors. There can also be failure of resistors due to mechanical stress and adverse environmental factors including humidity. If not enclosed, wire-wound resistors can corrode.

Variable resistors degrade in a different manner, typically involving poor contact between the wiper and the body of the resistance. This may be due to dirt or corrosion and is typically perceived as "crackling" as the contact resistance fluctuates; this is especially noticed as the device is adjusted. This is similar to crackling caused by poor contact in switches, and like switches, potentiometers are to some extent self-cleaning: running the wiper across the resistance may improve the contact. Potentiometers which are seldom adjusted, especially in dirty or harsh environments, are most likely to develop this problem. When self-cleaning of the contact is insufficient, improvement can usually be obtained through the use of contact cleaner (also known as "tuner cleaner") spray. The crackling noise associated with turning the shaft of a dirty potentiometer in an audio circuit (such as the volume control) is greatly accentuated when an undesired DC voltage is present, often implicating the failure of a DC blocking capacitor in the circuit.

In a low-noise amplifier or pre-amp the noise characteristics of a resistor may be an issue. The unwanted inductance, excess noise, and temperature coefficient are mainly dependent on the technology used in manufacturing the resistor. They are not normally specified individually for a particular family of resistors manufactured using a particular technology. A family of discrete resistors is also characterized according to its form factor, that is, the size of the device and position of its leads (or terminals) which is relevant in the practical manufacturing of circuits using them.

CAPACITORS

Definition and Use

A **capacitor** (formerly known as **condenser**) is a passive electronic component consisting of a pair of conductors separated by a dielectric (insulator) as shown in Figure 2.35. When there is a potential difference (voltage) across the conductors, a static electric field develops in the dielectric that stores energy and produces a mechanical force between the conductors. An ideal capacitor is characterized by a single constant value, capacitance, measured in farads. This is the ratio of the electric charge on each conductor to the potential difference between them.

Figure 2. 35 The basic capacitor

Capacitors are widely used in electronic circuits for blocking direct current while allowing alternating current to pass, in filter networks, for smoothing the output of power supplies, in the resonant circuits that tune radios to particular frequencies and for many other purposes. The effect is greatest when there is a narrow separation between large areas of conductor, hence capacitor conductors are often called "plates", referring to an early means of construction. In practice the dielectric between the plates passes a small amount of leakage current and also has an electric field strength limit, resulting in a breakdown voltage, while the conductors and leads introduce an undesired inductance and resistance.

Parallel Plate Model

A capacitor consists of two conductors separated by a non-conductive region called the dielectric medium though it may be a vacuum or a semiconductor depletion region chemically identical to the conductors. A capacitor is assumed to be self-contained and isolated, with no net electric charge and no influence from any external electric field. Charge separation in a parallel-plate capacitor causes an internal electric field as illustrated in Figure 2.36. A dielectric (orange) reduces the field and increases the capacitance. The conductors thus hold equal and opposite charges on their facing surfaces, and the dielectric develops an electric field. In SI units, a capacitance of one farad means that one coulomb of charge on each conductor causes a voltage of one volt across the device.

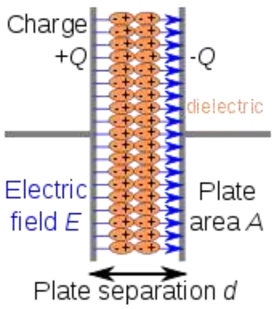


Figure 2. 36 Construction of a simple capacitor

The capacitor is a reasonably general model for electric fields within electric circuits. An ideal capacitor is wholly characterized by a constant capacitance *C*, defined as the ratio of charge ±*Q* on each conductor to the voltage *V* between them: $C = \frac{Q}{V}$

Sometimes charge build-up affects the capacitor mechanically, causing its capacitance to vary. In this case, capacitance is defined in terms of incremental changes: $C = \frac{dq}{dv}$.

Energy Storage

Work must be done by an external influence to "move" charge between the conductors in a capacitor. When the external influence is removed the charge separation persists in the electric field and energy is stored to be released when the charge is allowed to return to its equilibrium position. The work done in establishing the electric field, and hence the amount of energy stored, is given by: $W = \int_{q=0}^{Q} V dq = \int_{q=0}^{Q} \frac{q}{C} dq = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}CV^2 = \frac{1}{2}VQ$

Current-Voltage Relation

The current $i(t)$ through any component in an electric circuit is defined as the rate of flow of a charge $q(t)$ passing through it, but actual charges, electrons, cannot pass through the dielectric layer of a capacitor, rather an electron accumulates on the negative plate for each one that leaves the positive plate, resulting in an electron depletion and consequent positive charge on one electrode that is equal and opposite to the accumulated negative charge on the other. Thus the charge on the electrodes is equal to the integral of the current as well as proportional to the voltage as discussed above. As with any antiderivative, a constant of integration is added to represent the initial voltage $v(t_0)$. This is the integral form of the capacitor equation, $v(t) = \frac{q(t)}{C} = \frac{1}{C}\int_{t_0}^{t} i(\tau) d\tau + v(t_0)$.

Taking the derivative of this, and multiplying by *C*, yields the derivative form, $i(t) = \frac{dq(t)}{dt} = C\frac{dv(t)}{dt}$.

DC Circuits

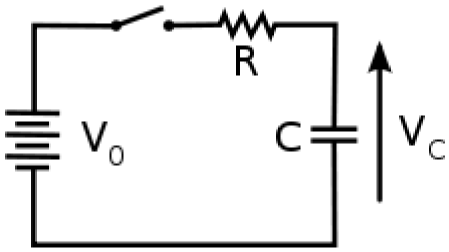


Figure 2. 37 A simple circuit demonstrating charging of a capacitor

A series circuit in Figure 2.37 containing only a resistor, a capacitor, a switch and a constant DC source of voltage V_0 is known as a *charging circuit*. If the capacitor is initially uncharged while the switch is open, and the switch is closed at t = 0, it follows from Kirchhoff's voltage law that

$$V_0 = v_{resistor}(t) + v_{capacitor}(t) = i(t)R + \frac{1}{C}\int_0^t i(\tau)d\tau.$$

Taking the derivative and multiplying by C, gives a first-order differential equation,

$$RC\frac{di(t)}{dt} + i(t) = 0.$$

At t = 0, the voltage across the capacitor is zero and the voltage across the resistor is V_0. The initial current is then i (0) =V_0 /R. With this assumption, the differential equation yields

$$i(t) = \frac{V_0}{R}e^{-t/\tau_0};\ v(t) = V_0(1 - e^{-\frac{t}{\tau_0}}),$$

where τ_0 = RC is the *time constant* of the system.

As the capacitor reaches equilibrium with the source voltage, the voltage across the resistor and the current through the entire circuit decay exponentially. The case of *discharging* a charged capacitor likewise demonstrates exponential decay, but with the initial capacitor voltage replacing V_0 and the final voltage being zero.

AC Circuits

Impedance, the vector sum of reactance and resistance, describes the phase difference and the ratio of amplitudes between sinusoidally varying voltage and sinusoidally varying current at a given frequency. Fourier analysis allows any signal to be constructed from a spectrum of frequencies, whence the circuit's reaction to the various frequencies may be found. The reactance and impedance of a capacitor are respectively

$$X = -\frac{1}{\omega C} = -\frac{1}{2\pi f C};\ Z = \frac{1}{j\omega C} = -\frac{j}{\omega C} = -\frac{j}{2\pi f C}$$

where j is the imaginary unit and ω is the angular velocity of the sinusoidal signal. The $-j$ phase indicates that the AC voltage $V = Z I$ lags the AC current by 90°: the positive current phase corresponds to increasing voltage as the capacitor charges; zero current corresponds to instantaneous constant voltage, etc.

Note that impedance decreases with increasing capacitance and increasing frequency. This implies that a higher-frequency signal or a larger capacitor results in a lower voltage amplitude per current amplitude—an AC "short circuit" or AC coupling. Conversely, for very low frequencies, the reactance will be high, so that a capacitor is nearly an open circuit in AC analysis—those frequencies have been "filtered out". Capacitors are different from resistors and inductors in that the impedance is *inversely* proportional to the defining characteristic, i.e. capacitance.

Non-Ideal Behavior

Capacitors deviate from the ideal capacitor equation in a number of ways. Some of these, such as leakage current and parasitic effects are linear, or can be assumed to be linear, and can be dealt with by adding virtual components to the equivalent circuit of the capacitor. The usual methods of network analysis can then be applied. In other cases, such as with breakdown voltage, the effect is non-linear and normal (i.e., linear) network analysis cannot be used, the effect must be dealt with separately. There is yet another group, which may be linear but invalidate the assumption in the analysis that capacitance is a constant. Such an example is temperature dependence.

Breakdown Voltage

Above a particular electric field, known as the dielectric strength E_{ds}, the dielectric in a capacitor becomes conductive. The voltage at which this occurs is called the breakdown voltage of the device, and is given by the product of the dielectric strength and the separation between the conductors, $V_{\text{bd}} = E_{\text{ds}}d$

The maximum energy that can be stored safely in a capacitor is limited by the breakdown voltage. Due to the scaling of capacitance and breakdown voltage with dielectric thickness, all capacitors made with a particular dielectric have approximately equal maximum energy density, to the extent that the dielectric dominates their volume.

For air dielectric capacitors the breakdown field strength is of the order 2 to 5 MV/m; for mica the breakdown is 100 to 300 MV/m, for oil 15 to 25 MV/m, and can be much less when other materials are used for the dielectric. The dielectric is used in very thin layers and so absolute breakdown voltage of capacitors is limited. Typical ratings for capacitors used for general electronics applications range from a few volts to 100V or so. As the voltage increases, the dielectric must be thicker, making high-voltage capacitors larger than those rated for lower voltages. The breakdown voltage is critically affected by factors such as the geometry of the capacitor conductive parts; sharp edges or points increase the electric field strength at that point and can lead to a local breakdown. Once this starts to happen, the breakdown will quickly "track" through the dielectric till it reaches the opposite plate and cause a short circuit.

The usual breakdown route is that the field strength becomes large enough to pull electrons in the dielectric from their atoms thus causing conduction. Other scenarios are possible, such as impurities in the dielectric, and, if the dielectric is of a crystalline nature, imperfections in the crystal structure can result in an avalanche breakdown as seen in semi-conductor devices. Breakdown voltage is also affected by pressure, humidity and temperature.

Equivalent Circuit

Two different equivalent circuit models of a capacitor is shown in Figure 2.38. An ideal capacitor only stores and releases electrical energy, without dissipating any. In reality, all capacitors have imperfections within the capacitor's material that create resistance. This is specified as the *equivalent series resistance* or **ESR** of a component. This adds a real component to the impedance: $R_C = Z + R_{ESR} = \frac{1}{j\omega C} + R_{ESR}$

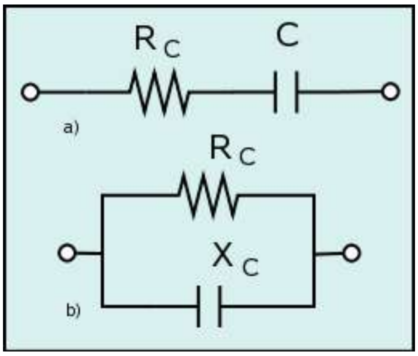


Figure 2. 38 Two different circuit models of a real capacitor

As frequency approaches infinity, the capacitive impedance (or reactance) approaches zero and the ESR becomes significant. As the reactance becomes negligible, power dissipation approaches $P_{RMS} = V_{RMS}^2 / R_{ESR}$.

Similarly to ESR, the capacitor's leads add *equivalent series inductance* or **ESL** to the component. This is usually significant only at relatively high frequencies. As inductive reactance is positive and increases with frequency, above a certain frequency capacitance will be canceled by inductance. High-frequency engineering involves accounting for the inductance of all connections and components.

If the conductors are separated by a material with a small conductivity rather than a perfect dielectric, then a small leakage current flows directly between them. The capacitor therefore has a finite parallel resistance, and slowly discharges over time (time may vary greatly depending on the capacitor material and quality).

Ripple Current

Ripple current is the AC component of an applied source (often a switched-mode power supply) whose frequency may be constant or varying. Certain types of capacitors, such as electrolytic tantalum capacitors, usually have a rating for maximum ripple current (both in frequency and magnitude). This ripple current can cause damaging heat to be generated within the capacitor due to the current flow across resistive imperfections in the materials used within the capacitor, more commonly referred to as equivalent series resistance (ESR). For example electrolytic tantalum capacitors are limited by ripple current and generally have the highest ESR ratings in the capacitor family, while ceramic capacitors generally have no ripple current limitation and have some of the lowest ESR ratings.

Capacitance Instability

The capacitance of certain capacitors decreases as the component ages. In ceramic capacitors, this is caused by degradation of the dielectric. The type of dielectric and the ambient operating and storage temperatures are the most significant aging factors, while the operating voltage has a smaller effect. The aging process may be reversed by heating the component above the Curie point. Aging is fastest near the beginning of life of the component, and the device stabilizes over time. Electrolytic capacitors age as the electrolyte evaporates. In contrast with ceramic capacitors, this occurs towards the end of life of the component.

Temperature dependence of capacitance is usually expressed in parts per million (ppm) per °C. It can usually be taken as a broadly linear function but can be noticeably non-linear at the temperature extremes. The temperature coefficient can be either positive or negative, sometimes even amongst different samples of the same type. In other words, the spread in the range of temperature coefficients can encompass zero. The leakage current section in the data sheet of respective capacitors contains examples of them.

Capacitors, especially ceramic capacitors, and older designs such as paper capacitors, can absorb sound waves resulting in a microphonic effect. Vibration moves the plates, causing the capacitance to vary, in turn inducing AC current. Some dielectrics also generate piezoelectricity. The resulting interference is especially problematic in audio applications, potentially causing feedback or unintended recording. In the reverse microphonic effect, the varying electric field between the capacitor plates exerts a physical force, moving them as a speaker. This can generate audible sound, but drains energy and stresses the dielectric and the electrolyte, if any.

Capacitor Types

Practical capacitors are available commercially in many different forms. The type of internal dielectric, the structure of the plates and the device packaging all strongly affect the characteristics of the capacitor, and its applications. Values available range from very low (picofarad range; while arbitrarily low values are in principle possible, stray (parasitic) capacitance in any circuit is the limiting factor) to about 5 kF super capacitors. Above approximately 1 μF electrolytic capacitors are usually used because of their small size and low cost compared with other technologies, unless their relatively poor stability, life and polarized nature make them unsuitable. Very high capacity super capacitors use a porous carbon-based electrode material.

Dielectric materials

Figure 2. 39 Various capacitors used in practice

Figure 2.39 shows various capacitors that are commonly used in practice. The capacitor materials from left: multilayer ceramic, ceramic disc, multilayer polyester film, tubular ceramic, polystyrene, metalized polyester film, aluminum electrolytic. Major scale divisions are in centimeters. Most types of capacitor include a dielectric spacer, which increases their capacitance. These dielectrics are most often insulators. However, low capacitance devices are available with a vacuum between their plates, which allows extremely high voltage operation and low losses. Variable capacitors with their plates open to the atmosphere were commonly used in radio tuning circuits. Later designs use polymer foil dielectric between the moving and stationary plates, with no significant air space between them. In order to maximize the charge that a capacitor can hold, the dielectric material needs to have as high a permittivity as possible, while also having as high a breakdown voltage as possible.

Several solid dielectrics are available, including paper, plastic, glass, mica and ceramic materials. Paper was used extensively in older devices and offers relatively high voltage performance. However, it is susceptible to water absorption, and has been largely replaced by plastic film capacitors. Plastics offer better stability and aging performance, which makes them useful in timer circuits, although they may be limited to low operating temperatures and frequencies. Ceramic capacitors are generally small, cheap and useful for high frequency applications, although their capacitance varies strongly with voltage and they age poorly. They are broadly categorized as **class 1** dielectrics, which have predictable variation of capacitance with temperature or **class 2** dielectrics, which can operate at higher voltage. Glass and mica capacitors are extremely reliable, stable and tolerant to high temperatures and voltages, but are too expensive for most mainstream applications. Electrolytic capacitors and super capacitors are used to store small and larger amounts of energy, respectively, ceramic capacitors are often used in resonators, and parasitic capacitance occurs in circuits wherever the simple conductor-insulator-conductor structure is formed unintentionally by the configuration of the circuit layout.

Electrolytic capacitors use an aluminum or tantalum plate with an oxide dielectric layer. The second electrode is a liquid electrolyte, connected to the circuit by another foil plate. Electrolytic capacitors offer very high capacitance but suffer from poor tolerances, high instability, gradual loss of capacitance especially when subjected to heat, and high leakage current. Poor quality capacitors may leak electrolyte, which is harmful to printed circuit boards. The conductivity of the electrolyte drops at low temperatures, which increases equivalent series resistance. While widely used for power-supply conditioning, poor high-frequency characteristics make them unsuitable for many applications. Electrolytic capacitors will self-degrade if unused for a period (around a year), and when full power is applied may short circuit, permanently damaging the capacitor and usually blowing a fuse or causing arcing. They can be restored before use (and damage) by gradually applying the operating voltage. Unfortunately, the use of this technique may be less satisfactory for some solid state equipment, which may be damaged by operation below its normal power range, requiring that the power supply first be isolated from the consuming circuits. Such remedies may not be applicable to modern high-frequency power supplies as these produce full output voltage even with reduced input.

Tantalum capacitors offer better frequency and temperature characteristics than aluminum, but higher dielectric absorption and leakage. OS-CON (or OC-CON) capacitors are a polymerized organic semiconductor solid-electrolyte type that offer longer life at higher cost than standard electrolytic capacitors. Several other types of capacitor are available for specialist applications. Supercapacitors store large amounts of energy. Supercapacitors made from carbon aero gel, carbon nanotubes, or highly porous electrode materials offer extremely high capacitance (up to 5 kF as of 2010) and can be used in some applications instead of rechargeable batteries. Alternating current capacitors are specifically designed to work on line (mains) voltage AC power circuits. They are commonly used in electric motor circuits and are often designed to handle large currents, so they tend to be physically large. They are usually ruggedly packaged, often in metal cases that can be easily grounded/earthed. They also are designed with direct current breakdown voltages of at least five times the maximum AC voltage.

Structure

Various axial and radial capacitors that are used in practice were shown in Figure 2.39. Figure 2.40 illustrates examples of capacitor packages: SMD ceramic at top left; SMD tantalum at bottom left; through-hole tantalum at top right; through-hole electrolytic at bottom right. Major scale divisions are cm. The arrangement of plates and dielectric has many variations depending on the desired ratings of the capacitor. For small values of capacitance (microfarads and less), ceramic disks use metallic coatings, with wire leads bonded to the coating. Larger values can be made by multiple stacks of plates and disks. Larger value capacitors usually use a metal foil or metal film layer deposited on the surface of a dielectric film to make the plates, and a dielectric film of impregnated paper or plastic – these are rolled up to save space. To reduce the series resistance and inductance for long plates, the plates and dielectric are staggered so that connection is made at the common edge of the rolled-up plates, not at the ends of the foil or metalized film strips that comprise the plates.

The assembly is encased to prevent moisture entering the dielectric – early radio equipment used a cardboard tube sealed with wax. Modern paper or film dielectric capacitors are dipped in a hard thermoplastic. Large capacitors for high-voltage use may have the roll form compressed to fit into a rectangular metal case, with bolted terminals and bushings for connections. The dielectric in larger capacitors is often impregnated with a liquid to improve its properties.

Figure 2. 40 Examples of capacitor packages

Capacitors may have their connecting leads arranged in many configurations, for example axially or radially. "Axial" means that the leads are on a common axis, typically the axis of the capacitor's cylindrical body – the leads extend from opposite ends. Radial leads might more accurately be referred to as tandem; they are rarely actually aligned along radii of the body's circle, so the term is inexact, although universal. The leads (until bent) are usually in planes parallel to that of the flat body of the capacitor, and extend in the same direction; they are often parallel as manufactured.

Small, cheap discoidal ceramic capacitors have existed since the 1930s, and remain in widespread use. Since the 1980s, surface mount packages for capacitors have been widely used. These packages are extremely small and lack connecting leads, allowing them to be soldered directly onto the surface of printed circuit boards. Surface mount components avoid undesirable high-frequency effects due to the leads and simplify automated assembly, although manual handling is made difficult due to their small size.

Mechanically controlled variable capacitors allow the plate spacing to be adjusted, for example by rotating or sliding a set of movable plates into alignment with a set of stationary plates. Low cost variable capacitors squeeze together alternating layers of aluminum and plastic with a screw. Electrical control of capacitance is achievable with varactors (or varicaps), which are reverse-biased semiconductor diodes whose depletion region width varies with applied voltage. They are used in phase-locked loops, amongst other applications.

Capacitor Markings

Most capacitors have numbers printed on their bodies to indicate their electrical characteristics. Larger capacitors like electrolytics usually display the actual capacitance together with the unit (for example, **220 µF**). Smaller capacitors like ceramics, however, use a shorthand consisting of three numbers and a letter, where the numbers show the capacitance in pF (calculated as XY x 10^{Z} for the numbers XYZ) and the letter indicates the tolerance (J, K or M for ±5%, ±10% and ±20% respectively). Additionally, the capacitor may show its working voltage, temperature and other relevant characteristics.

Example

A capacitor with the text **473K 330V** on its body has a capacitance of 47 x 10^{3} pF = 47 nF (±10%) with a working voltage of 330 V.

Applications of Capacitors

Capacitors have many uses in electronic and electrical systems. They are so common that it is a rare electrical product that does not include at least one for some purpose.

Energy Storage

A capacitor can store electric energy when disconnected from its charging circuit, so it can be used like a temporary battery. Capacitors are commonly used in electronic devices to maintain power supply while batteries are being changed. (This prevents loss of information in volatile memory.)

Pulsed Power and Weapons

Groups of large, specially constructed, low-inductance high-voltage capacitors (*capacitor**Error! Bookmark not defined.** banks*) are used to supply huge pulses of current for many pulsed power applications. These include electromagnetic forming, Marx generators, pulsed lasers (especially TEA lasers), pulse forming networks, radar, fusion research, and particle accelerators. Large capacitor banks (reservoir) are used as energy sources for the exploding-bridgewire detonators or slapper detonators in nuclear weapons and other specialty weapons. Experimental work is under way using banks of capacitors as power sources for electromagnetic armour and electromagnetic railguns and coilguns.

Power Conditioning

Reservoir capacitors are used in power supplies where they smooth the output of a full or half wave rectifier. They can also be used in charge pump circuits as the energy storage element in the generation of higher voltages than the input voltage. Figure 2.41 shows A

Figure 2. 41 A reservoir capacitor in an amplifier

10,000 μF capacitor in the power supply section of an amplifier.

Capacitors are connected in parallel with the power circuits of most electronic devices and larger systems (such as factories) to shunt away and conceal current fluctuations from the primary power source to provide a "clean" power supply for signal or control circuits. Audio equipment, for example, uses several capacitors in this way, to shunt away power line hum before it gets into the signal circuitry. The capacitors act as a local reserve for the DC power source, and bypass AC currents from the power supply. This is used in car audio applications, when a stiffening capacitor compensates for the inductance and resistance of the leads to the lead-acid car battery.

Power Factor Correction

In electric power distribution, capacitors are used for power factor correction. Such capacitors often come as three capacitors connected as a three phase load. Usually, the values of these capacitors are given not in farads but rather as a reactive power in volt-amperes reactive (VAr). The purpose is to counteract inductive loading from devices like electric motors and transmission lines to make the load appear to be mostly resistive. Individual motor or lamp loads may have capacitors for power factor correction, or larger sets of capacitors (usually with automatic switching devices) may be installed at a load center within a building or in a large utility substation.

Suppression and Coupling

Signal coupling

Because capacitors pass AC but block DC signals (when charged up to the applied dc voltage), they are often used to separate the AC and DC components of a signal. This method is known as *AC coupling* or "capacitive coupling". Here, a large value of capacitance, whose value need not be accurately controlled, but whose reactance is small at the signal frequency, is employed.

Decoupling

A decoupling capacitor is a capacitor used to protect one part of a circuit from the effect of another, for instance to suppress noise or transients. Noise caused by other circuit elements is shunted through the capacitor, reducing the effect they have on the rest of the circuit. It is most commonly used between the power supply and ground. An alternative name is *bypass capacitor* as it is used to bypass the power supply or other high impedance component of a circuit.

Noise filters and Snubbers

When an inductive circuit is opened, the current through the inductance collapses quickly, creating a large voltage across the open circuit of the switch or relay. If the inductance is large enough, the energy will generate a spark, causing the contact points to oxidize, deteriorate, or sometimes weld together, or destroying a solid-state switch. A snubber capacitor across the newly opened circuit creates a path for this impulse to bypass the contact points, thereby preserving their life; these were commonly found in contact breaker ignition systems, for instance. Similarly, in smaller scale circuits, the spark may not be enough to damage the switch but will still radiate undesirable radio frequency interference (RFI), which a filter capacitor absorbs. Snubber capacitors are usually employed with a low-value resistor in series, to dissipate energy and minimize RFI. Such resistor-capacitor**Error! Bookmark not defined.** combinations are available in a single package.

Capacitors are also used in parallel to interrupt units of a high-voltage circuit breaker in order to equally distribute the voltage between these units. In this case they are called grading capacitors.

In schematic diagrams, a capacitor used primarily for DC charge storage is often drawn vertically in circuit diagrams with the lower, more negative, plate drawn as an arc. The straight plate indicates the positive terminal of the device, if it is polarized.

Motor Starters

In single phase squirrel cage motors, the primary winding within the motor housing is not capable of starting a rotational motion on the rotor, but is capable of sustaining one. To start the motor, a secondary winding is used in series with a non-polarized *starting capacitor* to introduce a lag in the sinusoidal current through the starting winding. When the secondary winding is placed at an angle with respect to the primary winding, a rotating electric field is created. The force of the rotational field is not constant, but is sufficient to start the rotor spinning. When the rotor comes close to operating speed, a centrifugal switch (or current-sensitive relay in series with the main winding) disconnects the capacitor. The start capacitor is typically mounted to the side of the motor housing. These are called capacitor-start motors, that have relatively high starting torque.

There are also capacitor-run induction motors which have a permanently connected phase-shifting capacitor**Error! Bookmark not defined.** in series with a second winding. The motor is much like a two-phase induction motor.

Motor-starting capacitors are typically non-polarized electrolytic types, while running capacitors are conventional paper or plastic film dielectric types.

Signal Processing

The energy stored in a capacitor can be used to represent information, either in binary form, as in DRAMs, or in analogue form, as in analog sampled filters and Charge Coupled Devices (CCDs). Capacitors can be used in analog circuits as components of integrators or more complex filters and in negative feedback loop stabilization. Signal processing circuits also use capacitors to integrate a current signal.

Tuned Circuits

Capacitors and inductors are applied together in tuned circuits to select information in particular frequency bands. For example, radio receivers rely on variable capacitors to tune the station frequency. Speakers use passive analog crossovers, and analog equalizers use capacitors to select different audio bands. The resonant frequency *f* of a tuned circuit is a function of the inductance (*L*) and capacitance (*C*) in series, and is given by: $f = \frac{1}{2\pi\sqrt{LC}}$ where *L* is in henries and *C* is in farads.

Capacitive Sensing

The simplest capacitor consists of two parallel conductive plates separated by a dielectric of thickness *d* with permittivity *ε* (such as air) as illustrated in Figure 2.42. The model may also be used to make qualitative predictions for other device geometries. The plates are considered to extend uniformly over an area *A.* The capacitance is expressed as: $C = \frac{\varepsilon A}{d}$.

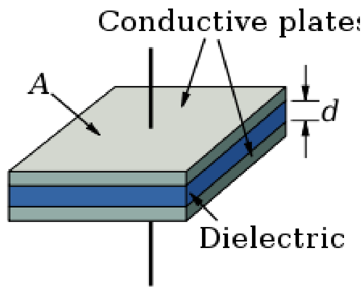


Figure 2. 42 A simple capacitor

Most capacitors are designed to maintain a fixed physical structure.

However, various factors can change the structure of the capacitor**Error! Bookmark not defined.**, and the resulting change in capacitance can be used to sense those factors.

Changing the Dielectric

The effects of varying the physical and/or electrical characteristics of the **dielectric** can be used for sensing purposes. Capacitors with an exposed and porous dielectric can be used to measure humidity in air. Capacitors are used to accurately measure the fuel level in airplanes; as the fuel covers more of a pair of plates, the circuit capacitance increases.

Changing the Distance Between the Plates

Capacitors with a flexible plate can be used to measure strain or pressure. Industrial pressure transmitters used for process control use pressure-sensing diaphragms, which form a capacitor plate of an oscillator circuit. Capacitors are used as the sensor in condenser microphones, where one plate is moved by air pressure, relative to the fixed position of the other plate. Some accelerometers use MEMS capacitors etched on a chip to measure the magnitude and direction of the acceleration vector. They are used to detect changes in acceleration, e.g. as tilt sensors or to detect free fall, as sensors triggering airbag deployment, and in many other applications. Some fingerprint sensors use capacitors. Additionally, a user can adjust the pitch of a theremin musical instrument by moving his hand since this changes the effective capacitance between the user's hand and the antenna.

Changing the Effective Area of the Plates

Capacitive touch switches are now used on many consumer electronic products.

Hazards and Safety

Capacitors may retain a charge long after power is removed from a circuit; this charge can cause dangerous or even potentially fatal shocks or damage connected equipment. For example, even a seemingly innocuous device such as a disposable camera flash unit powered by a 1.5 volt AA battery contains a capacitor which may be charged to over 300 volts. This is easily capable of delivering a shock. Service procedures for electronic devices usually include instructions to discharge large or high-voltage capacitors. Capacitors may also have built-in discharge resistors to dissipate stored energy to a safe level within a few seconds after power is removed. High-voltage capacitors are stored with the terminals shorted, as protection from potentially dangerous voltages due to dielectric absorption.

Some old, large oil-filled capacitors contain polychlorinated biphenyls (PCBs). It is known that waste PCBs can leak into groundwater under landfills. Capacitors containing PCB were labeled as containing "Askarel" and several other trade names. PCB-filled capacitors are found in very old (pre 1975) fluorescent lamp ballasts, and other applications.

High-voltage capacitors may catastrophically fail when subjected to voltages or currents beyond their rating, or as they reach their normal end of life. Dielectric or metal interconnection failures may create arcing that vaporizes dielectric fluid, resulting in case bulging, rupture, or even an explosion. Capacitors used in RF or sustained high-current applications can overheat, especially in the center of the capacitor rolls. Capacitors used within high-energy capacitor banks can violently explode when a short in one capacitor causes sudden dumping of energy stored in the rest of the bank into the failing unit. High voltage vacuum capacitors can generate soft X-rays even during normal operation. Proper containment, fusing, and preventive maintenance can help to minimize these hazards.

High-voltage capacitors can benefit from a pre-charge to limit in-rush currents at power-up of high voltage direct current (HVDC) circuits. This will extend the life of the component and may mitigate high-voltage hazards.

Supercapacitors - Electric Double-Layer Capacitors

An electric double-layer capacitor (EDLC), also known as supercapacitor, supercondenser, pseudocapacitor, electrochemical double layer capacitor, or ultracapacitor, is an electrochemical capacitor with relatively high energy density. Compared to conventional electrolytic capacitors the energy density is typically on the order of thousands of times greater. In comparison with conventional batteries or fuel cells, EDLCs also have a much higher power density.

A typical D-cell sized electrolytic capacitor displays capacitance in the range of tens of millifarads. The same size EDLC might reach several farads, an improvement of two orders of magnitude. EDLCs usually yield a lower working voltage; as of 2010 larger double-layer capacitors have capacities up to 5,000 farads.

EDLCs have a variety of commercial applications, notably in "energy smoothing" and momentary-load devices. They have applications as energy-storage devices used in vehicles, and for smaller applications like home solar energy systems where extremely fast charging is a valuable feature.

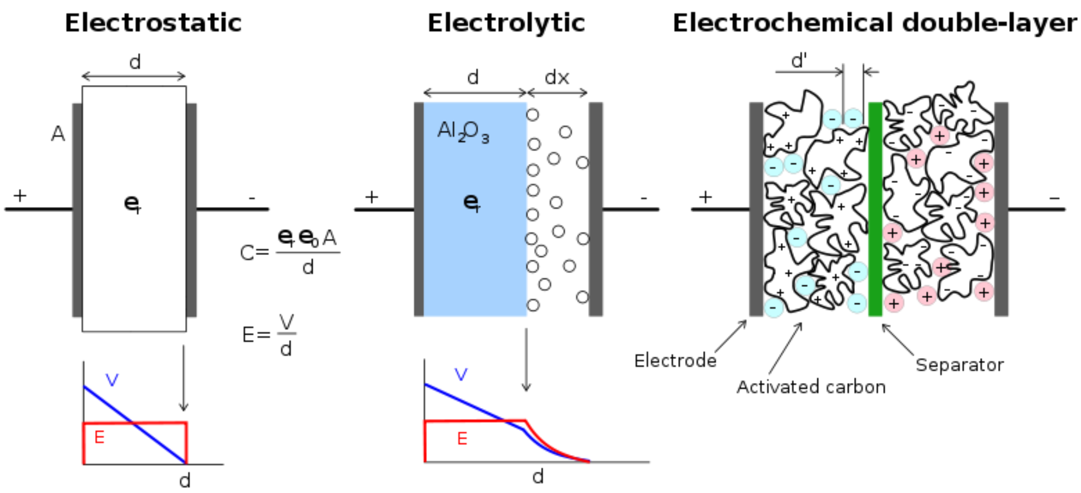


Figure 2. 43 Comparison of capacitors

Figure 2.43 shows a diagram comparing construction of three types of capacitors: electrostatic (normal), electrolytic (high capacity) and electrochemical (supercapacitors). In a conventional capacitor, energy is stored by the removal of charge carriers, typically electrons, from one metal plate and depositing them on another. This charge separation creates a potential between the two plates, which can be harnessed in an external circuit. The total energy stored in this fashion is proportional to both the amount of charge stored and the potential between the plates. The amount of charge stored per unit voltage is essentially a function of the size, the distance, and the material properties of the plates and the material in between the plates (the dielectric), while the potential between the plates is limited by breakdown of the dielectric. The dielectric controls the capacitor's voltage. Optimizing the material leads to higher energy density for a given size of capacitor.

EDLCs do not have a conventional dielectric. Rather than two separate plates separated by an intervening substance, these capacitors use "plates" that are in fact two layers of the same substrate, and their electrical properties, the so-called "electrical double layer", result in the effective separation of charge despite the vanishingly thin (on the order of nanometers) physical separation of the layers. The lack of need for a bulky layer of dielectric permits the packing of plates with much larger surface area into a given size, resulting in high capacitances in practical-sized packages.

In an electrical double layer, each layer by itself is quite conductive, but the physics at the interface where the layers are effectively in contact means that no significant current can flow between the layers. However, the double layer can withstand only a low voltage, which means that electric double-layer capacitors rated for higher voltages must be made of matched series-connected individual EDLCs, much like series-connected cells in higher-voltage batteries.

EDLCs have much higher power density than batteries. Power density combines the energy density with the speed that the energy can be delivered to the load. Batteries, which are based on the movement of charge carriers in a liquid electrolyte, have relatively slow charge and discharge times. Capacitors, on the other hand, can be charged or discharged at a rate that is typically limited by current heating of the electrodes. So while existing EDLCs have *energy* densities that are perhaps 1/10th that of a conventional battery, their *power* density is generally 10 to 100 times as great.

INDUCTORS

Definition and Use

The dual of the capacitor is the inductor, which stores energy in the magnetic field rather than the electric field. Its current-voltage relation is obtained by exchanging current and voltage in the capacitor equations and replacing *C* with the inductance *L*.

An **inductor** or a **reactor** is a passive electrical component that can store energy in a magnetic field created by the electric current passing through it. An inductor's ability to store magnetic energy is measured by its inductance, in units of henries. Typically an inductor is a conducting wire shaped as a coil; the loops help to create a strong magnetic field inside the coil due to Ampere's Law. Due to the time-varying magnetic field inside the coil, a voltage is induced, according to Faraday's law of electromagnetic induction, which by Lenz's Law opposes the change in current that created it. Inductors are one of the basic components used in electronics where current and voltage change with time, due to the ability of inductors to delay and reshape alternating currents. Inductors called chokes are used as parts of filters in power supplies or to block AC signals from passing through a circuit.

Overview

Inductance (*L*) results from the magnetic field forming around a current-carrying conductor which tends to resist changes in the current. Electric current through the conductor creates a magnetic flux proportional to the current, and a change in this current creates a corresponding change in magnetic flux which, in turn, by Faraday's Law generates an electromotive force (EMF) that opposes this change in current. Inductance is a measure of the amount of EMF generated per unit change in current. For example, an inductor with an inductance of 1 Henry produces an EMF of 1 volt when the current through the inductor changes at the rate of 1 ampere per second. The number of loops, the size of each loop, and the material it is wrapped around all affect the inductance. For example, the magnetic flux linking these turns can be increased by coiling the conductor around a material with a high permeability such as iron. This can increase the inductance by 2000 times.

Ideal and Real Inductors

An "ideal inductor" has inductance, but no resistance or capacitance, and does not dissipate or radiate energy. A real inductor may be partially modeled by a combination of inductance, resistance (due to the resistance of the wire and losses in core material), and capacitance. At some frequency, some real inductors behave as resonant circuits (due to their self capacitance). At some frequency the capacitive component of impedance becomes dominant. Energy is dissipated by the resistance of the wire, and by any losses in the magnetic core due to hysteresis. Practical iron-core inductors at high currents show gradual departure from ideal behavior due to nonlinearity caused by magnetic saturation. At higher frequencies, resistance and resistive losses in inductors grow due to skin effect in the inductor's winding wires. Core losses also contribute to inductor losses at higher frequencies. Practical inductors work as antennas, radiating a part of energy processed into surrounding space and circuits, and accepting electromagnetic emissions from other circuits, taking part in electromagnetic interference. Circuits and materials close to the inductor will have near-field coupling to the inductor's magnetic field, which may cause additional energy loss. Real-world inductor applications may consider the parasitic parameters as important as the inductance.

Applications of Inductors

Figure 2.44 shows an inductor with two 47mH windings, as may be found in a power supply. Inductors are used extensively in analog circuits and signal processing. Inductors in conjunction with capacitors and other components form tuned circuits which can emphasize or filter out specific signal frequencies. Applications range from the use of large inductors in power supplies, which in conjunction with filter capacitors remove residual hums known as the mains hum or other fluctuations from the direct current output, to the small inductance of the ferrite bead or torus installed around a cable to prevent radio frequency interference from being transmitted down the wire. Smaller inductor/capacitor combinations provide tuned circuits used in radio reception and broadcasting, for instance.

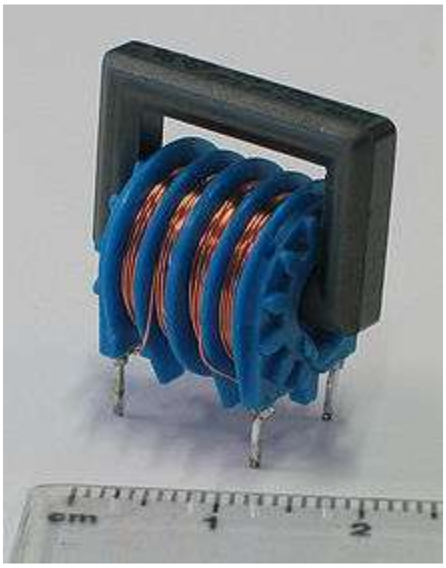


Figure 2. 44 A simple inductor

Two (or more) inductors that have coupled magnetic flux form a transformer, which is a fundamental component of every electric utility power grid. The efficiency of a transformer may decrease as the frequency increases due to eddy currents in the core material and skin effect on the windings. Size of the core can be decreased at higher frequencies and, for this reason, aircraft use 400 hertz alternating current rather than the usual 50 or 60 hertz, allowing a great saving in weight from the use of smaller transformers.

An inductor is used as the energy storage device in some switched-mode power supplies. The inductor is energized for a specific fraction of the regulator's switching frequency, and de-energized for the remainder of the cycle. This energy transfer ratio determines the input-voltage to output-voltage ratio. This X_L is used in complement with an active semiconductor device to maintain very accurate voltage control.

Inductors are also employed in electrical transmission systems, where they are used to depress voltages from lightning strikes and to limit switching currents and fault current. In this field, they are more commonly referred to as reactors. Larger value inductors may be simulated by use of gyrator circuits.

Inductor Construction

An inductor is usually constructed as a coil of conducting material, typically copper wire, wrapped around a core either of air or of ferromagnetic or ferromagnetic material. Core materials with a higher permeability than air increase the magnetic field and confine it closely to the inductor, thereby increasing the inductance. Low frequency inductors are constructed like transformers, with cores of electrical steel laminated to prevent eddy currents. 'Soft' ferrites are widely used for cores above audio frequencies, since they do not cause the large energy losses at high frequencies that ordinary iron alloys do. Inductors come in many shapes as illustrated in Figure 2.45. Most are constructed as enamel coated wire (magnet wire) wrapped around a ferrite bobbin with wire exposed on the outside, while some enclose the wire completely in ferrite and are referred to as "shielded". Some inductors have an adjustable core, which enables changing of the inductance. Inductors used to block very high frequencies are sometimes made by stringing a ferrite cylinder or bead on a wire.

Figure 2. 45 Types of inductors

Small inductors can be etched directly onto a printed circuit board by laying out the trace in a spiral pattern. Some such planar inductors use a planar core. Small value inductors can also be built on integrated circuits using the same processes that are used to make transistors. Aluminum interconnect is typically used, laid out in a spiral coil pattern. However, the small dimensions limit the inductance, and it is far more common to use a circuit called a "gyrator" that uses a capacitor and active components to behave similarly to an inductor.

Types of Inductors

Air Core Coil

The term *air core coil* describes an inductor that does not use a magnetic core made of a ferromagnetic material. The term refers to coils wound on plastic, ceramic, or other nonmagnetic forms, as well as those that actually have air inside the windings. Air core coils have lower inductance than ferromagnetic core coils, but are often used at high frequencies because they are free from energy losses called core losses that occur in ferromagnetic cores, which increase with frequency. A side effect that can occur in air core coils in which the winding is not rigidly supported on a form is 'microphony': mechanical vibration of the windings can cause variations in the inductance.

Radio Frequency Inductors

At high frequencies, particularly radio frequencies (RF), inductors have higher resistance and other losses. In addition to causing power loss, in resonant circuits this can reduce the Q factor of the circuit, broadening the bandwidth. In RF inductors, which are mostly air core types, specialized construction techniques are used to minimize these losses. The losses are due to these effects:

Skin effect: The resistance of a wire to high frequency current is higher than its resistance to direct current because of skin effect. Radio frequency alternating current does not penetrate far into the body of a conductor but travels along its surface. Therefore, in a solid wire, most of the cross sectional area of the wire is not used to conduct the current, which is in a narrow annulus on the surface. This effect increases the resistance of the wire in the coil, which may already have a relatively high resistance due to its length and small diameter.

Proximity effect: Another similar effect that also increases the resistance of the wire at high frequencies is proximity effect, which occurs in parallel wires that lie close to each other. The individual magnetic field of adjacent turns induces eddy currents in the wire of the coil, which causes the current in the conductor to be concentrated in a thin strip on the side near the adjacent wire. Like skin effect, this reduces the effective cross-sectional area of the wire conducting current, increasing its resistance.

Parasitic capacitance: The capacitance between individual wire turns of the coil, called parasitic capacitance, does not cause energy losses but can change the behavior of the coil. Each turn of the coil is at a slightly different potential, so the electric field between neighboring turns stores charge on the wire. So the coil acts as if it has a capacitor in parallel with it. At a high enough frequency this capacitance can resonate with the inductance of the coil forming a tuned circuit, causing the coil to become self-resonant.

To reduce parasitic capacitance and proximity effect, RF coils are constructed to avoid having many turns lying close together, parallel to one another. The windings of RF coils are often limited to a single layer, and the turns are spaced apart. To reduce resistance due to skin effect, in high-power inductors such as those used in transmitters the windings are sometimes made of a metal strip or tubing which has a larger surface area, and the surface is silver-plated.

Honeycomb coils: To reduce proximity effect and parasitic capacitance, multilayer RF coils are wound in patterns in which successive turns are not parallel but crisscrossed at an angle; these are often called *honeycomb* or *basket-weave* coils.

Spiderweb coils: Another construction technique with similar advantages is flat spiral coils. These are often wound on a flat insulating support with radial spokes or slots, with the wire weaving in and out through the slots; these are called *spiderweb* coils. The form has an odd number of slots, so successive turns of the spiral lie on opposite sides of the form, increasing separation.

Litz wire: To reduce skin effect losses, some coils are wound with a special type of radio frequency wire called litz wire. Instead of a single solid conductor, litz wire consists of several smaller wire strands that carry the current. Unlike ordinary stranded wire, the strands are insulated from each other, to prevent skin effect from forcing the current to the surface, and are braided together. The braid pattern ensures that each wire strand spends the same amount of its length on the outside of the braid, so skin effect distributes the current equally between the strands, resulting in a larger cross-sectional conduction area than an equivalent single wire.

Ferromagnetic Core Coil

Ferromagnetic-core or iron-core inductors use a magnetic core made of a ferromagnetic or ferrimagnetic material such as iron or ferrite to increase the inductance. A magnetic core can increase the inductance of a coil by a factor of several thousand, by increasing the magnetic field due to its higher magnetic permeability. However the magnetic properties of the core material cause several side effects which alter the behavior of the inductor and require special construction:

Core losses: A time-varying current in a ferromagnetic inductor, which causes a time-varying magnetic field in its core, causes energy losses in the core material that are dissipated as heat, due to two processes:

Eddy currents: From Faraday's law of induction, the changing magnetic field can induce circulating loops of electric current in the conductive metal core. The energy in these currents is dissipated as heat in the resistance of the core material. The amount of energy lost increases with the area inside the loop of current.

Hysteresis: Changing or reversing the magnetic field in the core also causes losses due to the motion of the tiny magnetic domains it is composed of. The energy loss is proportional to the area of the hysteresis loop in the BH graph of the core material. Materials with low coercivity have narrow hysteresis loops and so low hysteresis losses.

For both of these processes, the energy loss per cycle of alternating current is constant, so core losses increase linearly with frequency.

Nonlinearity: If the current through a ferromagnetic core coil is high enough that the magnetic core saturates, the inductance will not remain constant but will change with the current through the device. This is called nonlinearity and results in distortion of the signal. For example, audio signals can suffer intermodulation distortion in saturated inductors. To prevent this, in linear circuits the current through iron core inductors must be limited below the saturation level. Using a powdered iron core with a distributed air gap allows higher levels of magnetic flux which in turn allows a higher level of direct current through the inductor before it saturates.

Laminated Core Inductor

Low-frequency inductors are often made with laminated cores to prevent eddy currents, using construction similar to transformers. The core is made of stacks of thin steel sheets or laminations oriented parallel to the field, with an insulating coating on the surface. The insulation prevents eddy currents between the sheets, so any remaining currents must be within the cross sectional area of the individual laminations, reducing the area of the loop and thus the energy loss greatly. The laminations are made of low-coercivity silicon steel, to reduce hysteresis losses.

Ferrite-Core Inductor

For higher frequencies, inductors are made with cores of ferrite. Ferrite is a ceramic ferrimagnetic material that is nonconductive, so eddy currents cannot flow within it. The formulation of ferrite is $xxFe_2O_4$ where xx represents various metals. For inductor cores soft ferrites are used, which have low coercivity and thus low hysteresis losses. Another similar material is powdered iron cemented with a binder.

Toroidal Core Coils

In an inductor wound on a straight rod-shaped core, the magnetic field lines emerging from one end of the core must pass through the air to reenter the core at the other end. This reduces the field, because much of the magnetic field path is in air rather than the higher permeability core material. A higher magnetic field and inductance can be achieved by forming the core in a closed magnetic circuit. The magnetic field lines form closed loops within the core without leaving the core material. The shape often used is a toroidal or doughnut-shaped ferrite core. Because of their symmetry, toroidal cores allow a minimum of the magnetic flux to escape outside the core (called *leakage flux*), so they radiate less electromagnetic interference than other shapes. Toroidal core coils are manufactured of various materials, primarily ferrite, Kool Mu MPP, powdered iron and laminated cores.

Variable Inductor

A variable inductor can be constructed by making one of the terminals of the device a sliding spring contact that can move along the surface of the coil, increasing or decreasing the number of turns of the coil included in the circuit. An alternate construction method is to use a moveable magnetic core, which can be slid in or out of the coil. Moving the core farther into the coil increases the permeability, increasing the inductance. Many inductors used in radio applications (usually less than 100 MHz) use adjustable cores in order to tune such inductors to their desired value, since manufacturing processes have certain tolerances (inaccuracy).

Inductors in Electric Circuits

Current and Voltage Relations

The effect of an inductor in a circuit is to oppose changes in current through it by developing a voltage across it proportional to the rate of change of the current. An ideal inductor would offer no resistance to a constant direct current; however, only superconducting inductors have truly zero electrical resistance.

The relationship between the time-varying voltage $v(t)$ across an inductor with inductance L and the time-varying current $i(t)$ passing through it is described by the differential equation: $v(t) = L\frac{di(t)}{dt}$

When there is a sinusoidal alternating current (AC) through an inductor, a sinusoidal voltage is induced. The amplitude of the voltage is proportional to the product of the amplitude (I_P) of the current and the frequency (f) of the current.

$$i(t) = I_P \sin(2\pi f t);\ \frac{di(t)}{dt} = s\pi f I_P \cos(2\pi f t);\ v(t) = 2\pi f L I_P \cos(2\pi f t)$$

In this situation, the phase of the current lags that of the voltage by $\pi/2$.

If an inductor is connected to a direct current source with value I via a resistance R, and then the current source is short-circuited, the differential relationship above shows that the current through the inductor will discharge with an exponential decay: $i(t) = Ie^{-\left(\frac{R}{L}\right)t}$

Stored Energy

The energy (measured in joules, in SI) stored by an inductor is equal to the amount of work required to establish the current through the inductor, and therefore the magnetic field. This is given by: $E_{stored} = \frac{1}{2}LI^2$

where *L* is inductance and *I* is the current through the inductor.

This relationship is only valid for linear (non-saturated) regions of the magnetic flux linkage and current relationship.

Q Factor

An ideal inductor will be lossless irrespective of the amount of current through the winding. However, typically inductors have winding resistance from the metal wire forming the coils. Since the winding resistance appears as a resistance in series with the inductor, it is often called the *series resistance*. The inductor's series resistance converts electric current through the coils into heat, thus causing a loss of inductive quality. The quality factor (or *Q*) of an inductor is the ratio of its inductive reactance to its resistance at a given frequency, and is a measure of its efficiency. The higher the Q factor of the inductor, the closer it approaches the behavior of an ideal, lossless, inductor.

The Q factor of an inductor can be found through the following formula, where *R* is its internal electrical resistance and ω*L* is capacitive or inductive reactance at resonance: $Q = \frac{\omega L}{R}$

By using a ferromagnetic core, the inductance is greatly increased for the same amount of copper, multiplying up the Q. Cores however also introduce losses that increase with frequency. A grade of core material is chosen for best results for the frequency band. At VHF or higher frequencies an air core is likely to be used.

Inductors wound around a ferromagnetic core may saturate at high currents, causing a dramatic decrease in inductance (and Q). This phenomenon can be avoided by using a (physically larger) air core inductor. A well designed air core inductor may have a Q of several hundred.

An almost ideal inductor (Q approaching infinity) can be created by immersing a coil made from a superconducting alloy in liquid helium or liquid nitrogen. This supercools the wire, causing its winding resistance to disappear. Because a superconducting inductor is virtually lossless, it can store a large amount of electrical energy within the surrounding magnetic field. Bear in mind that for inductors with cores, core losses still exist.

TRANSFORMER

Definition and Use

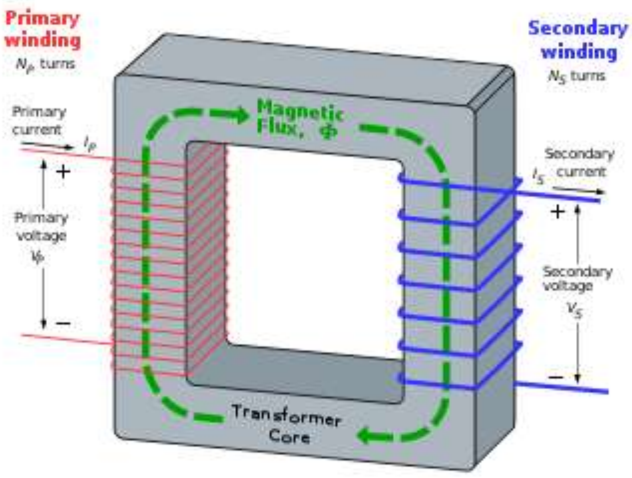


Figure 2. 46 A trnasformer

A transformer is a device that transfers electrical energy from one circuit to another through inductively coupled conductors—the transformer's coils as illustrated in Figure 2.46. A varying current in the first or primary winding creates a varying magnetic flux in the transformer's core and thus a varying magnetic field through the secondary winding. This varying magnetic field induces a varying electromotive force (EMF), or "voltage", in the secondary winding. This effect is called mutual induction.

In the vast majority of transformers, the windings are coils wound around a ferromagnetic core, air-core transformers being a notable exception. Transformers range in size from a thumbnail-sized coupling transformer hidden inside a stage microphone to huge units weighing hundreds of tons used to interconnect portions of power grids. All operate with the same basic principles, although the range of designs is wide. While new technologies have eliminated the need for transformers in some electronic circuits, transformers are still found in nearly all electronic devices designed for household ("mains") voltage. They are also used extensively in electronic products to step down the supply voltage to a level suitable for the low voltage circuits they contain. The transformer also electrically isolates the end user from contact with the supply voltage.Transformers are essential for high-voltage electric power transmission, which makes long-distance transmission economically practical.

The Ideal Transformer as a Circuit Element

If a load is connected to the secondary, an electric current will flow in the secondary winding and electrical energy will be transferred from the primary circuit through the transformer to the load. In an ideal transformer, the induced voltage in the secondary winding (*Vs*) is in proportion to the primary voltage (*Vp*), and is given by the ratio of the number of turns in the secondary (*Ns*) to the number of turns in the primary (*Np*). By appropriate selection of the ratio of turns, a transformer thus allows an alternating current (AC) voltage to be "stepped up" by making *Ns* greater than *Np*, or "stepped down" by making *Ns* less than *Np*. Ideally, the transformer is perfectly efficient; all the incoming energy is transformed from the primary circuit to the magnetic field and into the secondary circuit. If this condition is met, the incoming electric power must equal the outgoing power:

$$P_{incoming} = I_p V_p = P_{outgoing} = I_s V_s$$

giving the ideal transformer equation

$$\frac{V_s}{V_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s}$$

Transformers normally have high efficiency, so this formula is a reasonable approximation.

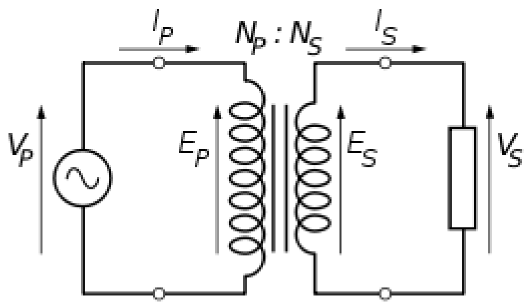


Figure 2. 47 A transformer as a circuit element

If the voltage is increased, then the current is decreased by the same factor. The impedance in one circuit is transformed by the *square* of the turns ratio. For example, if an impedance Z_s is attached across the terminals of the secondary coil, it appears to the primary circuit to have an impedance of $(N_p/N_s)^2 Z_s$. This relationship is reciprocal, so that the impedance Z_p of the primary circuit appears to the secondary to be $(N_s/N_p)^2 Z_p$.

Operation and Practical Considerations

The simplified description above neglects several practical factors, in particular the primary current required to establish a magnetic field in the core, and the contribution to the field due to current in the secondary circuit.

Leakage Flux of a Transformer

The ideal transformer model assumes that all flux generated by the primary winding links all the turns of every winding, including itself. In practice, some flux traverses paths that take it outside the windings as shown in Figure 2.48. Such flux is termed *leakage flux*, and results in leakage inductance in series with the mutually coupled transformer windings. Leakage results in energy being alternately stored in and discharged from the magnetic fields with each cycle of the power supply. It is not directly a power loss (see "Stray losses" below), but results in inferior voltage regulation, causing the secondary voltage to fail to be directly proportional to the primary, particularly under heavy load. Transformers are therefore normally designed to have very low leakage inductance.

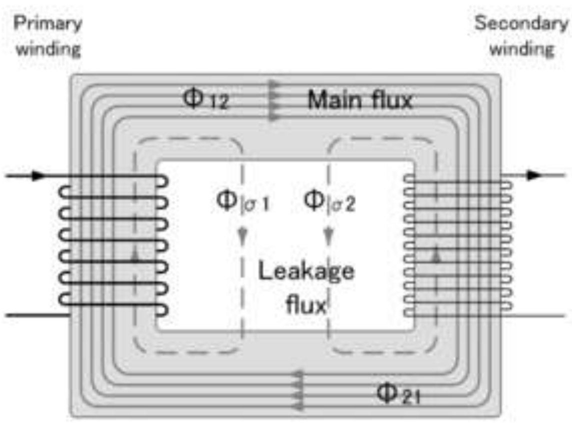


Figure 2. 48Leakage flux of a transformer

However, in some applications, leakage can be a desirable property, and long magnetic paths, air gaps, or magnetic bypass shunts may be deliberately introduced to a transformer's design to limit the short-circuit current it will supply. Leaky transformers may be used to supply loads that exhibit negative resistance, such as electric arcs, mercury vapor lamps, and neon signs; or for safely handling loads that become periodically short-circuited such as electric arc welders.

Effect of Frequency

The EMF of a transformer at a given flux density increases with frequency. By operating at higher frequencies, transformers can be physically more compact because a given core is able to transfer more power without reaching saturation and fewer turns are needed to achieve the same impedance. However, properties such as core loss and conductor skin effect also increase with frequency. Aircraft and military equipment employ 400 Hz power supplies which reduce core and winding weight. Conversely, frequencies used for some railway electrification systems were much lower (e.g. 16.7 Hz and 25 Hz) than normal utility frequencies (50 – 60 Hz) for historical reasons concerned mainly with the limitations of early electric traction motors. As such, the transformers used to step down the high over-head line voltages (e.g. 15 kV) are much heavier for the same power rating than those designed only for the higher frequencies.

Operation of a transformer at its designed voltage but at a higher frequency than intended will lead to reduced magnetizing current; at lower frequency, the magnetizing current will increase. Operation of a transformer at other than its design frequency may require assessment of voltages, losses, and cooling to establish if safe operation is practical. For example, transformers may need to be equipped with "volts per hertz" over-excitation relays to protect the transformer from overvoltage at higher than rated frequency. Knowledge of natural frequencies of transformer windings is of importance for the determination of the transient response of the windings to impulse and switching surge voltages.

Energy Losses

An ideal transformer would have no energy losses, and would be 100% efficient. In practical transformers energy is dissipated in the windings, core, and surrounding structures. Larger transformers are generally more efficient, and those rated for electricity distribution usually perform better than 98%. Experimental transformers using superconducting windings achieve efficiencies of 99.85%. The increase in efficiency can save considerable energy, and hence money, in a large heavily-loaded transformer; the trade-off is in the additional initial and running cost of the superconducting design.

Transformer losses are divided into losses in the windings, termed copper loss, and those in the magnetic circuit, termed iron loss. Losses in the transformer arise from:

- **Winding resistance**: Current flowing through the windings causes resistive heating of the conductors. At higher frequencies, skin effect and proximity effect create additional winding resistance and losses.

- **Hysteresis losses**: Each time the magnetic field is reversed, a small amount of energy is lost due to hysteresis within the core. For a given core material, the loss is proportional to the frequency, and is a function of the peak flux density to which it is subjected.

- **Eddy currents**: Ferromagnetic materials are also good conductors, and a core made from such a material also constitutes a single short-circuited turn throughout its entire length. Eddy currents therefore circulate within the core in a plane normal to the flux, and are responsible for resistive heating of the core material. The eddy current loss is a complex function of the square of supply frequency and inverse square of the material thickness. Eddy current losses can be reduced by making the core of a stack of plates electrically insulated from each other, rather than a solid block; all transformers operating at low frequencies use laminated or similar cores as shown in Figure 2.49.

Figure 2. 49 A transformer with laminated steel core

- **Magnetostriction**: Magnetic flux in a ferromagnetic material, such as the core, causes it to physically expand and contract slightly with each cycle of the magnetic field, an effect known as magnetostriction. This produces the buzzing sound commonly associated with transformers, and can cause losses due to frictional heating.

- **Mechanical losses**: In addition to magnetostriction, the alternating magnetic field causes fluctuating forces between the primary and secondary windings. These incite vibrations within nearby metalwork, adding to the buzzing noise, and consuming a small amount of power.

- **Stray losses**: Leakage inductance is by itself largely lossless, since energy supplied to its magnetic fields is returned to the supply with the next half-cycle. However, any leakage flux that intercepts nearby conductive materials such as the transformer's support structure will give rise to eddy currents and be converted to heat. There are also radiative losses due to the oscillating magnetic field, but these are usually small.

PROBLEMS

Review Questions

1. What are the subatomic particles that contribute to the electrical activities within an atom?
2. What do you understand from energy of an orbit for an electron?
3. What generates the electrical field?
4. What is the relationship between the electrical filed and electrical potential?
5. Define electrical conduction and electrical current.
6. What generates the magnetic field?
7. Describe the effect of an external magnetic field on a current carrying conductor.
8. What is the electromagnetism?
9. How an electric arc is generated and what is the spark gap?
10. How electrical energy is generated from fossil fuels and renewable sources?
11. Why the electrical energy is preferred over other forms of energies overwhelmingly?
12. Define in precise terms conductors, semiconductors and insulators.
13. What is a superconductor and how it is generated?
14. Why the elements named as "conductors" conduct electricity easily?
15. What are the three best conductors?
16. Why copper is the mostly used conductor?
17. Why the bare copper wire is not used (why it is used with some sort of covering/coating)?
18. Why a stranded wire is preferred to solid core wire?
19. Why we use twisted pairs of wires?
20. What is a transmission line and how it differs from an ordinary wire?
21. Why we use shielded wires?
22. Why we use constant spacing between pairs of signal wires?
23. Why we don't use thick solid conductors at high frequency AC applications?
24. What is the wire gage and how it is used to select the wire size for a given application?
25. What determines the current carrying capacity (ampacity) of a wire conductor?
26. Express the resistance of a wire in terms of its length and diameter.
27. What is the meaning of "positive temperature coefficient" for a resistive wire?
28. How is the flow (current) through a resistance is related to the effort (voltage) applied?
29. What is the difference between a potentiometer and a rheostat?
30. What is the difference between a carbon composition resistor and a carbon film resistor? What are the advantages and limitations of both types?
31. State the advantages of metal film resistors over carbon composition and film resistors.
32. What is the a wire-wound resistor and how the inductive effect is minimized?

33. State a few resistive sensors with their areas of applications.
34. Illustrate the markings for a four-band resistor with an example.
35. How the markings for low-value resistors differ from the regular ones?
36. Illustrate the markings for a five-band resistor with an example.
37. What are the differences in identification of the value of a resistor between a four-band and a five-band marking?
38. What is an SMD resistor and how it is identified?
39. List the preferred values of resistors in one decade for E12 and E24 series.
40. What is a heat sink and how it improves the power rating of a resistor?
41. How an axial resistor behave at high frequencies?
42. Why the resistors generate noise and which types are preferred in preamplifiers?
43. What are the failure modes for resistors?
44. What is the function a capacitor?
45. How is the flow (current) through a capacitor is related to the effort (voltage) applied?
46. Explain the behavior of a capacitor in AC and DC circuits.
47. How the electrolytic and non-electrolytic capacitors differ from each other?
48. State the non-ideal behaviors of capacitors.
49. Explain the effect of the dielectric on the performance of the capacitor.
50. What is the breakdown voltage and how effective it is in choosing a capacitor for a specific application?
51. What is the ripple current?
52. Describe the capacitor marking commonly used in identifying the capacitors with examples.
53. List applications of capacitors.
54. Explain the terms "signal coupling" and "decoupling" and the function of the capacitors in achieving them.
55. State a few capacitive sensors with their areas of applications.
56. How you can select the proper capacitor for a given application?
57. What are the hazards related to capacitors and required safety measures?
58. What is a supercapacitors and how it differs from a regular electrolytic capacitor?
59. What is the function an inductor?
60. How is the flow (current) through an inductor is related to the effort (voltage) applied?
61. Explain the behavior of an inductor in AC and DC circuits.
62. What are the salient features of radio frequency inductors?
63. State the non-ideal behaviors of inductors.
64. Explain the effect of the core on the performance of an inductor.

65. What is the Q factor of an inductor?
66. What basic function a transformer performs in electrical circuits?
67. Explain the behavior of a transformer in low frequency and high frequency applications.
68. How a practical transformer differs from the ideal one?
69. What is the efficiency of a transformer?

General Questions

1. No. 14 gage copper wire is used for house wiring. It's weight is 18.5 gram/meter. It's resistance is 0.00827 Ω/m at 20 °C. The temperature coefficient of copper is 0.004 /°C.
    a. What will be the resistance of 10 m wire at 20 °C and at 60 °C
    b. How much is the voltage drop across the wire in the above question is the current is 4 A at 20 °C and at 60 °C
    c. Assume that the wire was warming up by 2 °C as the current through it was 1 A. How much is the maximum current allowed if the plastic covering melts at 60 °C?

BIBLIOGRAPHY

Further Reading

Useful Websites

(Visited February 23, 2011)

http://www.sciencedaily.com/articles/e/electricity_generation.htm

http://www.facstaff.bucknell.edu/mastascu/elessonshtml/TOC_BasicConcepts.html

http://www.need.org/needpdf/infobook_activities/SecInfo/Elec3S.pdf

http://en.wikipedia.org/wiki/Electrical_resistivities_of_the_elements_(data_page)

http://www.allaboutcircuits.com/vol_1/chpt_12/6.html

http://en.wikipedia.org/wiki/Resistor;

http://www.doctronics.co.uk/resistor.htm

http://www.kpsec.freeuk.com/components/resist.htm;

http://wiki.answers.com/Q/How_is_the_resistor_behaved_at_high_frequency#ixzz1Eh5bUyyk

http://www.vishay.com;

http://en.wikipedia.org/wiki/Capacitor

http://en.wikipedia.org/wiki/Electric_double-layer_capacitor

http://en.wikipedia.org/wiki/Inductor

http://en.wikipedia.org/wiki/Transformer

MEASUREMENT AND ERROR

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Express the need for measurement and analysis of measured data
2. Define technical terms related to a measurement such as accuracy, precision, resolution, error, tolerance, etc.
3. Describe the input/output relationship for a measuring equipment (static calibration)
4. Analyze the accuracy and precision of a measurement.
5. Compare and contrast the accuracy and precision for a measurement.
6. Use significant figures to express the precision of a measurement.
7. Classify the measurement errors and list ways of reducing them
8. Analyze the measured data using statistical measures such as the mean values and deviations from the mean.
9. Determine the probability of errors using statistical distribution functions.
10. Analyze the uncertainties in meter readings for analog and digital displays.
11. Calculate the limiting and probable errors in a set of measurement.
12. Infer propagation of errors as the result of a measurement is used in calculations.
13. Identify the number of samples needed to infer the population statistics.

INTRODUCTION

An instrument is a device designed to collect data from an environment, or from a unit under test, and to display information to a user based on the collected data. Such an instrument may employ a transducer to sense changes in a physical parameter, such as temperature or pressure, and to convert the sensed information into electrical signals, such as voltage or frequency variations. The term instrument may also cover, and for purposes of this description it will be taken to cover, a physical or software device that performs an analysis on data acquired from another instrument and then outputs the processed data to display or recording means. This second category of instruments would, for example, include oscilloscopes, spectrum analyzers and digital multimeters. The types of source data collected and analyzed by instruments may thus vary widely, including both physical parameters such as temperature, pressure, distance, and light and sound frequencies and amplitudes, and also electrical parameters including voltage, current, and frequency.

An engineer has to make a lot of measurements, collect and analyze data, and make decisions about the validity of his approaches and procedures. He must have a clear idea about the results he is going to obtain. In this respect, he may develop models of his expectations and compare the outcomes from the experiments to those from the model. He uses various measuring instruments whose reliabilities have outmost importance in successes of his decisions. Characteristics of measuring instruments that are used in selecting the proper ones are reviewed in the first section. Section 2 deals with analyses of measurement data. Section 3 handles the analyses of uncertainties and establishment of engineering tolerances.

CHARACTERISTICS OF MEASURING INSTRUMENTS

Definition of Terms

The characteristics of measuring instruments are specified using terms shortly defined below. The full description of some of these terms will be provided later with examples.

True value: standard or reference of known value or a theoretical value

Accuracy: closeness to the true value; closeness with which an instrument reading approaches the true or accepted value of the variable (quantity) being measured. It is considered to be an indicator of the total error in the measurement without looking into the sources of errors.

Precision: a measure of the reproducibility of the measurements; given a fixed value of a variable, precision is a measure of the degree to which successive measurements differ from one another i.e., a measure of reproducibility or agreement with each other for multiple trials.

Sensitivity: the ability of the measuring instrument to respond to changes in the measured quantity. It is expressed as the ratio of the change of output signal or response of the instrument to a change of input or measured variable.

Resolution: the smallest change in measured value to which the instrument will respond, i.e. the smallest incremental quantity that can be reliably measured.

Error: deviation from the true value of the measured variable.

Linearity: the percentage of departure from the linear value, i.e., maximum deviation of the output curve from the best-fit straight line during a calibration cycle.

Tolerance: maximum deviation allowed from the conventional true value. It is not possible to build a perfect system or make an exact measurement. All devices deviate from their ideal (design) characteristics and all measurements include uncertainties (doubts). Hence, all devices include tolerances in their specifications. If the instrument is used for high-precision applications, the design tolerances must be small. However, if a low degree of accuracy is acceptable, it is not economical to use expensive sensors and precise sensing components.

Static Calibration

The static calibration for a multi-input instrument is carried out by keeping all inputs except one at some constant values. The single input under study is varied over some range of constant values, causing the output(s) to vary over some range of constant values. The input-output relation developed in this way is called "static calibration". A calibration curve for a dual-input single-output system is shown in Figure 3.1. The static sensitivity (S) is the slope of the calibration curve and is

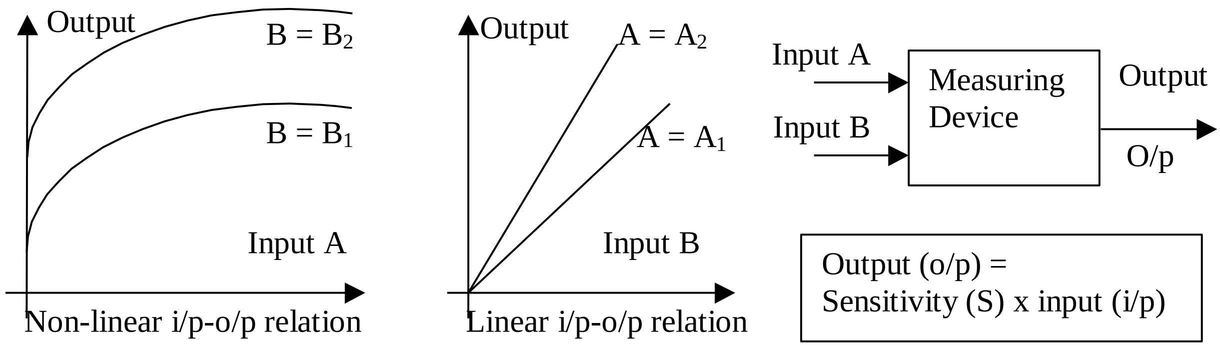


Figure 3.1 Static calibration curves for a multi-input single-output system

defined as,

$$S = \partial(\text{output})/\partial(\text{input})$$

S is a constant for linear relation. Otherwise, S is a function of the input.

Accuracy and Precision

A measurement isn't very meaningful without an error estimate! No measurement made is ever exact. The accuracy (correctness) and precision (number of significant figures) of a measurement are always limited by apparatus used, skill of the observer and the basic physics in the experiment and the experimental technique used to access it. The goal of the experimenter is to obtain the best possible value of some quantity or to validate/falsify a theory. What comprises a deviation from a theory? Every measurement must give the range of possible value. In this section we will discuss the accuracy and precision with examples.

Accuracy

Accuracy is defined as the degree of conformity of a measured value to the true (*conventional true value* – CTV) or accepted value of the variable being measured. It is a measure of the total error in the measurement without looking into the sources of the errors. Mathematically it is expressed as the maximum absolute deviation of the readings from the CTV. This is called the absolute accuracy.

$$Absolute\,accuracy = \left|\max imum\,devation\,from\,CTV\right|$$

$$\mathrm{Re}\,lative\,accuracy = \frac{absolute\,accuracy}{CTV};$$

$$Percentage(\%)\,accuracy = relative\,accuracy \times 100$$

Example 3.1

A voltmeter is used for reading on a standard value of 50 volts, the following readings are obtained: 47, 52, 51, 48

- Conventional true value (CTV) = 50 volts,
- Maximum (V_{MAX}) = 52 volts and minimum (V_{MIN}) = 47 volts.
- CTV – V_{MIN} = 50 – 47 = 3 volts; V_{MAX} – CTV = 52 – 50 = 2 volts.
- Absolute accuracy = max of {3, 2} = 3 volts.
- Relative accuracy = 3/50 = 0.06 and % accuracy = 0.06x100 = 6%

Precision

Precision is composed of two characteristics as *conformity* and the number of *significant figures*.

Conformity

The conformity is the ability of an instrument to produce the same reading, or it is the degree of agreement between individual measurements. So, it is also called repeatability or reproducibility. Mathematically it is expressed as "the absolute maximum deviation from the average of the readings", i.e. *Precision (Pr) = max* $\{(V_{AV} - V_{MIN}), (V_{MAX} - V_{AV})\}$

Bias

The difference between CTV and average value (V_{AV}) is called the bias. Ideally, the bias should be zero. For a high quality digital voltmeter, the loading error is negligible yielding bias very close to zero.

Bias = CTV - V_{AV} (3.6)

In the previous example the average (V_{AV}) = (47+48+51+52)/4 = 49.5 V

Pr = max {(49.5 – 47), (52 – 49.5)} = 2.5 volts. Thus, Bias = 50 – 49.5 = 0.5 volt.

A consistent bias can be due to the presence of a systematic error or instrument loading. Hence, eliminating the causes removes the bias. However, if the bias is consistent and causes cannot be identified and/or eliminated, the bias can be removed by re-calibrating the instrument.

Example 3.2

A known voltage of 100 volts (CTV = 100 V) is read five times by a voltmeter and following readings are obtained: 104, 103, 105, 103, 105

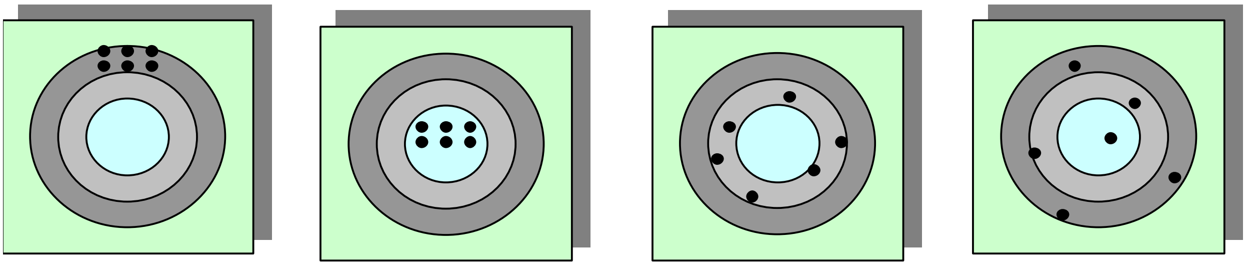
- Average reading = (1/5)x(104+103+105+103+105) = 104 volts
- Pr = max $\{(V_{AV} - V_{MIN}), (V_{MAX} - V_{AV})\}$ = max {(104 – 103), (105 – 104)} = 1 volt
- Accuracy = max $\{(CTV - V_{MIN}), (V_{MAX} - CTV)\}$ = max {(100 – 103), (105 – 100)} =5 V
- Bias = CTV – average= 100 – 104 = -4 volts.

If we re-calibrated the instrument to remove the bias, then the average reading = CTV. The new readings would be 100, 99, 101, 99, 101. Hence, after re-calibration, average = CTV = 100 volts, and accuracy = precision = 1 volt.

Accuracy versus Precision

The distinction between accuracy and precision can be illustrated by an example: two voltmeters of the same make and model may be compared. Both meters have knife-edge pointers and mirror backed scales to avoid parallax, and they have carefully calibrated scales. They may therefore be read to the same precision. If the value of the series resistance in one meter changes considerably, its readings may be in error by a fairly large amount. Therefore the accuracy of the two meters may be quite different. To determine which meter is in error, a comparison measurement with a standard meter should be made.

Accuracy refers to the degree of closeness or conformity to the true value at the quantity under measurement. Precision refers to the degree of agreement within a group of measurements or instruments. The target-shooting example shown in Figure 3.2 illustrates the difference. The high accuracy, poor precision situation occurs when the person hits all the bullets on a target plate on the outer circle and misses the bull's eye. In the second case, all bullets hit the bull's eye and spaced



Poor accuracy High precision

High accuracy High precision

Average accuracy Poor precision

Poor accuracy Poor precision

Figure 3.2 An illustration of accuracy and precision

closely enough leading to high accuracy and high precision. The bullet hits are placed symmetrically with respect to the bull's eye in the third case but spaced apart yielding average accuracy but poor precision. In the last example, the bullets hit in a random manner, hence poor accuracy and poor precision.

The scatter graph in Figure 3.3 shows an alternative way of presenting the accuracy and

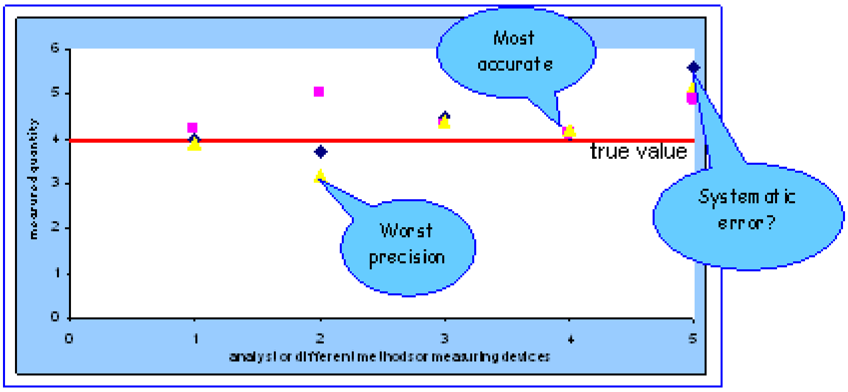


Figure 3.3 An illustration of accuracy and precision by a scatter graph

precision. Same quantity was measured three times by 5 different analyst or methods or measuring instruments. Distribution of readings around the true value indicates the most accurate, most precise and least accurate and least precise readings. The last reading is too far away from the true value and from other readings that may indicate a systematic error.

Precision is composed of two characteristics as stated: conformity and the number of significant figures to which a measurement may be made. Consider, for example, that the insulation resistance of a transformer has the true value 2,475,653 Ω. It is measured by an ohmmeter, which consistently and repeatedly indicates 2.5 MΩ. But can the observer "read" the true value from the scale? His estimates from the scale reading consistently yield a value of 2.5 MΩ. This is as close to the true value as he can read the scale by estimation. Although there are no deviations from the observed value, the error produced by the limitation of the scale reading is a precision error. The example illustrates that conformity is a necessary, but not sufficient, condition for precision because of the lack of significant figures obtained. Similarly, precision is a necessary, but not sufficient condition for accuracy.

Too often the beginning student is inclined to accept instrument readings at face value. He is not aware that the accuracy of a reading is not necessarily guaranteed by its precision. In fact, good measurement technique demands continuous skepticism as to the accuracy of the results.

In critical work, good practice dictates that the observer make an independent set of measurements, using different instruments or different measurement techniques, not subject to the same systematic errors. He must also make sure that the instruments function properly and are calibrated against a known standard, and that no outside influence affects the accuracy of his measurements.

Significant Figures

An indication of the precision of the measurement is obtained from the number of significant figures in which the result is expressed. Significant figures convey actual information regarding the magnitude and the measurement precision of a quantity. The more significant figures the greater the precision of measurement.

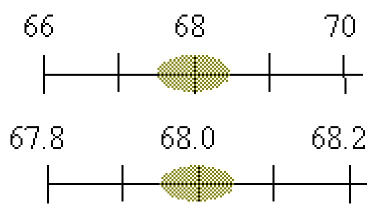


Figure 3.4 An illustration of significant figures

Figure 3.4 illustrates the importance of significant figures with an example. If a resistor is specified as having a resistance of 68 Ω, its resistance should be closer to 68 Ω than to 67 Ω or 69 Ω. If the value of the resistor is described as 68.0 Ω, it means that its resistance is closer to 68.0 Ω than it is to 67.9 Ω or 68.1 Ω. In 68 Ω there are two significant figures; in 68.0 Ω there are three. The latter, with more significant figures, expresses a measurement of greater precision than the former.

It is customary to record a measurement with all the digits of which we are sure nearest to the true value. For example in reading a voltmeter, the voltage may be read as 117.1 V. This simply indicates that the voltage, read by the observer to best estimation, is closer to 117.1 V than to 117.0 V or 117.2 V. Another way of expressing this result is that it indicates the range of possible error. The voltage may be expressed as 117.1 ± 0.05 V, indicating that the value of the voltage lies between 117.05 V and 117.15 V.

When two or more measurements with different degrees of accuracy are added, the result is only as accurate as the least accurate measurement. Consider the following example:

Example 3.3

Two resistors, R_1 and R_2, are connected in series. Individual resistance measurements using a digital multimeter, yield R_1 = 18.7 Ω and R_2 = 3.624 Ω. Calculate the total resistance to the appropriate number of significant figures.

SOLUTION

$R_1 = 18.\mathit{7}\ \Omega$ (three significant figures)

$R_2 = 3.62\mathit{4}\ \Omega$ (four significant figures)

$R_T = R_1 + R_2 = 22.\mathit{324}\ \Omega$ (five significant figures) $= 22.\mathit{3}\ \Omega$

The doubtful figures are written in *italic.* Any digit in the result is doubtful if it's computation involves doubtful digits. In the addition of R_1 and R_2 the last three digits of the sum are doubtful figures. There is no value whatsoever in retaining the last two digits (the 2 and the 4) because one of the resistors is accurate only to three significant figures or tenths of an ohm. The result should therefore also be reduced to three significant figures or the nearest tenth. i.e., 22.3 Ω. Note that if extra digits accumulate in the answer, they should be discarded or rounded off. In the usual practice, if the digit in the first place to be discarded (most significant of digits to be discarded) is less than five, it and the following digits are dropped from the answer as it was done in the example. If the digit in the first place to be discarded is five or greater, the previous digit is increased by one. For three-digit precision, therefore, 22.*324* should be rounded off to 22.3; and 22.*354* to 22.4.

Types of Errors (Uncertainties)

No measurement can be made with perfect accuracy, but it is important to find out what the accuracy actually is and how different errors have entered into the measurement. A study of errors is a first step in finding ways to reduce them. Such a study also allows us to determine the accuracy of the final test result. Errors may come from different sources and are usually classified under three main headings as:

Gross errors: largely human errors, among them misreading of instruments, incorrect adjustment and improper application of instruments, and computational mistakes.

Systematic (determinate) errors: shortcomings of the instruments, such as defective or worn parts, and effects of the environment on the equipment or the user. They are sometimes called bias due to error in one direction- high or low. They are generally originated from a known cause such as result from mis-calibrated device, experimental technique that always gives a measurement higher (or lower) than the true value, operator's limitations and calibration of glassware, sensor, or instrument. Their effects can be minimized by trying a different method for the same measurement. They can be corrected when determined.

Systematic errors may be of a constant or proportional nature as illustrated in figure 3.5. The constant error influences the **intercept** while the proportional error influences the **slope**.

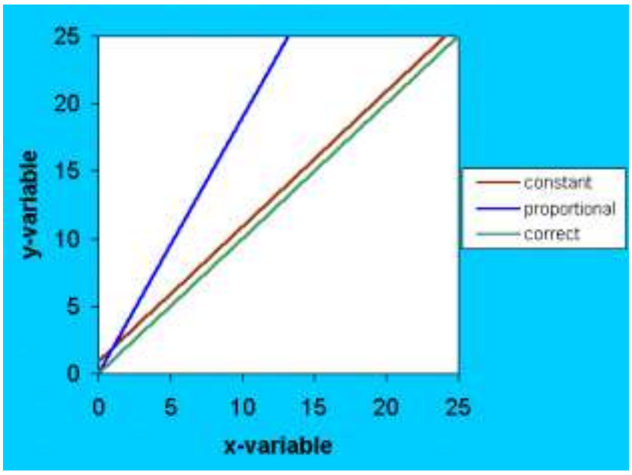


Figure 3.5 Constant and proportional type errors

Random (indeterminate) errors: those due to causes that cannot be directly established because of random variations in the parameter or the system of measurement. Hence, we have no control over them. Their random nature causes both high and low values to average out. Multiple trials help to minimize their effects. We deal with them using statistics. Figure 3.6 provides a schematic summary of errors and their possible means of elimination. For example, errors caused by the loading effect of the voltmeter can be avoided by using it intelligently. A low resistance voltmeter should not be used to measure voltages at the input of a voltage amplifier. In this particular measurement, a high input impedance voltmeter (such as a digital voltmeter - DVM) is required. Gross and systematic errors cannot be treated mathematically. They can be avoided only by taking care in reading and recording the measurement data. Good practice requires making more than one reading of the same quantity, preferably by a different observer. Never place complete dependence on one reading but take at least three separate readings, preferably under conditions in which instruments are switched off/on.

The error may be originated from the sampling of the source, preparation of the samples and measurement and analysis of the measurand. Care must be taken so that the sample is representative of the whole population (homogeneous vs. heterogeneous). No unwanted additions or deletions are allowed during the preparatory phase. Finally, calibration of the measuring instrument using standard measurands or standard solutions is done as frequent as defined by the equipment manufacturer. One way to assess total error is to treat a reference standard as a sample. The reference standard would be carried through the entire process to see how close the results are to the reference value.

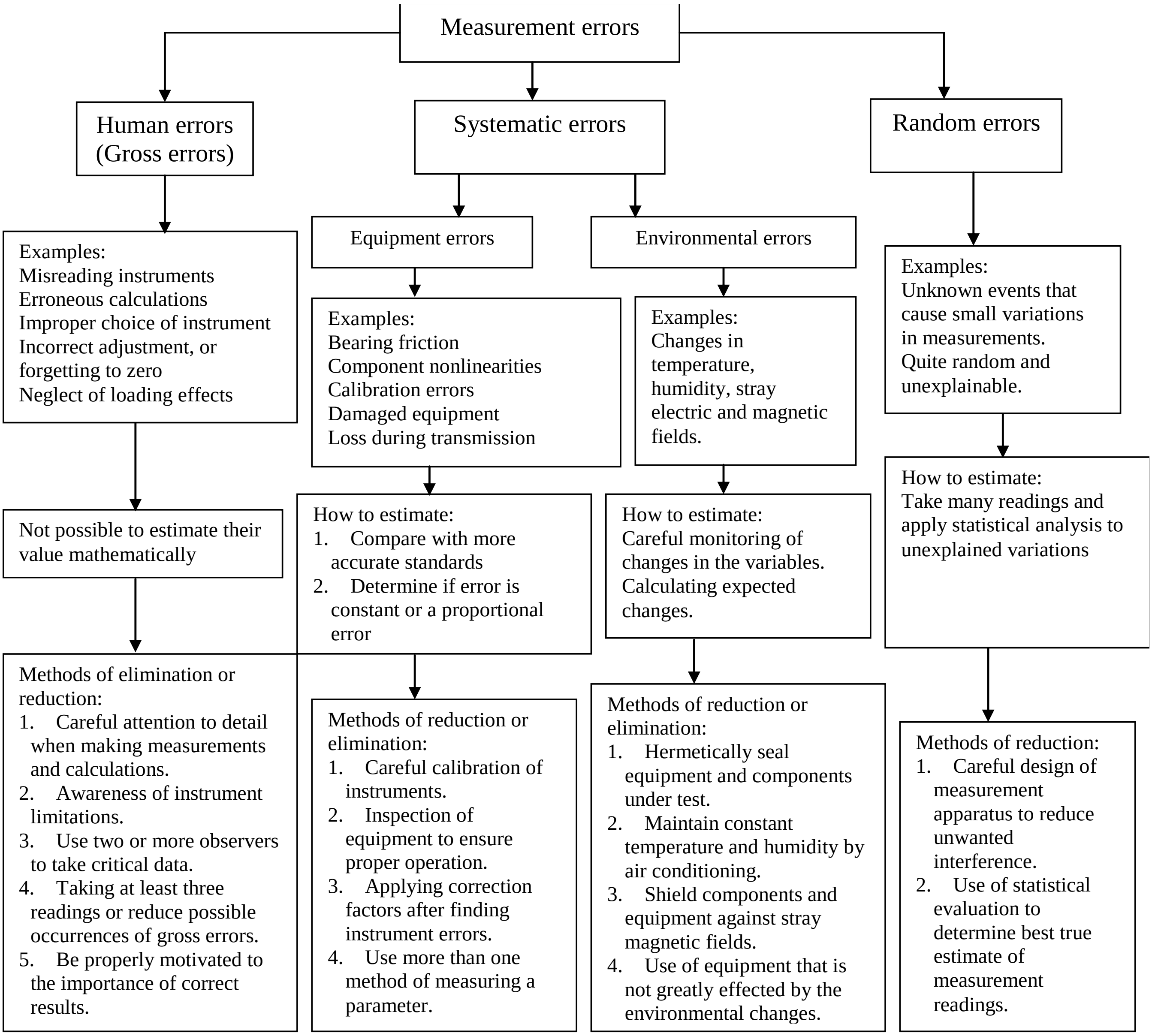


Figure 3.6 A schematic summary of measurement errors

ANALYSIS OF MEASUREMENT DATA

Table 3.1. Deviations around mean

d_1 = 12.8 - 12.65 = 0.15 mA
d_2 = 12.2 - 12.65 = -0.45 mA
d_3 = 12.5 - 12.65 = -0.15 mA
d_4 = 13.1 - 12.65 = 0.45 mA
d_5 = 12.9 - 12.65 = 0.25 mA
d_6 = 12.4 - 12.65 = -0.25 mA

A statistical analysis of measurement data is common practice because it allows an analytical determination of the uncertainty of the final test result. The outcome of a certain measurement method may be predicted on the basis of sample data without having detailed information on all the disturbing factors. To make statistical methods and interpretations meaningful, a large number of measurements are usually required. Also, systematic errors should be small compared with residual or random errors, because statistical treatment of data cannot remove a fixed bias contained in all the measurements.

Arithmetic Mean

The most probable value of a measured variable is the arithmetic mean of the number of readings taken. The best approximation will be made when the number of readings of the same quantity is very large. Theoretically, an infinite number of readings would give the best result although in practice only a finite number of measurements can be made. The arithmetic mean is given by:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x}{n}$$

where $\bar{x}$ = arithmetic mean, $x_1 \ldots x_n$ = readings taken, and n = number of readings.

Example 3.4

A set of independent current measurements was taken by six observers and recorded as 12.8 mA, 12.2 mA, 12.5 mA, 13.1 mA, 12.9 mA, and 12.4 mA. Calculate the arithmetic mean.

$$\bar{x} = \frac{12.8 + 12.2 + 12.5 + 13.1 + 12.9 + 12.4}{6} = 12.65mA$$

Deviation from the Mean

In addition to knowing the mean value of a series of measurements, it is often informative to have some idea of their range about the mean. Deviation is the departure of a given reading from the arithmetic mean of the group of readings. If the deviation of the first reading x_1 is called d_1, and that of the second reading, x_2 is called d_2 and so on, then the deviations from the mean can be expressed as

$$d_1 = x_1 - \bar{x}\,;\; d_2 = x_2 - \bar{x}\,;\; \ldots;\; d_n = x_n - \bar{x}$$

The deviation from the mean may have a positive or a negative value and that the algebraic sum of all the deviations must be zero. The computation of deviations for the previous example is given in Table 3.1.

Average Deviation

The average deviation is an indication of the precision at the instruments used in making the measurements. Highly precise instruments will yield a low average deviation between readings. By definition average deviation is the sum of the absolute values of the deviations divided by the number of readings. The absolute value of the deviation is the value without respect to sign. Average deviation may be expressed as

$$D = \frac{|d_1| + |d_2| + |d_3| + \cdots + |d_n|}{n} = \frac{\sum d}{n}$$

Example 3.5

The average deviation for the data given in the above example:

$$D = \frac{0.15 + 0.45 + 0.15 + 0.45 + 0.25 + 0.25}{6} = 0.283mA$$

Standard Deviation

The range is an important measurement. It indicates figures at the top and bottom around the average value. The findings farthest away from the average may be removed from the data set without affecting generality. However, it does not give much indication of the spread of observations about the mean. This is where the standard deviation comes in.

In statistical analysis of random errors, the root-mean-square deviation or standard deviation is a very valuable aid. By definition, the standard deviation σ of a finite number of data is the square root of the sum of all the individual deviations squared, divided by the number of readings minus one. Expressed mathematically:

$$\sigma = \sqrt{\frac{d_1^2 + d_2^2 + d_3^2 + \cdots + d_n^2}{n-1}} = \sqrt{\frac{\sum d_i^2}{n-1}}$$

Another expression for essentially the same quantity is the variance or mean square deviation, which is the same as the standard deviation except that the square root is not extracted. Therefore

variance (V) = mean square deviation = σ^2

The variance is a convenient quantity to use in many computations because variances are additive. The standard deviation however, has the advantage of being of the same units as the variable making it easy to compare magnitudes. Most scientific results are now stated in terms of standard deviation.

Probability of Errors

Normal Distribution of Errors

A practical point to note is that, whether the calculation is done on the whole "population" of data or on a sample drawn from it, the population itself should at least approximately fall into a so called "normal (or Gaussian)" distribution.

For example, 50 readings of voltage were taken at small time intervals and recorded to the nearest 0.1 V. The nominal value of the measured graphically in the form of a block diagram or histogram in which the number of observations is plotted against each observed voltage reading. The histogram and the table data are given in Figure 3.7. The figure shows that the largest number of readings (19) occurs at the central value of 100.0 V while the other readings are placed more or less symmetrically on either side of the central value. If more readings were taken at smaller increments, say 200 readings at 0.05-V intervals, the distribution of observations would remain approximately symmetrical about the central value and the shape of the histogram would be about the same as before. With more and more data taken at smaller and smaller increments, the contour of the histogram would finally become a smooth curve as indicated by the dashed line in the figure. This bell shaped curve is known as a Gaussian curve. The sharper and narrower the curve, the more definitely an observer may state that the most probable value of the true reading is the central value

Tabulation of Voltage Readings

Voltage reading (volts)	# of reading
99.7	1
99.8	4
99.9	12
100.0	19
100.1	10
100.2	3
100.3	1

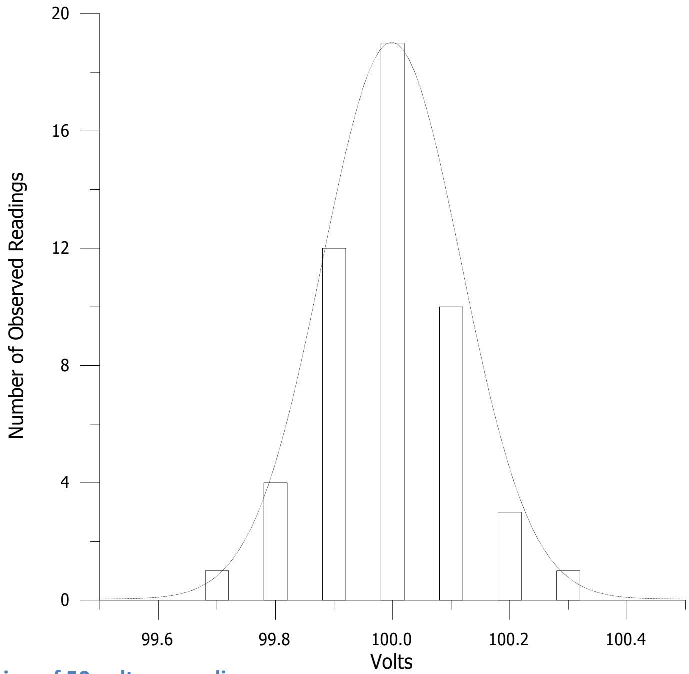


Figure 3.7 Distribution of 50 voltage readings

or mean reading.

For unbiased experiments all observations include small disturbing effects, called random errors. Random errors undergo a Normal (Gaussian) law of distribution shown in Figure 3.8. They can be positive or negative and there is equal probability of positive and negative random errors. The error distribution curve indicates that:

- Small errors are more probable than large errors.
- Large errors are very improbable.
- There is an equal probability of plus and minus errors so that the probability of a given error will be symmetrical about the zero value.

$$\Pr obability of\, error = \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{x^2}{2\sigma})$$

Area Under the Probability Curve	
Deviation $\pm\sigma$	Fraction of total area
0.6745	0.5000
1.0	0.6828
2.0	0.9546
3.0	0.9972

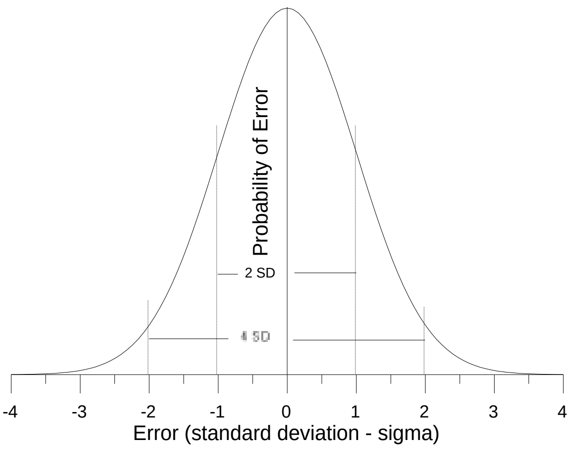


Figure 3.8 The error distribution curve for a normal (Gaussian) distribution

Table 3.2 Deviations in readings

Reading, x	Deviation	
	d	d^2
101.	-0.1	0.01
101.7	0.4	0.16
101.3	0.0	0.00
101.0	-0.3	0.09
101.5	0.2	0.04
101.3	0.0	0.00
101.2	-0.1	0.01
101.4	0.1	0.01
101.3	0.0	0.00
101.1	-0.2	0.04
Σx=1013.0	$\Sigma\|d\|$=1.4	Σd^2=0.36

The error distribution curve in Figure 3.8 is based on the Normal (Gaussian) law and shows a symmetrical distribution of errors. This normal curve may be regarded as the limiting form of the histogram in which the most probable value of the true voltage is the mean value of 100.0V. Table 3.2 lists the readings, deviations and deviation squares of readings from the mean value. The reason why the standard deviation is such a useful measure of the scatter of the observations is illustrated in the figure. If the observations follow a "normal" distribution, a range covered by one standard deviation above the mean and one standard deviation below it (i.e. $\bar{x} \pm 1$ SD) includes about 68% of the observations. A range of 2 standard deviations above and below ($\bar{x} \pm 2$ SD) covers about 95% of the observations. A range of 3 standard deviations above and below ($\bar{x} \pm 3$ SD) covers about 99.72% of the observations.

Range of a Variable

If we know the mean and standard deviation of a set of observations, we can obtain some useful information by simple arithmetic. By putting 1, 2, or 3 standard deviations above and below the mean we can estimate the ranges that would be expected to include about 68%, 95% and 99.7% of observations. Ranges for ± SD and ± 2 SD are indicated by vertical lines. The table in the inset (next to the figure) indicates the fraction of the total area included within a given standard deviation range.

Acceptable range of possible values is called the confidence interval. Suppose we measure the resistance of a resistor as (2.65 ± 0.04) kΩ. The value indicated by the color code is 2.7 kΩ. Do the two values agree? Rule of thumb: if the measurements are within 2 SD, they agree with each other. Hence, ± 2 SD around the mean value is called the range of the variable.

Probable Error

The table also shows that half of the cases are included in the deviation limits of $\pm 0.6745\sigma$. The quantity *r* is called the *probable error* and is defined as

$$\textit{probable error } r = \pm 0.6745\sigma$$

This value is *probable* in the sense that there is an even chance that any one observation will have a random error no greater than $\pm r$. Probable error has been used in experimental work to some extent in the past, but standard deviation is more convenient in statistical work and is given preference.

Example 3.6

Ten measurements of the resistance of a resistor gave 101.2 Ω, 101.7 Ω, 101.3 Ω, 101.0 Ω, 101.5 Ω, 101.3 Ω, 101.2 Ω, 101.4 Ω, 101.3 Ω, and 101.1 Ω. Assume that only random errors are present. Calculate the arithmetic mean, the standard deviation of the readings, and the probable error.

SOLUTION: With a large number of readings a simple tabulation of data is very convenient and avoids confusion and mistakes.

Arithmetic mean, $\bar{x} = \dfrac{\sum x}{n} = \dfrac{1013.0}{10}$ = 101.3 Ω

Standard deviation, $\sigma = \sqrt{\dfrac{d^2}{n-1}} = \sqrt{\dfrac{0.36}{9}}$ = 0.2 Ω

Probable error = 0.6745 σ = 0.6745 x 0.2 = 0.1349 Ω

Some MS Excel Functions

The electronic spreadsheet program Microsoft Excel offers many built-in statistical functions that can be used in data analysis. They can be easily accessed from the "insert function" menu. The salient ones are:

=SUM(A2:A5) Find the sum of values in the range of cells A2 to A5.

.=AVERAGE(A2:A5) Find the average of the numbers in the range of cells A2 to A5.

=AVEDEV(A2:A5) Find the average deviation of the numbers in the range of cells A2 to A5.

=STDEV(A2:A5) Find the sample standard deviation (unbiased) of the numbers in the range of cells A2 to A5.

=STDEVP(A2:A5) Find the sample standard deviation (biased) of the numbers in the range of cells A2 to A5.

Determining Random Errors

Random errors are due to random variations in the parameter or the system of measurement as mentioned before. We deal with them using statistics and multiple trials generally help to minimize their effects. One of their primary causes can be pinpointed to instrument limit of error and least count. The least count is the smallest division that is marked on the instrument. The **instrument limit of error** is the precision to which a measuring device can be read, and is always equal to or smaller than the least count. The estimation of the uncertainty is important. For example, assume a volt meter may give us 3 significant digits, but we observe that the last two digits oscillate during the measurement. What is the error? Average deviation or standard deviation based on repeated measurements of the same quantity are used in determining the uncertainty.

Uncertainties in Reading Digital Displays

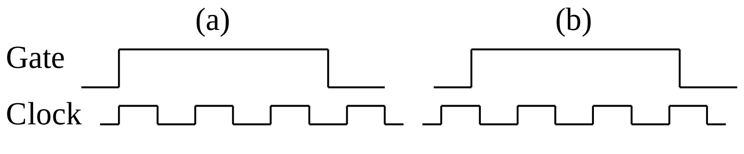


Figure 3.9 Unsynchronized gate and clock signals

A digital meter involves counting from a clock signal during the gate interval as depicted in Figure 3.9. As the gate and clock signals are not synchronized and combined in an AND gate, case (b) results 4 pulses while case (a) supplies only 3 pulses. Hence, a digital read-out has an uncertainty of $\pm$1 digit.

Uncertainties in Reading Analog Displays

The uncertainty in analog displays depends upon the organization of display screen and capabilities of the reader. In analog multi meters it is accepted as ± ½ scale divisions (the least count). In oscilloscope displays, it depends upon the thickness of the trace and it is around ±½ mm. For both analog and digital displays, it is recommended to take the measurement as close to full scale as possible to minimize the effect of the reading error. The following example illustrates the uncertainties in analog meter readings.

Example 3.7

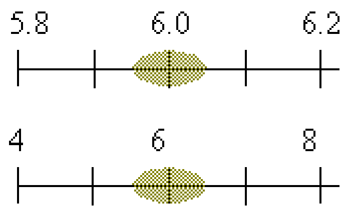


Figure for example 2.7.

An analog voltmeter is used to measure a voltage. It has 100 divisions on the scale. The voltage read is 6 volts and the meter has two ranges as 0 – 10 volts and 0 – 100 volts. Find the uncertainty in the measured value in both ranges.

Uncertainty = ± ½ V_{FSD} / # of divisions, where V_{FSD} is the voltage measured at full-scale deflection of the meter.

On 10 V range, uncertainty = ± ½ 10 / 100 = ± 0.05 V yielding V = 6 ± 0.05 volt.

On 100 V range, uncertainty = ± ½ 100 / 100 = ± 0.5 V yielding V = 6 ± 0.5 volt.

Relative uncertainty: on 10 V range, 0.05/6 = 1/120 = 0.0083;

on 100 V range, 0.5/6 = 1/12 = 0.083

Percentage uncertainty: on 10 V range, (0.05/6)x100 = 0.83%, and

on 100 V range, (0.5/6)x100 = 8.3%

Exercise(adapted from http://www.hep.vanderbilt.edu/~julia/VUteach/PHY225a)

For each of the three rulers in Figure 3.10, determine and record

- The least count of the scale (smallest division) – scales are all in cm
- Length of the gray rods
- Uncertainties in your readings

Compare your result with those of the student next to you

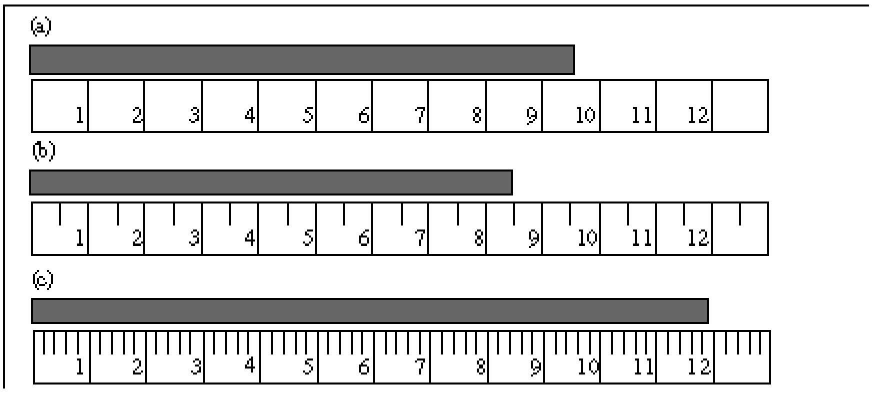


Figure 3.10 Three rulers with different divisions

UNCERTAINTY ANALYSIS

Any system that relies on a measurement system will involve some amount of uncertainty (doubt). The uncertainty may be caused by individual inaccuracy of sensors, limitations of the display devices, random variations in measurands, or environmental conditions. The accuracy of the total system depends on the interaction of components and their individual accuracies. This is true for measuring instruments as well as production systems that depend on many subsystems and components. Each component will contribute to the overall error, and errors and inaccuracies in each of these components can have a large cumulative effect.

Mathematical Analysis of the Uncertainty

If an experiment has number of component sources, each being measured individually using independent instruments, a procedure to compute the total accuracy is necessary. Let

$R = f(x_1, x_2, x_3, \ldots, x_n)$

where $x_1, x_2, x_3, \ldots, x_n$ are independent variables. Each variable is defined as

$$x_i = \bar{x}_i \pm \Delta x_i$$

$i = 1, 2, \ldots, n$; $\bar{x}_i$ is known as the nominal value; Δx_i is known as the uncertainty in the variable x_1;

then $R = \bar{R} \pm \Delta R$ where $\bar{R} = f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$

The uncertainty $\Delta R = w_R$ can be computed using Taylor's series expansion and statistical analysis. All partial derivatives of R are taken. The partial derivative $\frac{\partial f}{\partial x_i}$ shows the sensitivity of R to variable x_i. Since the measurements have been taken, the x_i values are known and can be substituted into the expressions for the partial derivatives and partial derivatives are evaluated at known values of $x_1, x_2, \ldots, x_n$.

Limiting Error

Two methods are commonly used for determining the uncertainty. The first one is called the *method of equal effects* and it yields the limiting (guarantied) error (maximum uncertainty possible).

$$\Delta R = \omega_R = \sum_{i=1}^{n} \left| [\frac{\partial f}{\partial x_i}]_o \Delta x_i \right|$$

where $(\frac{\partial f}{\partial x_i})_o$ is the partial derivative of the function with respect to x_I calculated at the nominal value. The absolute value is used because some of the partial derivatives may be negative and would have a canceling effect. If one of the partial derivative is high compared to the others, then a small uncertainty in the corresponding variable has large effect on the total error. Hence, the equation also illustrates which of the variable exerts strongest influence on the accuracy of the overall results.

Example 3.8

The voltage generated by a circuit is equally dependent on the value of three resistors and is given by the following equation: $V_0 = I(R_1R_2/R_3)$

If the tolerance of each resistor is 1 per cent, what is the maximum error of the generated voltage?

SOLUTION: Let us find the sensitivities first.

$$\frac{\partial V_0}{\partial R_1} = I\frac{R_2}{R_3} = \frac{V_0}{R_1}\ ;\ \frac{\partial V_0}{\partial R_2} = I\frac{R_1}{R_3} = \frac{V_0}{R_2}\ ;\ \frac{\partial V_0}{\partial R_3} = -I\frac{R_1R_2}{R_3^2} = -\frac{V_0}{R_3}$$

All tolerances are given as 1%, therefore: $\Delta R_1 = 0.01R_1$; $\Delta R_2 = 0.01R_2$; $\Delta R_3 = 0.01R_3$

$$\Delta V_0 = \frac{\partial V_0}{\partial R_1}\Delta R_1 + \frac{\partial V_0}{\partial R_2}\Delta R_2 + \frac{\partial V_0}{\partial R_3}\Delta R_3$$ That yields $\Delta V_0 = 0.03V_0$

The total variation of the resultant voltage is ±0.3 per cent, which is the algebraic sum of the three tolerances. This is true in the first approximation. The maximum error is slightly different from the sum of the individual tolerances. On the other hand, it is highly unlikely that all three components of this example would have the maximum error and in such a fashion to produce the maximum or minimum voltage. Therefore, the statistical method outlined below is preferred.

Expected Value of Uncertainty

The second method is called the *square root of sum of squares*. It is based on the observations stated before for the random errors. It yields the expected value of the uncertainty and computed as

$$(\Delta R)^2 = (\omega_R)^2 = \sum_{i=1}^{n}[(\frac{\partial f}{\partial x_i})^2]_o(\Delta x_i)^2$$

This will be used throughout the course unless the question asks the limiting error, or maximum possible uncertainty.

Example 3.9

P = *VI*, if *V* = 100 ± 2 volt (measured) and *I* = 10 ± 0.2 Amp (measured), determine the maximum allowable uncertainty, and the expected uncertainty in power.

SOLUTION: $\Delta P_m = w_{Pm} = \left|\frac{\partial P}{\partial V}\Delta V\right| + \left|\frac{\partial P}{\partial I}\Delta I\right| = 10x2 + 100x0.2 = 40$ watts is the limiting value of the uncertainty.

However, the expected uncertainty $\Delta P = w_P = \sqrt{(\frac{\partial P}{\partial V}\Delta V)^2 + (\frac{\partial P}{\partial I}\Delta I)^2}$

$w_P = \sqrt{(Ix2)^2 + (Vx0.2)^2} = \sqrt{(10x2)^2 + (100x0.2)^2} = \sqrt{100x8} = 10\sqrt{8} = 28.3$ watts.

The nominal value of power = 100x10 = 1000 watts

Percentage uncertainty = (28.3/1000)x100 = 2.83%, and P = 1000 ± 28.3 watt.

Example 3.10

The resistance of a certain size of copper wire is given by $R = Ro[1+\alpha(T-20)]$. The resistance at 20°C is Ro = 6Ω ± 0.3%, temperature coefficient α = 0.004/°C ± 1%, temperature T = 30°C ± 1°C. Compute the uncertainty in the resistance R.

SOLUTION: the nominal value of R, $\overline{R} = 6[1+(0.004)(30-20)] = 6.24\Omega$

$\partial R/\partial R_0 = 1 + \alpha(T - 20) = 1 + 0.004(30 - 20) = 1.04$

$\partial R/\partial \alpha = R_0\,[T - 20] = 6(30 - 20) = 60$; $\partial R/\partial T = R_0\,\alpha = 6x0.004 = 0.024$

Uncertainty in the nominal value of R_0 = percentage uncertainty of R_0 x nominal R_0

$\omega_{Ro} = (0.3/100) \times 6 = 0.018$; $\omega_\alpha = (1/100)(0.004) = 4x10^{-5}/°C$; $\omega_T = 1°C$

The uncertainty in the resistance R is given by: $\omega_R = \sqrt{(1.04x0.018)^2 + (60x4x10^{-5})^2 + (0.024x1)^2}$

$= 0.0305\Omega \rightarrow (0.0305/6.24)x100 = 0.49\%$

If the maximum error in the resistance is asked, it can be found as:

$$\Delta R_m = 1.04x0.018 + 60x4x10^{-5} + 0.024x1 = 0.045\Omega$$

Special Case

$R = Y_1^l Y_2^n Y_3^k$, then $(\frac{\omega_R}{Ro})^2 = l^2(\frac{\omega_{Y_1}}{Y_{1o}})^2 + n^2(\frac{\omega_{Y_2}}{Y_{2o}})^2 + k^2(\frac{\omega_{Y_3}}{Y_{3o}})^2$

Series and Parallel Analysis

Example 3.11

Two resistors R_1 and R_2 are connected first in series, then in parallel. Let R_1 = 10 Ω ± 0.5Ω and R_2 = 10 Ω ± 0.5Ω. Find the maximum and expected values for the uncertainty in the combination.

Series analysis

$R_s = R_1 + R_2$; $\partial R_s/\partial R_1 = \partial R_s/\partial R_2 = 1$; $\overline{R_s} = \overline{R_1} + \overline{R_2}$ = 10 + 10 = 20Ω

The limiting error (maximum uncertainty) = $\Delta R_{sm} = \left| \frac{\Delta R_s}{\Delta R_1}\Delta R_1 + \frac{\Delta R_s}{\Delta R_2}\Delta R_2 \right| = \frac{1}{2} + \frac{1}{2} = 1\Omega$

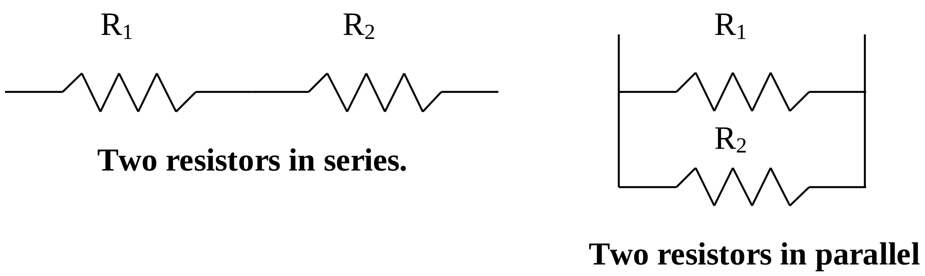


Figure for example 3.11 Series and parallel connected resistors

The uncertainty: $(\Delta R_s)^2 = (\frac{\Delta R_s}{\Delta R_1})^2(\Delta R_1)^2 + (\frac{\Delta R_s}{\Delta R_2})^2(\Delta R_2)^2 = (1)^2(\tfrac{1}{2})^2 + (1)^2(\tfrac{1}{2})^2 = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}$

yielding $\Delta R_s \approx 0.7\Omega$. The relative uncertainty = 0.7/20 = 0.035, and the percentage uncertainty = 3.5%. Therefore, R_s = 20Ω ± 0.7Ω = 20Ω ± 3.5%

Parallel analysis

$R_p = \frac{R_1 R_2}{R_1 + R_2} = \overline{R_p} \pm \Delta R_p$; $\overline{R_p} = \frac{10x10}{10+10} = 5\Omega$

$$\frac{\partial R_p}{\partial R_1} = \frac{(R_1 + R_{2)})R_2 - R_1 R_2}{(R_1 + R_2)^2} = \frac{R_2{}^2}{(R_1 + R_2)^2}$$

$$\frac{\partial R_p}{\partial R_1} = \frac{(R_1 + R_{2)})R_2 - R_1 R_2}{(R_1 + R_2)^2} = \frac{R_2{}^2}{(R_1 + R_2)^2} \frac{\partial R_p}{\partial R_2} = \frac{(R_1 + R_{2)})R_1 - R_1 R_2}{(R_1 + R_2)^2} = \frac{R_1^2}{(R_1 + R_2)^2}$$

Hence, $\frac{\partial R_p}{\partial R_1} = \frac{10^2}{(10+10)^2} = \frac{100}{400} = \frac{1}{4} = \frac{\partial R_p}{\partial R_2}$

$$(\Delta R_p)^2 = (\frac{\partial R_p}{\partial R_1})^2 (\Delta R_1)^2 + (\frac{\partial R_p}{\partial R_2})^2 (\Delta R_2)^2 = (\frac{1}{4})^2 (\frac{1}{2})^2 + (\frac{1}{4})^2 (\frac{1}{2})^2 = (\frac{1}{16})(\frac{1}{4})(1+1) = \frac{1}{32}$$

Therefore the uncertainty in R_p is : $\Delta R_p = \sqrt{\frac{1}{32}} = 0.175\Omega$

The nominal value of R_p = 5Ω, the percentage uncertainty = (0.175/5)x100=3.5%

Then $R_p = 5 \pm 0.175\ \Omega = 5\Omega \pm 3.5\%$

Limiting error in R_p = $\frac{1}{4}(\frac{1}{2} + \frac{1}{2}) = 0.25\Omega$

Summary of how to propagate the errors

Assume that z = f(x,y); table summarizes the relationship between z, x and y.

	Function	Relation between Δz, Δx and Δy	Comment
1	z = x + y	$(\Delta z)^2 = (\Delta x)^2 + (\Delta y)^2$	Addition and subtraction (x+y; x-y): add absolute errors
2	z = x - y		
3	z = xy	$\left(\frac{\Delta z}{z}\right)^2 = \left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2$	Multiplication and division: add relative errors Multiplication by an exact number (a*x): multiply absolute error by the number
4	z = x/y		
5	$z = x^n$	$\frac{\Delta z}{z} = n\frac{\Delta x}{x}$	
6	z = ln x	$\Delta z = \frac{\Delta x}{x}$	
7	$z = e^x$	$\frac{\Delta z}{z} = \Delta x$	

Further explanations can be obtained from MathWorld -

http://mathworld.wolfram.com/ErrorPropagation.html.

Sample and Population Statistics

In many instances, we take samples from a population and infer the population statistics as illustrated in Figure 3.10. Suppose we want to know the average weight of adults. It is not feasible to weigh every single adult and then take the average of all the weights. All adults are called the population. Instead, we decide to take a small fraction of the adults, say 1 out of every 1000, and average these weights. This small fraction is called our sample population. Now we have an average weight for our sample population. We want to know if our sample population average weight is a good estimation of the population average weight. In addition, measurement is a costly process. Hence, we also want to know the minimum sample size that yields uncertainties within the tolerance range.

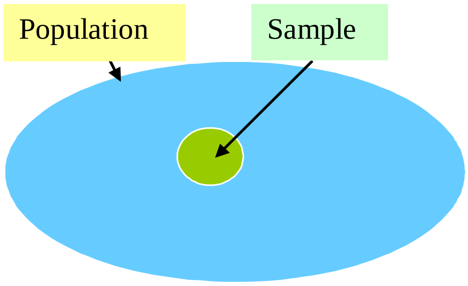


Figure 3.10 Population and sample

Figure 3.11 illustrates the distribution for the population and the sample. For the normal

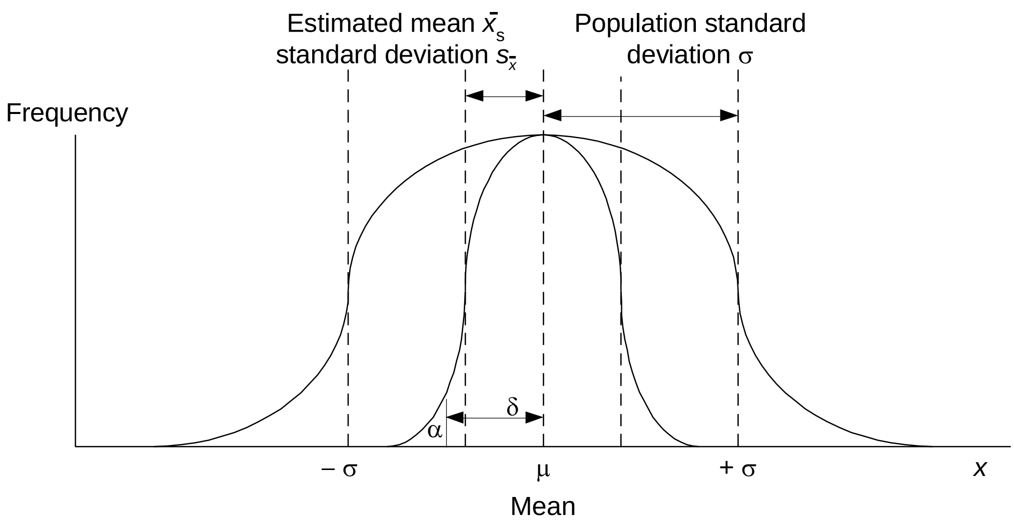


Figure 3.11 Normal distribution curves for population and sample

distribution, 68% of the data lies within ±1 standard deviation. By measuring samples and averaging, we obtain the estimated mean $\bar{x}_s$, which has a smaller standard deviation s_x. α is the tail probability that x_s does not differ from μ by more than δ.

The population standard deviation is

$$\sigma_{population} = \sqrt{\frac{(deviations)^2}{n}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

And the sample standard deviation is

$$\sigma_{sample} = \sigma_s = s = \sqrt{\frac{(deviations)^2}{n-1}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

The sample standard deviation allows for more variation in the sample compared to the population, since sample is only part of population. Dividing by (n-1) increases the estimate of the population variation. This attempts to eliminate the possibility of bias.

The estimated sample standard deviation is a measure of the spread of data about the mean. The standard deviation of the mean $\bar{x}$ is

$$\sigma_{\bar{x}} = \frac{\sigma_s}{\sqrt{n-1}}.$$

The above equation illustrates an important fact. The standard deviation doesn't change much, but the error on the mean improves dramatically! It goes as $\approx \frac{\sigma_s}{\sqrt{n}}$, where n is the number of measurements. As a rule of thumb, the range *R* of the random variable *x* can be roughly taken as $R \approx 4\sigma$. If Δ is the error that can be tolerated in the measurement, then the number of samples required to achieve the desired: $n \geq \sigma^2 / \Delta^2$. Then $\mu = \bar{x} \pm \Delta$

THE EXPERIMENTAL METHOD

Need for the Experiment

A well-planned, thoughtfully conducted, carefully analyzed and intuitively interpreted experiment is a must for a successful engineering work. This is indicated by ABET in student outcome 3(b) as: The Graduate of the Electrical and Computer Engineering at King Abdulaziz University is expected to demonstrate an ability to design and conduct experiments, analyze and interpret data. This means an engineer must

- **Design** the experiment from a problem description
- **Conduct** the experiment; use proper equipment and procedures to collect data
- **Analyze** and **interpret** data; write analysis reports on data collected from the field.

The assessment of this student outcome can be done by verifying the achievement of following indicators:

- **Identify** the constraints and assumptions for the experiment (cost, time, equipment), and apply them into experimental design.
- **Determine** proper data to collect and predicts experimental uncertainties.
- **Design** the experiment and report the results of the design.
- **Use** suitable measurement techniques to collect data.
- **Conduct** (or simulate) the experiment and report the results.
- **Select** and explain different methods of analysis (descriptive and inferential) and depth of analysis needed.
- **Use** proper tools to analyze data and self-explanatory graph formats to present the data.
- **Apply** statistical procedures where appropriate.
- **Verify** and validate experimental data.
- **Develop** mathematical models or computer simulations to correlate or interpret experimental results.
- **List** and discuss several possible reasons for deviations between predicted and measured results from an experiment, choose the most likely reason and justify the choice, and formulate a method to validate the explanation.

Experiments are carried out in various phases of an engineering project for various reasons including:

- To be familiar with test equipment, experimental set-up and procedures; *i.e.* to gain hands-on experience.
- To verify data available in literature.
- To obtain information that is not otherwise available.
- To test the proposed solution in the laboratory by controlling the experimental factors.
- To test the proposed solution in the field under naturally set conditions.

However, the experimental programs are costly and time-consuming, and require a lot of data processing during and after the experiments. Interpretation of the results obtained is a skill in itself. Hence, before you decide for experiments you have to double think on the reason for doing them. If you are absolutely sure that you need them, then you have to make a lot of preparations before you attempt.

Several experimental conditions must be satisfied before you decide for the experiment including:

- The system to be studied must be physically available to the experimenter;

- The problem to be studied should be possible to formulate with quantitative concepts that can be accurately formulated;
- There should be no political or social constraints to carry the experiments.

Design of the Experiment

An experiment is a series of trials that enable you to gather the required information. Careful planning is essential to obtain most of the information with least effort and cost. Important steps in an experiment is shown in Figure 3.12.

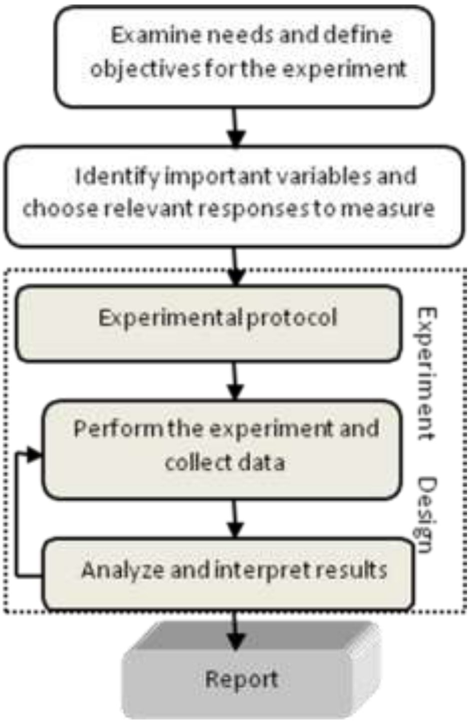


Figure 3.12 Important steps in an experiment

In an experimental work, firstly, you establish the need for the experiment and define the objectives for the experiments. Secondly, you identify the important variables (both independent variables and responses) and decide about the responses you want to measure. Then, the stages for the experiment design come. The last stage is the reporting of the results of the experiment.

An experimental protocol is very helpful in this respect. The protocol contains a list of equipments, devices and components to be used in the experiment, an experimental procedure that records the sequence of events during the experiment, and an indication of types of experimental errors and ways to avoid them. The next step is performing the experiment and collecting the data. Repetition is essential for reliability of the results and statistical analysis. The processing of data collected, error analysis and presenting the results are important ingredients for the success of the experiments.

Optimization

Experimental design has two meanings:

1. To plan an experiment and build possible equipment;
2. To deal with assigning the most suitable combination of factors under which the observation should be made.

The first one involves specialized measuring and statistical analysis techniques. It is partly dealt with throughout this work and there is a vast amount of literature about it. The second one requires optimization. It is exemplified by Figure 3.13 that shows a patient undergoing examination by radioisotopes. There are three essential factors to consider as:

1. The cost factor, C;

$$C = k_1 T \qquad (1)$$

where k_1 is a constant and T is the time the instrument in use.

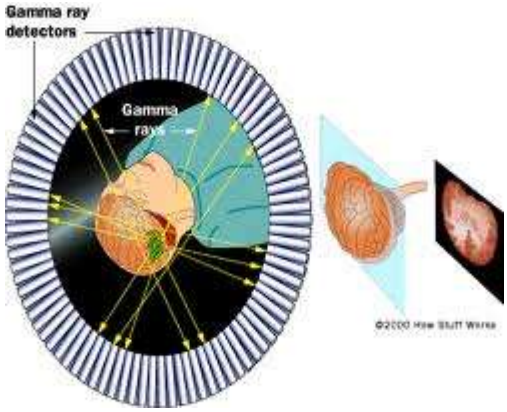


Figure 3.13 A patient undergoing examination by radioisotopes

2. Accuracy, $(1/\Delta)$ expressed in the uncertainty or error Δ:

$$\Delta/R = \left[2\sqrt{n}\right]^{-1}$$

where R is a constant of a particular instrument. $R \approx 4r$,
where r is the standard deviation and R is the range of measured variable in case of random variables. Hence, it can be rewritten as

$$\Delta = k_2/\sqrt{n} \qquad (2)$$

with k_2 a constant. n can be related to the time as

$$n = n_0 T \qquad (3)$$

where n_0 is a constant related to the original number of radioactive isotopes.

Therefore,

$$\Delta = k_2/\sqrt{n} = k_2/\sqrt{n_0 T} \qquad (4)$$

3. Damage factor, b

Gamma rays penetrating through the tissue may cause damage to the tissue. The damage is proportional to the total number of detected gamma particles. Thus:

$$b = k_3 n \qquad (5)$$

where k_3 is a constant.

Combining (2) and (5) and eliminating n yields:

$$b\Delta^2 = k_2 k_3 \qquad (6)$$

Safety is the most important aspect and $b \le b_0$. In this case, (6) can be rewritten as:

$$1/\Delta^2 = b_0 / k_2 k_3$$

The cost of the experiment in (1) can be related to the accuracy with the help of (4). It becomes

$$C = k_1 T = n_0 \left(k_1 / k_2 \right) \left(1/\Delta \right)^2 \qquad (7)$$

Important Reminder

Preferably use a hardbound notebook. Write down all you plan and you do. Never erase anything or discard a page by tearing it off. Rather, cross out what you don't want. Then, write down the steps you will follow in an experiment and even prepare a protocol. Remember that, an hour of hard work at the desk saves hours of frustrations in the laboratory. Also, hours of carefully planned experiments save the whole of the design from disasters.

Before you use any instrument, make sure that it is in working order, well calibrated and ready to use with all of its peripherals. If you are not very well-informed with any equipment or device, run a familiarization tests that yield known results before you attempt to use them in real experiments.

PROBLEMS

Review Questions

1. Why we need to make measurements?
2. What are the basic functions of a measuring instrument?
3. What do you understand from analysis of measured data?
4. What is the true value of a measurement and how it is established?
5. What is the accuracy of a measurement and what are the factors affecting it?
6. What is the precision of a measurement and how it differs from the accuracy?
7. What is the bias and how it effects the measurement?
8. What is the tolerance? Is it the result or precondition of a measurement?
9. What is the static calibration and how it is done?
10. What is the significant figure and how it is determined?
11. What is the gross error and how it can be eliminated?
12. What is the systematic error and how it can be minimized?
13. What is the random error and how it effects the measurement?

14. What are the errors that can be treated mathematically?
15. What is the arithmetic mean?
16. What is the significance of the standard deviation?
17. What specifies a normal (Gaussian) distribution?
18. What is the range of a variable and the probable error?
19. What determines the uncertainty in a digital readout?
20. How the uncertainty in an analog reading is specified?
21. How do you determine the total error based on the errors of component sources?
22. What is the limiting error?
23. How the population and sample statistics differ from each other?
24. What is the error of the mean and how it is effected by the sample size?

Solved Examples

1. A digital thermometer is used to measure the boiling point of water (100.0°C). The measurement is repeated 5 times and following readings are obtained: 99.9, 101.2, 100.5, 100.8, 100.1. Determine the accuracy, the precision and the bias of the thermometer.

T_{CTV} = 100.0°C; T_{AV} =(99.9 + 101.2 + 100.5 + 100.8 + 100.1)/5 = 100.5°C.

Accuracy = max of [(101.2 – 100.0), (100.0 – 99.9)] = 1.2°C; % acc. = 1.2%

Pr = max of [(101.2 – 100.5), (100.5 – 99.9)] = 0.7°C

Bias = T_{CTV} - T_{AV} = -0.5°C.

2. A digital voltmeter uses 4½ digit display (it can display up to 19999). It is used to measure a voltage across a standard cell whose value is 1.2341 volt 4 times and following readings are obtained: 1.2202, 1.2115, 1.2456, 1.2218. Determine the accuracy, the precision and the bias of the voltmeter.

CTV =1.2341 volt; V_{AV} = 0.25x(1.2202 + 1.2115 + 1.2456 + 1.2218) = 1.2248V.

The accuracy = |1.2341 – 1.2115| = 0.0226V; % accuracy = 1.83% ,

pr = max[(1.2456 – 1.22248), (1.2248 – 1.2115)] = 0.0208 V;

Bias = 1.2341 – 1.2248 = - 0.0093 V

3. A recently calibrated digital voltmeter is used to read a voltage and it consistently yields 75 volts. Another meter in the lab is also used five times to measure the same voltage and following readings are obtained: 77, 75, 74, 76, 77. For the second meter,

a. Find the absolute accuracy, relative accuracy and percentage accuracy.

The recently calibrated meter presumably reads the conventional true value. Therefore CTV = 75 V, yielding absolute accuracy = max {(77 - 75), (75 – 74)} = 2 volts,

The relative accuracy = 2/75 = 0.027, The % accuracy = 2.7%

b. Find the precision.

V_{AV} = (1/5)(77+75+74+76+77) = 75.8 volts. Pr = max {(77 – 75.8), (75.8 – 74)}=1.8V

c. Calculate the bias. Bias = $V_{CTV} - V_{AV}$ = 75 – 75.8 = - 0.8 volt.

4. The gain of the amplifier is defined in dB by: $G = 20\log_{10}(\frac{V_2}{V_1})$. Show that the uncertainty in the gain is given by: $(\omega_G)^2 = (20\log_{10}e)^2[(\frac{\omega_{V_1}}{V_1})^2 + (\frac{\omega_{V_2}}{V_2})^2]$

Hint: $\log_{10}a = (\log_e a)/(\log_e 10) = (\log_{10}e)(\log_e a)$, $\log_e a = \ln(a)$ and $d(\ln x)/dx = 1/x$. $\log_{10}e = 0.434$

$G = 20\log_{10}(\frac{V_2}{V_1}) = (20\log_{10}e)[\ln(V_2) - \ln(V_1)$; $\frac{\partial G}{\partial V_1} = -20\log_{10}e\frac{1}{V_1}$ and $\frac{\partial G}{\partial V_2} = 20\log_{10}e\frac{1}{V_2}$

yielding the uncertainty as defined above.

5. Five resistors are available, one of 20 Ω and four of 10 Ω each. The uncertainty of the 20 Ω resistor is 5% and that of each 10 Ω resistor is 10%. 3 possible connections using these resistors are shown below. Which one would you use to obtain a 30 Ω resistance with the least uncertainty? What is the uncertainty of this best connection?

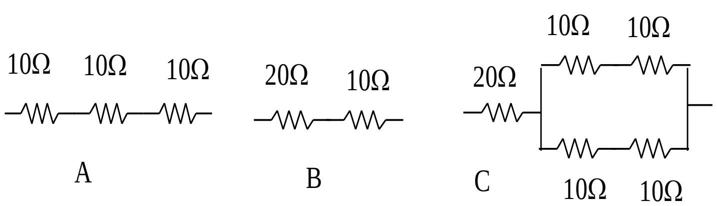


Figure solved example 5.

$\omega_{10\Omega}$ = 1 Ω; $\omega_{20\Omega}$ = 1 Ω; $(\omega_A)^2 = 3x(1)^2$ → ω_A = 1.73 Ω; $(\omega_B)^2 = (1)^2 + (1)^2 = 2$ → ω_B =1.414 Ω; in (C), R_P = 10 Ω ± ΔR_P = (20 ± 1.414)// (20 ± 1.414); $(\Delta R_P)^2 = 2x(1/4)^2x(1.414)^2$ → ΔR_P = 0.25 Ω. Yielding $(\omega_C)^2 = (1)^2 + (0.25)^2$ → ω_C =1.031 Ω , hence (C) has the least uncertainty.

6. The DC current in a resistance R = 10 kΩ ± 0.5% is measured to be I = 10 mA ± 1%. Find the power dissipated in this resistance with its uncertainty and limiting error.

$P = I^2R$; $\delta P/\delta I = 2IR$; $\delta P/\delta R = I^2$, $\therefore \overline{P} = 10x10^3x(10x10^{-3})^2 = 1\,W$

$(\Delta P)^2 = (2IR)^2(\Delta I)^2 + (I^2)^2(\Delta R)^2$, with ΔI = 10^{-4} A, ΔR = 50 Ω , $(\Delta P)^2$ = 4.25x10^{-4} and ΔP = 20.6 mW yielding $\%\omega_P$ = 2.06% and P = 1 W ± 2.06%

Limiting error = $\Delta P_m = \left|2IR\Delta I + I^2\Delta R\right| = 2x20x10^{-3}x10x10^3x10^{-4} + 10^{-8}x50 \cong 40mW$

7. A metallic resistance thermometer has a linear variation of resistance with temperature as $R = R_0[1+\alpha_0(T-T_0)]$. The resistance at T_0 =280K ± 0.01K is R_0 = 20 kΩ ± 0.1%, while at a temperature T the resistance R is R = 30 kΩ ± 0.1%. The coefficient α_0 = 0.00392/K.
   a. Write down an explicit expression for T.

$$R = R_0 + \alpha_0 R_0 T - \alpha_0 R_0 T_0 \rightarrow T = \frac{R-R_0}{\alpha_0 R_0} + T_0 = \frac{1}{\alpha_0}(\frac{R}{R_0} - 1) + T_0$$

   b. Show that the uncertainty ΔT in T is given by:

$$(\Delta T)^2 = (\Delta T_0)^2 + \frac{1}{\alpha_0^2}(\frac{R}{R_0})^2[(\frac{\Delta R_0}{R_0})^2 + (\frac{\Delta R}{R})^2]$$

First, we calculate the sensitivity of T to R, R_0, α_0, and T_0

$\frac{\partial T}{\partial R} = \frac{1}{\alpha_0 R_0}$, $\frac{\partial T}{\partial R_0} = -\frac{R}{\alpha_0 R_0^2}$, $\frac{\partial T}{\partial \alpha_0} = \frac{1}{\alpha_0^2}(1 - \frac{R}{R_0})$, and $\frac{\partial T}{\partial T_0} = 1$

$(\Delta T)^2 = (\Delta T_0)^2 + (\frac{1}{\alpha_0 R_0})^2(\Delta R)^2 + (\frac{R}{\alpha_0 R_0^2})^2(\Delta R_0)^2$ Reorganizing yields the answer.

   c. Calculate the nominal value of T and its uncertainty.

$\overline{T} = \frac{1}{0.00392}(\frac{30}{20} - 1) + 280 = 407.6K$; $(\Delta T)^2 = 10^{-4} + (\frac{1.5}{0.00392})^2(10^{-6} + 10^{-6}) = 0.29295$

yielding ΔT = ± 0.54K

   d. Find the static sensitivity $\frac{\partial R}{\partial T}$ of the thermometer.

$$\frac{\partial R}{\partial T} = \alpha_0 R_0 = 0.00392x20x10^3 = 78.4\,\Omega/K$$

   e. Calculate the maximum error in T.

$$\Delta T_m = \left|\Delta T_0\right| + \frac{1}{\alpha_0 R_0}\left|\Delta R\right| + \frac{R}{\alpha_0 R_0^2}\left|\Delta R_0\right| = 0.01 + \frac{1.5}{0.00392}(0.001 + 0.001) = 0.7753K$$

General Questions

True-False

Please answer the following True or False questions.

Question	True	False
Systematic errors can be eliminated by recalibrating the equipment		
Systematic errors can be eliminated by making multiple measurements		
Accuracy of a measurement is an indication of how close the reading is to the average value		
Accuracy of a measurement is an indication of total errors in the measurement		
The smallest incremental quantity that can be measured is the resolution		
The precision is an indicator of consistency in a set of measurements		
The result of 10.5+ 1.267 (with significant figures only) is 11.8		
Gross (human) errors can be treated mathematically		
The current in a 10-Ω resistor is measured as 0.25 A ±1%. The power dissipated by the resistor is 625 ± 12.5 mW.		

Multiple-Choice Questions

Please **choose and CICRLE** the most appropriate statement in the following questions

1. Gross (human) errors
    a. Are due to equipment failures
    b. Can be minimized by making multiple measurements
    c. Cannot be treated mathematically
    d. Do not affect the accuracy of the measurement
2. Resolution is
    a. An indicator of how close the reading to the true value
    b. The smallest incremental quantity that we can identify
    c. The difference between the minimum and maximum values of the measurement
    d. The total error in the measurement
3. Systematic errors
    a. Cannot be treated mathematically
    b. Can be eliminated by making multiple measurements
    c. Indicate the accuracy of the measurement

d. Are due to environmental factors upsetting the user and the equipment

4. Accuracy of a measurement is an indication of
    a. How far the reading is away from the average value
    b. How many digits we use to display the data
    c. How close the reading is to the conventional true value
    d. The smallest incremental quantity that we can identify
5. Precision is
    a. An indicator of how close the reading is to the true value
    b. The total error in the measurement
    c. An indicator of how close the reading is to the average value
    d. The smallest incremental quantity that we can identify
6. What is the result of 1.264+ 10.5 (use significant figures only)
    a. 12
    b. 11.8
    c. 11.7
    d. 11.764
7. Mathematical treatment of errors is possible for
    a. Systematic and random errors
    b. Human and systematic errors
    c. Human and random errors
    d. Errors that are small

General Questions

1. Define the following terms shortly:
    a. Random error
    b. Instrumental error
    c. Calibration error
    d. Environmental error
    e. Limiting error
2. A digital voltmeter has three ranges as 0 to 1.999V, 0 to 19.99V, and 0 to 199.9V. Determine:
    a. The resolution in volt in each range
    b. The uncertainty in reading in volts in each range
    c. Percentile error in measuring 1.5 V in each range
3. A resistor is measured by the voltmeter-ammeter method. The voltage reading is 123.4 V on the 250-V scale and the ammeter reading is 283.5 mA on the 500-mA scale. Both meters are guaranteed to be accurate within ±1% of full-scale reading. Calculate

a. The indicated value of the resistance
b. The expected error in the resistance
c. The limits within which you guarantee the result

4. Four capacitors are placed in parallel. The values are (in μF) 47.23, 2.35, 18.026 and 0.428, with an uncertainty of one digit in the last place. Find the total capacitor and express the result using significant figures only. Also prove your result using uncertainty analysis.
5. Two resistors have values R_1 = 47 Ω ± 2% and R_2 = 82 Ω ± 5% Calculate
   a. The magnitude of error in each resistor
   b. The limiting error in ohms and in percent when the resistors are connected in series
   c. The value of the equivalent resistor and expected error in percent when the resistors are connected in parallel.
6. The potential of an electrical power source is measured 12.47 volts by a recently calibrated digital voltmeter. Two other voltmeters are used in the lab to measure the same voltage by six different observers in a short interval of time and following results (in volts) are recorded:

Meter-1: 11.456, 11.324, 11.562, 11.243, 11.472, and 11.376

Meter-2: 12.45, 12.34, 12.67, 12.76, 12.21, and 12.54

   a. Determine the resolution of each meter in volt. Which one has a better resolution?
   b. Determine the accuracy and precision of each meter. How much is the bias in each meter? Which one is more precise? Which one is more accurate?
7. The following values were obtained from the measurements for a resistor in ohms: 220.2, 119.5, 221.1, 119.9, 220.0, 220.5, 119.8, 220.1, 220.4, and 119.8. Calculate
   a. The arithmetic mean
   b. The average deviation
   c. The standard deviation
   d. The probable error of the average of the ten readings.
8. A metallic resistance thermometer has a linear variation of temperature with resistance as $T = \frac{1}{\alpha_0}(\frac{R}{R_0} - 1) + T_0$. The temperature at R_0 = 5 kΩ ± 1% is T_0 =25°C ± 0.1°C, while at a T the resistance R is found to be R = 6 kΩ ± 1%. α_0 = 0.004/°C.
   a. Calculate the static sensitivity $\partial T / \partial R$ at R_0 of the thermometer.
   b. Calculate the nominal value of T.
   c. Show that the limiting error ΔT_m in T is given by: $\Delta T_m = |\Delta T_0| + \frac{1}{\alpha_0}\frac{R}{R_0}[\left|\frac{\Delta R_0}{R_0}\right| + \left|\frac{\Delta R}{R}\right|]$

d. Calculate the limiting error and uncertainty in T.

9. A digital thermometer is used to measure the boiling water whose temperature is 96.2°C. The measurement is repeated 5 times and following readings are obtained: 95.9, 96.2, 96.5, 95.8, 96.1. Determine the percentile accuracy, the precision and the bias of the thermometer.

10. The following values were obtained from the measurements for the line voltage in Jeddah: 125.2, 125.5, 126.1, 126.2, 126.0, 125.8, 125.7, 126.1, 126.3, and 125.6. Write down the formulas and calculate

    a. The arithmetic mean

    b. The standard deviation and the probable error of the average of the ten readings.

11. The boiling temperature of water is measured 15 times using two thermometers A and B, and the readings presented in the graph are obtained. Conventional value for the boiling temperature of water is 96.2°C.

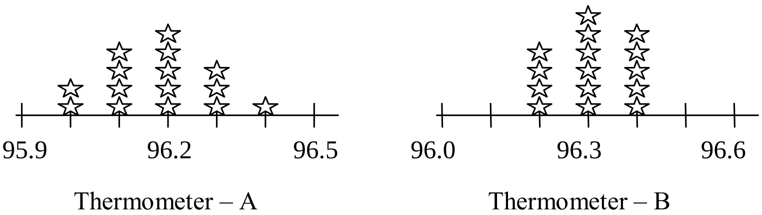


Figure problem 11.

    a. Which thermometer (A or B) is more precise, why?

    b. Calculate the percentage accuracy and bias of thermometer – A.

12. What is the addition of 12.5 and 1.364 with each having the last digit doubtful?

13. For the electronic counter show that the uncertainty in the period measurement can be reduced by a factor of $\frac{1}{\sqrt{N}}$ if the average of N time periods is taken. Hint: $T_{AV} = \frac{1}{N}(T_1 + T_2 + \cdots + T_N)$

    The T_i's are statistically independent, $T_i = T \pm \omega_T$,∀i

14. What is the systematic error, from where it comes and how it can be eliminated?

15. Three resistors are in series. The values are (in kΩ) 47.23, 2.205, and 180.2, with an uncertainty of one digit in the last place. Find the total resistor and express the result using significant figures only.

16. The potential of an electrical power source is measured 124.7 volts by a recently calibrated digital voltmeter. A voltmeter in the lab is used to measure the same voltage by six different observers in a short interval of time and following results (in volts) are recorded: 124.5, 123.4, 126.7, 127.6, 122.1, and 125.4. For the meter in the lab, determine the resolution in volt, the accuracy, the precision, and the bias?

17. Two resistors have values R1 = 56 Ω ± 5% and R2 = 120 Ω ± 2% Calculate

    a. The magnitude of error in each resistor

b. The limiting error in ohms and in percent when the resistors are connected in series.
c. The value of the equivalent resistor and expected error in percent when the resistors are connected in parallel.

18. There are 1500 chickens in a poultry farm. 15 chickens are randomly selected and weighted. The average value is 950 grams and the standard deviation is 60 grams.
    a. How much is the error expected in the average value?
    b. How many chickens we will have weighing between 890 grams and 1010 gram
    c. How many chickens must be weighted to reduce the error in the average value down to 5 grams?
19. What is the systematic error, from where it comes and how it can be eliminated?
20. Three resistors are connected in series. The values are (in kΩ) 1.205, 39.24 and 150.3, with an uncertainty of one digit in the last place. Find the total resistor and express the result using significant figures only.
21. The potential of a lithium-ion battery is measured 3.72 volts by a recently calibrated digital voltmeter. A voltmeter in the lab is used to measure the same voltage by six different observers in a short interval of time and following results (in volts) are recorded: 3.69, 3.72, 3.75, 3.67, 3.70, and 3.73. For the meter in the lab, determine the resolution in volt, the accuracy, the precision, and the bias?
22. A 5 mV signal is measured with a meter ten times resulting in the following sequence of readings: 5 mV, 6 mV, 9 mV, 8 mV, 4 mV, 7 mV, 5 mV, 7 mV, 8 mV, 11 mV ?
    a. What is the average measured value?
    b. What is the percentile accuracy of the meter?
    c. What is the precision of the meter?
    d. What is the bias (systematic error) of the meter?
23. A meter is rated at 8-bits and has a full-scale range of ±5 V. What is the measurement resolution of this meter?
24. A signal is to measured with a resolution of ±0.5 μV. How many bits of resolution are required by a meter having a ±1 V full-scale range?

BIBLIOGRAPHY

Further Reading

Useful Websites

MEASUREMENT OF ELECTRICAL QUANTITIES

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Illustrate principles of voltage and current measurements.
2. Discuss principles of moving coil instruments.
3. Describe the galvanometer and its use as a measuring instrument.
4. Describe the operation of MC based ammeters and voltmeters.
5. Devise MC based multi-range ammeters and voltmeters.
6. Demonstrate measurement of resistors and design of MC based ohmmeters and VOM meters.
7. Discuss the effect of instrument loading.
8. Calculate errors introduced by loading errors in ammeters and voltmeters.
9. Explain the defining features of AC and DC voltages.
10. Calculate the RMS and average values of AC waveforms.
11. Discuss means of obtaining DC equivalents of AC waveforms
12. Determine the form factors of AC waveforms and calculate the waveform errors.
13. Discuss the operational principles and use of clamp-on meters.
14. Discuss the need for true RMS meters and identify ways of realizing the true RMS measures.
15. Compare and contrast oscilloscopes, electronic counters and digital voltmeters as measuring instruments.
16. Illustrate principles of time and frequency measurements.
17. Discuss devices that are commonly used in electronic measuring instruments
18. Explain operation of counters in frequency, time-period and time-interval modes.
19. Calculate errors in measurements using counters.
20. Express the principles measurement of rotative speed.
21. Express the use, advantages and operation of the digital voltmeter (DVM).
22. Explain digitization of analog signals and the use of sample and hold circuits.
23. Explain the principles of operation of integrating and successive approximation type analog to digital converters and their applications in digital voltmeters.
24. Compare and contrast single slope and dual slope integration type digital voltmeters.
25. Discuss utilization of electrical energy and measurement of electric power.
26. Compare and contrast electricity measuring in resistive and reactive loads.
27. Compare and contrast analog multiplier based and digital sampling based electricity measuring devices.
28. Describe analog multiplication techniques TDM, Hall effect and transconductance as used in measuring the electrical power.
29. Describe the digital sampling type electricity measurement and state its advantages.

PRINCIPLES OF MEASUREMENTS

Electrical voltage and current are two important quantities in an electrical network. The voltage is the effort variable without which no current is available. It is measured across an electrical circuit element or branch of a circuit. The device that measures the voltage is the voltmeter. The current is the flow variable that represents net motion of the charged particles (electrons in solids, ions in a liquid) in a given direction. The product of the two yields the instantaneous electrical power. The ratio of the voltage to the current is the impedance.

The current is measured by an ammeter (also called an ampermeter). Ammeters are connected in series with the load to measure the current in the load. Eventually, the ammeters require breaking the current loop to place it into the circuit. The voltmeter connection is rather easy since it is connected without disturbing the circuit layout. Therefore, most electrical measurements require determination of the voltage rather than the current due the ease of measurement. Connections of ammeters and voltmeters are illustrated in Figure 4.1.

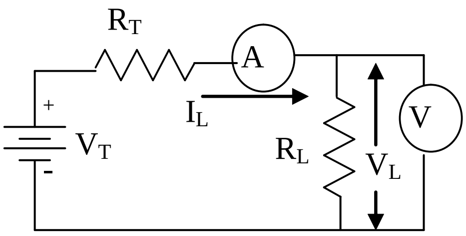


Figure 4. 1 Connections for an ammeter and a voltmeter

The current generates a magnetic field around the current carrying conductor. It is also possible to check out the size of the current by sensing the magnetic field strength. This is carried out by clamp-on type ammeters that will be shown later in the chapter. The electrical resistance of a circuit component is measured using an ohmmeter that applies a voltage across and determines the current passing through the component.

Voltmeters and ammeters display the results as deflections of dials on calibrated screens or numerical values on alphanumeric displays as illustrated in Figure 4.2. Both types are connected to the circuit via sensing leads and indicate the voltage. However, their internal operations and user interfaces are different. The first type forms the analog meters that will be discussed firstly in this chapter. The second category will be discussed later in the chapter under the title of digital voltmeters. Many measuring instruments use operational amplifiers and similar electronic devices for signal amplification and processing. A short theory about the operational amplifiers is given in Appendix-B.

Figure 4. 2 Analog and digital voltmeters

MOVING COIL IN MEASURING INSTRUMENTS

Magnetic field generated by a current carrying conductor and force exerted on such a conductor as it is inserted in a magnetic field were discussed in Chapter 2 and illustrated by Figures 2.4 – 2.8. The magnitude of the force on the conductor depends on the magnitude of the current which it carries. The force is a maximum when the current flows perpendicular to the field and it is zero when it flows parallel to the field as illustrated in diagrams A and B respectively in Figure 4.3.

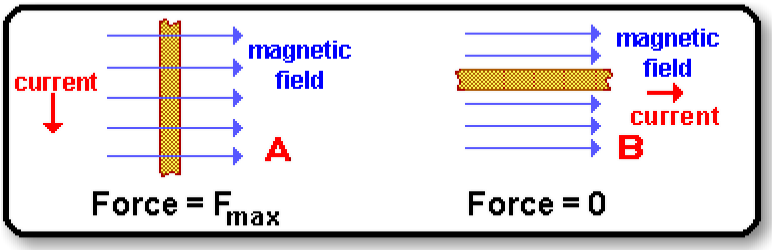


Figure 4. 3 Force exerted on a current carrying conductor in a magnetic field

Balancing the Electromagnetic Torque by a Spring Torque

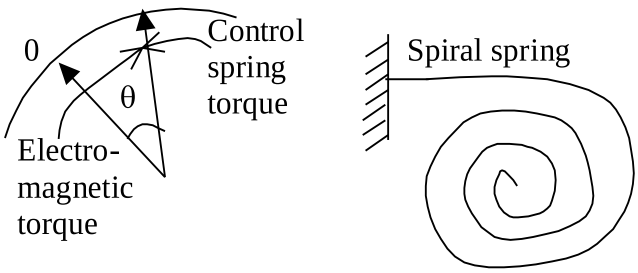


Figure 4. 4 Compensating electromagnetic torque by the torque of control springs

The coil is suspended in a uniform magnetic field and rotates due to the electromagnetic torque T_{EM}. This torque is opposed by spiral control springs (Figure 4.4) mounted on each end of the coil. The torque put forth on the control spring is $T_{SP} = k\theta$ where θ is the angle of rotation (degrees) and k is spring constant (N-m/degree). At equilibrium (at balance)

$T_{EM} = T_{SP}$ yielding $NBIA = k\theta$

The equation can be rearranged for θ,

$$\theta = \left(\frac{NAB}{k}\right) I = SI$$

where S is the sensitivity

$$S = \frac{\partial\theta}{\partial I} = \left(\frac{NAB}{k}\right)\left(\frac{\deg ree}{Amp}\right)$$

which is constant for a specific equipment provided that B is constant. In this respect, the moving coil instrument can be considered as a transducer that converts the electrical current to angular displacement. The linear relation between θ and I indicate that we have a linear (uniform) scale as shown in Figure 4.5.

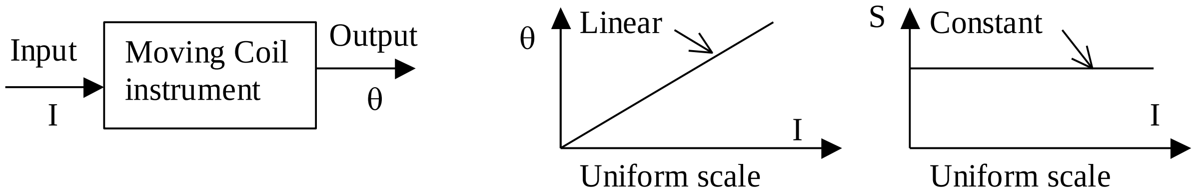


Figure 4. 5 Model of a moving coil instrument

Examples 4.1

A moving coil has following parameters: Area A= 2 cm^2, N=90 turns, B= 0.2 Tesla, coil resistance = 50 Ω, current I= 1 mA. Calculate:

a. Power dissipated by the coil;

✓ $P = I^2 x R_m$ = 50 μW.

b. The electromagnetic torque established;

✓ $T_{EM}=NBAI$ = 90x0.2x2x10^{-4}x10^{-3} = 3.6x10^{-6} N-m

c. Assume that the electromagnetic torque of the coil is compensated by a spring torque and the spring constant k = 3.6x10^{-8} N-m/degrees. Find the angle of deflection of the coil at equilibrium.

✓ $\theta = T_{EM} / k$ = 100 °

Example 4.2

A moving coil instrument has the following data: # of turns of the coil = 100, width of the coil = 2 cm, length of the coil = 3 cm, flux density in the air gap = 0.1 Wb/m^2 (Tesla). Calculate the deflection torque when carrying a current of 10 mA. Also calculate the deflection (angle) if the control spring constant is 20x10^{-7} N-m/degree.

✓ A = 6 cm^2 and T_{EM} = 60x10^{-6} N-m

✓ $\theta = T_{EM} / k$ = 30 °

The D'Arsonval Meter Movement

A Permanent Magnet Moving Coil (PMMC) meter that consists of a moving coil suspended between the poles of a horseshoe type permanent magnet is called the D'Arsonval meter as shown in Figure 4.6. It is an analog electromechanical transducer that produces a rotary

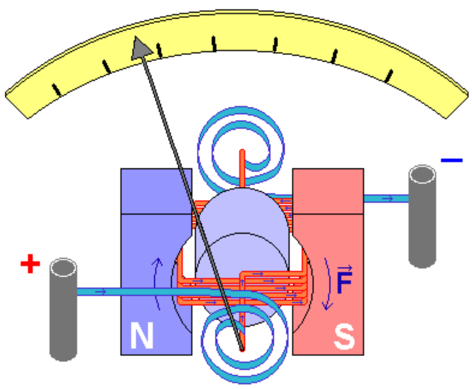


Figure 4. 6 The basic PMMC meter

deflection of some type of pointer in response to electric current flowing through its coil. Shoe poles are curved to have a uniform magnetic field through the coil. The coil is suspended between to pivots and can rotate easily. Iron core and permanent magnet are fixed. Coil axes and pointer is the moving parts. The principle of operation is similar to the general moving coil instrument explained above. There are mechanical stops at both ends to limit the movement of the pointer beyond the scale. The amount of the DC current that causes maximum allowable deflection on the screen is called the full-scale deflection current I_{FSD} and it is specified for all meters by the manufacturer.

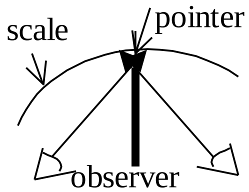


Figure 4. 7 The Parallax error

The moving coil instrument provides a unidirectional movement of the pointer as the coil moves against the control springs. It can be used to display any electrical variable that can be converted to a DC current within the range of I_{FSD}. The screen is calibrated in a curvilinear fashion it has a mirror-backed scale to identify the position of the pointer. The reading must be done under reasonable lighting conditions and just above the pointer. Otherwise, there will be parallax errors in the reading as shown in Figure 4.7. Under the best measurement conditions, the reading can be interpreted by the user within ± ½ small (minor) scale division.

The Galvanometer

The galvanometer is a moving coil instrument in which position of the pointer can be biased so that it stays in the middle of the scale to indicate zero current as shown in Figure 4.8. It can deflect in both directions to show the negative and positive values. It is commonly used in bridge measurements where zeroing (balancing – null) of the display is important for a very accurate measurement of the

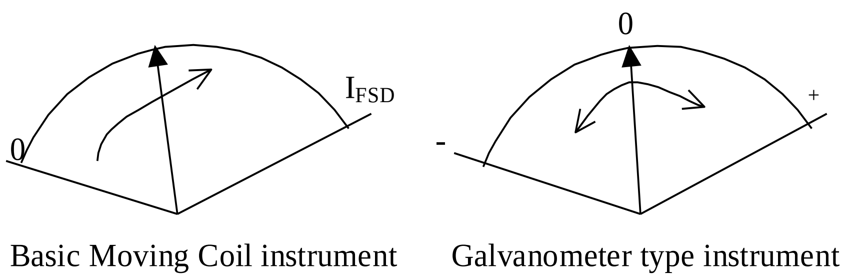


Figure 4. 8 Basic moving coil and galvanometer type displays

variable. It is also used in mechanical recorders in which a pen assembly is attached to the tip of the pointer and it marks on the paper passing underneath.

Neither the standard moving coil instrument nor the galvanometer can be used for AC measurement directly since the AC current produces positive deflection with the positive alternate and negative deflection with the negative alternate. Thus, a stable position on the scale can't be obtained to indicate the magnitude of the current.

MC BASED MEASURING INSTRUMENTS

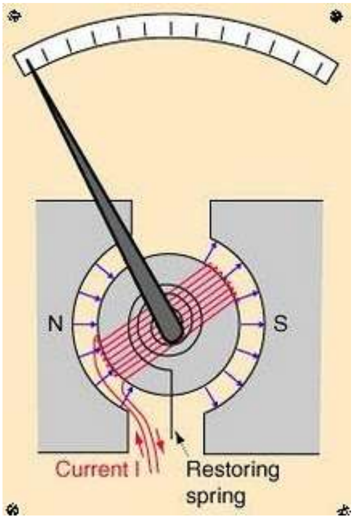


Figure 4. 9 A simplified view of the PMMC meter

MC in Analog Electrical Measuring Instruments

Figure 4.9 shows another simplified illustration of a PMMC meter. The standard MC instrument indicates positive DC currents (I_{MC}) as deflection on the scale. The galvanometer displays both positive and negative currents. The moving coil is usually made up of a very thin wire. The maximum current that gives full-scale deflection I_{FSD} is in the order of 0.1 to 10 mA and coil resistance R_{MC} 10 to 1000 Ω. The maximum deflection angle is about 100°. The current through the moving coil I_{MC} is limited by the I_{FSD}. A voltage drop $V_{MC} = I_{MC}R_{MC}$ occurs across the coil. The moving coil can represented by the full-scale deflection current I_{FSD} and coil resistance R_{MC} as shown in Figure 4.10.

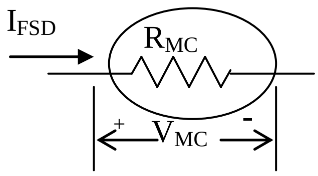


Figure 4. 10 Model of MC based instrument

Basic DC Ammeter (Ampermeter)

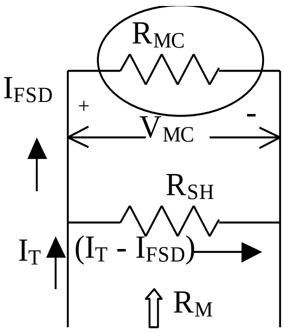


Figure 4. 11 DC Ammeter

The current capacity of the meter can be expended by adding a resistor in parallel with the meter coil as shown in Figure 4.11. The input current is shared between the coil resistance R_{MC} and the parallel resistance that is called the shunt R_{SH}. As the maximum input current I_T flows in, the coil takes I_{FSD} and remaining (I_T - I_{FSD}) is taken by the shunt resistor. Voltage developed across the meter is

$$V_{MC} = I_{FSD}R_{MC} = (I_T - I_{FSD})R_{SH}$$

The meter resistance R_M seen between the input terminals is

$$R_M = \frac{V_{MC}}{I_T} = R_{MC} // R_{SH}$$

Example 4.3

Calculate the multiplying power of a shunt of 20 Figure 4. 11 0 Ω resistance used with a galvanometer of 1000 Ω resistance. Determine the value of shunt resistance to give a multiplying factor of 50.

I_{fsd}x1000 = ($I_T - I_{fsd}$)x200 yielding $I_T = 6xI_{fsd}$.

For I_T=$50xI_{fsd}$, 1000xI_{fsd}=(50-1)xI_{fsd}xR_{sh} yielding R_{sh} =1000/49 = 20.41 Ω

Multi-Range Ammeters

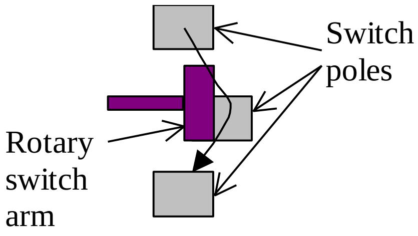


Figure 4. 12 Make-before-break type switch

The parallel resistance (shunt) can be changed to suit different full-scale current requirements as indicated in the previous example. The function can be accommodated by using a set of resistors and selecting them one by one. The switch however must be of make-before-break type (Figure 4.12) that makes the contact with the new position before it breaks the old connection. This eliminates the chance of forcing the full input current through the moving coil during changing the position of the switch.

Example 4.4

Design a multi-range DC ammeter using the basic movement with an internal resistance R_{MC}= 50 Ω and full-scale deflection current I_{MC}= I_{FSD}= 1 mA. The ranges required 0-10 mA, 0-50 mA, 0-100 mA and 0-500 mA as illustrated in Figure 4.13. V_{MC} = I_{MC}xR_{MC} = 50 mV

- For range-1 (0-10 mA) R_{SH1}= 50/9 =5.56 Ω
- For range-2 (0-50 mA) R_{SH2}= 50/49 =1.02 Ω
- For range-3 (0-100 mA) R_{SH3}= 50/99 =0.505 Ω
- For range-4 (0-500 mA) R_{SH4}= 50/499 =0.1 Ω

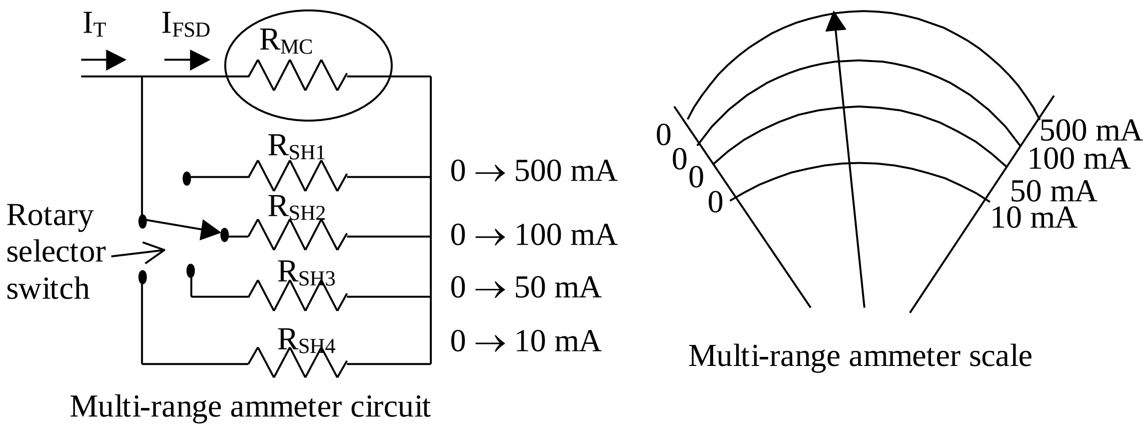


Figure 4. 13 A multi-range ammeter circuit and scale for example 4.4

Example 4.5

Design a multi-range DC ammeter using the basic movement with an internal resistance R_{MC}= 50 Ω and full-scale deflection current I_{MC}= I_{FSD}= 10 mA. The ranges required 0-0.1 A, 0-1 A, 0-10 A and 0-100 A.

$V_{MC} = I_{MC} x R_{MC}$ = 500 mV

- For range-1 (0-0.1 A) R_{SH1}= 500/90 = 5.56 Ω
- For range-2 (0-1 A) R_{SH2}= 0.5/0.99 = 0.505 Ω
- For range-3 (0-10 A) R_{SH3}= 0.5/9.99 = 0.05 Ω
- For range-4 (0-100 A) R_{SH4}= 0.5/99.99 = 0.005 Ω

A Basic DC Voltmeter

The moving coil can be used as a voltmeter by adding a series resistance R_S as illustrated in Figure 4.14. The input voltage is divided between the coil resistance R_{MC} and R_S. Current passing through both resistors is I_{MC} which is limited by the full-scale deflection current I_{FSD} of the coil. The full-scale input voltage

$$V_M = I_{FSD}(R_S+R_{MC})$$

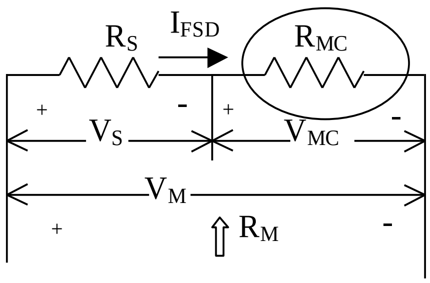


Figure 4. 14 Basic DC voltmeter

The input impedance seen is: $R_M = R_S + R_{MC}$

However, with R_S>>R_{MC}, R_M is approximately equal to R_S and $V_M \cong I_{FSD} R_S$.

Example 4.6

The coil of a moving coil voltmeter is 4 cm long and 3 cm wide and has 100 turns on it. The control spring exerts a torque of $2.4x10^{-4}$ N-m when the deflection is 100 divisions on the full scale. If the flux density of the magnetic filed in the air-gap is 0.1 Wb/m^2, estimate the resistance that must be put in series with the coil to give one volt per division. The resistance of the voltmeter coil may be neglected.

$T_{EM} = T_{SP} \Rightarrow 2.4x10^{-4} = 100x0.1x12x10^{-4} x I_{FSD} \Rightarrow I_{FSD}$ =20 mA. Therefore, current per division is 0.2 mA.

Assuming that R_{MC} is negligibly small compared to R_S : R_S = 5 kΩ

Example 4.7

A moving coil instrument gives full-scale deflection of 10 mA when the potential difference across its terminals is 100 mV. Calculate:

The shunt resistance for a full scale corresponding to 100 mA;

R_{SH} = 100 / 90 = 1.11 Ω

The resistance for full scale reading with 1000 V;

R_{MC} = 100 /10 = 10 Ω; $R_S + R_{MC}$ = (1000 / 10) kΩ = 100 kΩ yielding R_S = 100 kΩ (R_{MC} is negligible)

The power dissipated by the coil and by the external resistance in each case.

Power dissipated by the coil, $P_C = I_M^2 x R_{MC}$ = 1 mW;

$P_{SH} = V_M^2/R_{SH}$ = 9 mW

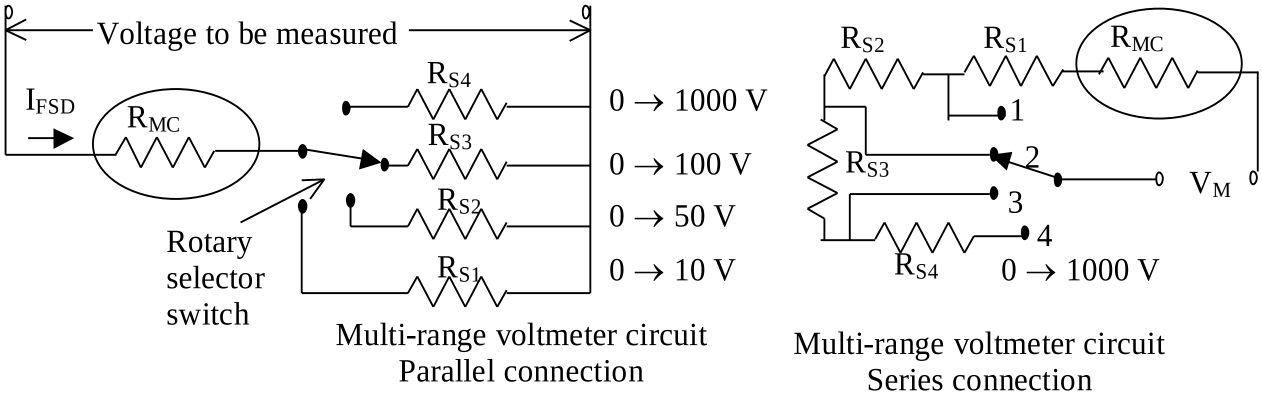


Figure 4. 15 Parallel and series resistance connections for a multi-range voltmeter

$P_S = V_M^2/R_S$ = 10 W.

Multi-Range Voltmeters

The series resistance can be changed to suit different full-scale voltage requirements as shown in Figure 4.15. Resistors are organized either in parallel fashion (conventional connection) as in the case of ammeter and selecting them one by one or all connected in series like a voltage divider (modified connection). The switch however must be of break-before-make type (Figure 4.16) that

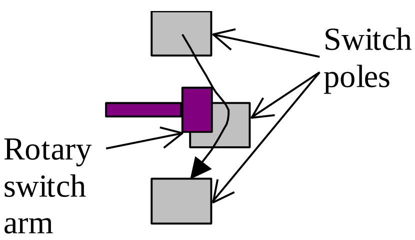


Figure 4. 16 Break-before-make type switch

breaks the contact with the old position before it makes it with the new position. This eliminates the chance of forcing a current larger than the full-scale current through the moving coil during changing the position of the switch.

The resistors are also called the multiplier resistors. Resistance seen by the input terminals of the device $R_M = V_M/I_{FSD}$ and written on the face of the scale as Ω/V. The contribution of the coil resistance R_{MC} can be ignored if it is too small compared to R_M. Following examples illustrate the selection of multiplier resistors.

Example 4.8

A multi-range DC voltmeter is designed using a moving coil with full-scale deflection current 10 mA and coil resistance 50 Ω. Ranges available: 0 – 10V, 0 – 50V, 0 – 100V, 0 - 1000V. Determine the multiplier resistors and input resistance of the meter using:

- Conventional connection
- Modified connection

In conventional connection, resistors are selected one-by-one to satisfy,

$V_M = I_{FSD} (R_{MC} + R_S) = V_{MC} + I_{FSD}R_S$ where V_M is the full-scale voltage of the selected range. V_{MC} = (10 mA)(50Ω) = 0.5V. Hence, $R_S = (V_M - 0.5)/10$ kΩ. Meter resistance seen between the input terminals is $R_M = R_{MC} + R_S$

- Range 1 (0 – 10V): R_{S1} = 9.5/10 = 0.95 kΩ = 950 Ω; R_{M1} = 950 Ω + 50 Ω = 1000 Ω
- Range 2 (0 – 50V): R_{S2} = 49.5/10 =4.95 kΩ; R_{M2} = 4.95 kΩ +0.05 kΩ = 5 kΩ
- Range 3 (0 – 100V): R_{S3} = 99.5/10 =9.95 kΩ; R_{M3} = 9.95 kΩ +0.05 kΩ = 10 kΩ
- Range 4 (0 – 1000V): R_{S4} = 999.5/10 =99.95 kΩ; R_{M4} = 99.95 kΩ +0.05 kΩ = 100 kΩ

For the alternative modified arrangement, the resistor for the lowest range is determined and others calculated as added to the total of the previous value. The total resistance seen from the input in all ranges will be the same as those in the previous case. Resistors between stages can be computed as $R_{Sn} = R_{Mn} - R_{M(n-1)}$

- Range 1 (0 – 10V): R_{M1} = 1000 Ω; R_{S1} = 1000 Ω - 50 Ω = 950 Ω
- Range 2 (0 – 50V): R_{M2} = 5 kΩ; R_{S2} = 5 kΩ - 1 kΩ = 4 kΩ;
- Range 3 (0 – 100V): R_{M3} = 10 kΩ; R_{S3} == 10 kΩ - 5 kΩ = 5 kΩ;
- Range 4 (0 – 1000V): R_{M4} = 100 kΩ; R_{S4} = 100 kΩ - 10 kΩ = 90 kΩ;

Example 4.9

A basic D'Arsonval meter movement with an internal resistance R_{MC}= 100 Ω, full scale current I_{FSD}= 1 mA, is to be converted into a multi-range DC voltmeter with ranges 0-10 V, 0-50 V, 0-250 V and 0-500 V. Find the values of multiplier resistors using the potential divider arrangement.

Four resistors R_{S1}-R_{S4} are added in series with R_{MC}.

- In the first range (0-10 V) only R_{S1} is used and the maximum voltage drop on R_{S1} is 10-0.1=9.9 V. Thus, R_{S1} = 9.9V/1mA = 9.9 kΩ
- In the 2nd range (0-50 V) $R_{S1}+R_{S2}$ is used and the maximum voltage drop on R_{S2} is 50-10= 40 V. Thus, R_{S2} = 40V/1mA = 40 kΩ
- In the 3rd range (0-250 V) $R_{S1}+R_{S2}+R_{S3}$ is used and the maximum voltage drop on R_{S3} is 250-50= 200 V. Thus, R_{S3} = 200V/1mA = 200 kΩ
- In the 4th range (0-500 V) $R_{S1}+R_{S2}+R_{S3}+R_{S4}$ is used and the maximum voltage drop on R_{S4} is 500-250= 250 V. Thus, R_{S4} = 250V/1mA = 250 kΩ

Ohm and VOM Meters

The Analog Ohmmeter

Analog ohmmeter can be designed simply by adding a battery and a variable resistor in series with the moving coil instrument as shown in Figure 4.17. The unknown resistance is connected to the terminals of the device to complete the electrical circuit. The output terminals are shorted together

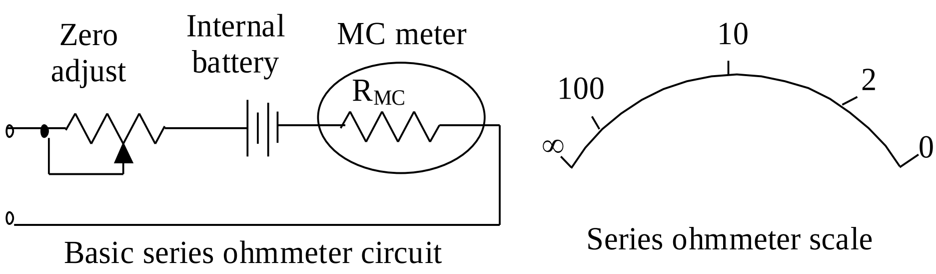


Figure 4. 17 Circuit and scale of a basic ohmmeter

with the leads (wires) used in connecting the external resistor. The variable resistance is adjusted until the full-scale deflection current passes through the coil. This is marked as the "0" resistance. When the leads are separated from each other, no current flows indicating an open-circuit which means "infinite - ∞" resistance. Hence, the scale is non-linear with resistance increases on the right side (opposite to ammeter). Multi-range ohmmeters can be obtained by combining the circuits of a series ohmmeter and a multi-range ammeter.

The VOM Meter

The functions of ammeter, voltmeter and ohmmeter can be combined in a multipurpose meter called a VOM (volt-ohm-milliampere) meter, or shortly “the VOM”. It has several multiple scales, usually color-coded in some way to make it easier to identify and read. Generally, it has a single multipurpose switch to select the function and the range.

Example 4.10

A moving coil has 100 turns, 5 cm^2 coil area, and air-gap magnetic flux density of 0.1 Tesla (Wb/m^2). The control spring exerts a torque of $5x10^{-6}$ N-m at the full-scale deflection of 90°. The potential difference across the coil terminals at the full-scale deflection is 100 mV. Using the above movement, design a multi-range DC ammeter with ranges 0-50 mA, 0-1 A and multi-range DC voltmeter with ranges 0-10 V and 0-200 V.

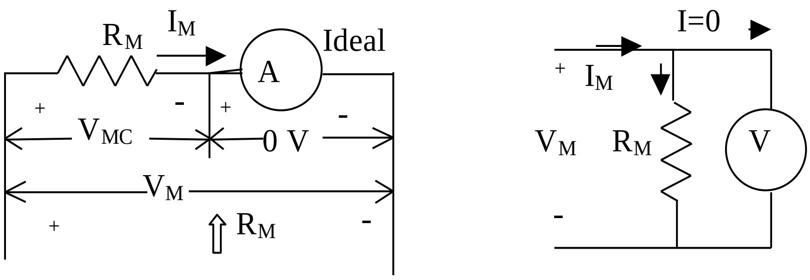
- $I_{FSD}=T_{SP}/NBA$ = 1 mA, therefore $R_{MC}= V_{MC} / I_{FSD}$ =100 Ω
- For ammeter ranges: R_{SH1}= 100 mV/ (50-1) mA = 2.04 Ω and R_{SH2} = 100/999 = 0.1 Ω
- For voltmeter ranges: R_{S1} = (10-0.1)V/1mA = 9.9 kΩ and R_{S2} = 199.9 kΩ

LOADING ERRORS

Instrument Loading

All measuring instruments draw energy from the source of measurement. This is called “the loading effect of the instrument”. Hence, all measurements include errors due to instrument loading. If the energy taken by the instrument is negligibly small compared to the energy exists in the source (of course of type measured), then the measurement is assumed to be close to perfect, and the loading error is ignored.

Ideal ammeter has zero internal resistance and no voltage across it. Ideal voltmeter has infinite internal (meter) resistance and draws no current from the circuit. The practical ammeter can be represented by an ideal ammeter with added series resistance that represent the meter resistance. Similarly, the practical voltmeter can be represented by an ideal voltmeter in parallel with the meter resistance. These two models are illustrated in Figure 4.18.



Practical ammeter Practical voltmeter

Figure 4. 18 Representations of practical ammeters and voltmeters

Loading Errors in Ammeters

Any electrical circuit can be modeled by a voltage source V_T and a series resistance R_T. The circuit is completed when the load resistance R_L is connected across the output terminals and a load current R_L flows through the load. An ammeter can be placed in series with the load to measure this current as shown in Figure 4.19. Current in the circuit can be calculated as

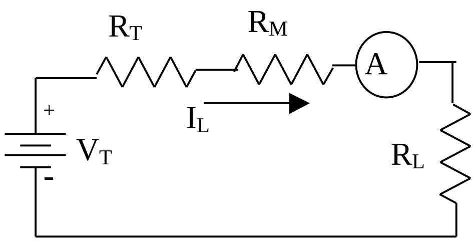


Figure 4. 19 Ammeter loading

$$I_L = \frac{V_T}{R_T + R_L + R_M}$$

In ideal condition, $R_M = 0$ and the true value of the current is

$$I_{LT} = \frac{V_T}{R_T + R_L}$$

The error is the difference between the measured value and the true value, and generally expressed as the percentile error which is:

$$\% loading error = \frac{measured value - true value}{true value} x100$$

Hence, the loading error due to the ammeter can be found as:

% loading error for ammeter = $\dfrac{\dfrac{V_T}{R_T + R_L + R_M} - \dfrac{V_T}{R_T + R_L}}{\dfrac{V_T}{R_T + R_L}} x100 = \dfrac{-100 R_M}{R_T + R_L + R_M}$

Loading error can be ignored if $R_M \ll (R_T + R_L)$ which is satisfied in most applications.

Loading Errors in Voltmeters

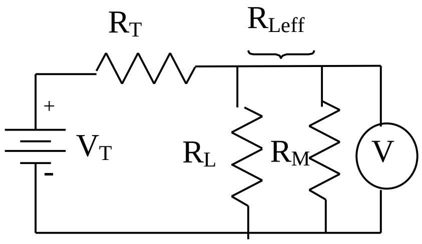


Figure 4. 20 Voltmeter loading

In voltage measurement, the meter is connected in parallel with load resistor as shown in Figure 4.20. The true value of the voltage across the resistor is (without the meter)

$$V_{LT} = \frac{V_T R_L}{R_T + R_L}$$

As the meter is connected, R_M becomes in parallel with R_L and effective load resistance becomes

$$R_{Leff} = \frac{R_L R_M}{R_L + R_M}$$

$R_{Leff} \cong R_L$ if $R_M >> R_L$. The voltage measured by the meter is

$$V_L = V_{Lind} = \frac{V_T \dfrac{R_L R_M}{R_L + R_M}}{R_T + \dfrac{R_L R_M}{R_L + R_M}}$$

$$\% loading error = \frac{V_{Lind} - V_{LT}}{V_{LT}} x100$$

Examples 4.11

A 150-V DC voltage source is coupled to a 50 kΩ load resistor through a 100 kΩ source resistance. Two voltmeters (A) and (B) are available for the measurement. Voltmeter-A has a sensitivity 1000 Ω/V, while voltmeter-B has a sensitivity 20000 Ω/V. Both meters have 0 – 50 V range.

- Calculate reading of each voltmeter.
- Calculate error in each reading expressed in a percentage of the true value.

$$V_{LT} = \frac{150}{(100+50)} x50 = 50\,V$$

Input resistance of voltmeter-A = sensitivity x range = (1000 Ω/V)x(50 V) = 50 kΩ and the effective value of the load resistance is 50//50 = 25 kΩ

Voltage indicated by voltmeter-A; $V_{LA} = \dfrac{150x25}{100+25} = 30\text{V}$

% age loading error = $\dfrac{30-50}{50}x100 = -40\%$

Input resistance of voltmeter-B = (20000 Ω/V)x(50 V) = 1000 kΩ and the effective value of the load resistance is 50//1000 = 48 kΩ

Voltage indicated by voltmeter-B; $V_{LB} = \dfrac{150x48}{100+48} = 48.5$ V

% age loading error = $\dfrac{48.5-50}{50}x100 = -3\%$

Example 4.12

A voltmeter has a resistance of 20 kΩ/V is used to measure the voltage on the circuit shown on a 0 - 10 V range. Find the percentage loading error.

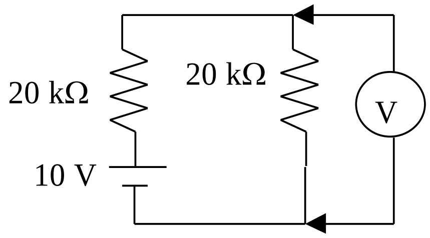


Figure for example 4.12

V_{TRUE} = 10x20/40 = 5 V. With R_M = 200 kΩ, the effective load resistance R_{Leff} = (400/22) = 18.18 kΩ. Therefore, V_{MEAS} = 10x18.18/38.18 = 4.76 V.

% loading error can be found as: %error = 100x(4.76 – 5)/5 = -4.8%

Example 4.13

A generator produces 100 volts DC and has an internal resistance of 100 kΩ as shown in the figure. The output voltage is measured using several voltage indicating devices. Calculate the output voltage and the percentage loading error for each of the following cases:

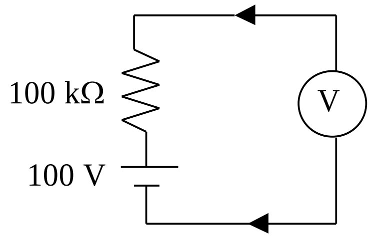


Figure for example 4.13

- An ideal voltmeter (Ri →∞) Vo = 100 V, Error = 0 %
- A digital voltmeter with Ri = 10 MΩ; Vo = 100x10/10.1 = 99 volts, % error = -1%
- An oscilloscope (Ri = 1 MΩ); Vo = 100x1/1.1 = 90.9 volts, % error = -9.1%
- A moving coil type analog voltmeter with 1 kΩ/V in 0 – 100 volt range
- Meter resistance is 100x1 kΩ = 100 kΩ, yielding Vo =50 volts, % error = 50 %

Example 4.14

A D'Arsonval movement gives full-scale deflection of 1 mA when a voltage of 50 mV is applied across its terminals. **Calculate** the resistance that should be added in series with this movement to convert it into a 0 – 100 V voltmeter. The above 0 – 100 V voltmeter is used to measure the voltage across the 10 kΩ resistor in the circuit shown. **Determine** the percentage loading error.

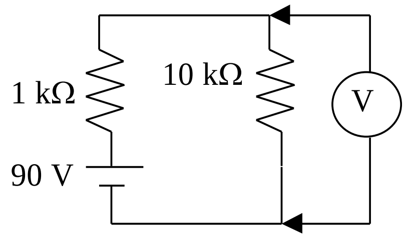


Figure for example 4.14

- Meter coil resistance R_M= 50 mV / 1 mA = 50 Ω and it's effect can be ignored in finding the series resistance of the voltmeter. Then, R_S= 100 V / 1 mA = 100 kΩ.
- True value of the voltage on the 10 kΩ resistance (without voltmeter loading) V_{true}= (10/11)x90 = 81.82 V
- With the voltmeter connected, 10 kΩ resistance will experience a 100 kΩ meter resistance in parallel yielding 9.09 kΩ at the output. The measured output voltage becomes: V_M = 90x (9.09/10.09) = 81. 08 V. The % error = 100x(81.08 - 81.82)/81.82 = - 0.9 %

AC VOLTMETERS

The voltmeter based on the permanent magnet moving coil (PMMC or D'Arsonval) and digital voltmeter that will be discussed later cannot be directly used to measure the alternating voltages. When measuring the value of an alternating current signal it is often necessary to convert the signal into a direct current signal of equivalent value (known as the root mean square, RMS value). This process can be quite complex. Most low cost instrumentation and signal converters carry out this conversion by rectifying and filtering the signal into an average value and applying a correction factor. Hence, we can classify the AC voltmeters in two broad categories as the averaging and true RMS types.

Average and RMS Values

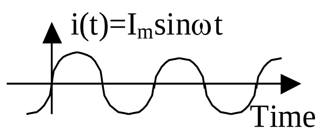


Figure 4. 21 Alternating current (AC) waveform

The moving coil instrument reads the average of an AC waveform. The average of the current waveform i(t) shown in Figure 4.21 is:

$$I_{AV} = \frac{1}{T}\int_0^T I_m \sin\omega t dt = 0$$

where T is the period and ω = $2\pi/T$ = radial frequency (rad/sec).

However, if this current is applied to a resistor R, the instantaneous power on the resistor $p(t) = i^2(t)R$

The average power over the period T becomes:

$$P_{AV} = \frac{R}{T}\int_0^T I_m{}^2 \sin\ \omega t dt = \frac{I_m{}^2 R}{2}$$

Hence, the average power is equivalent to the power that would be generated by a DC current called the effective current that is

$$I_{eff} = I_{RMS} = \sqrt{\frac{1}{T}\int_0^T i^2(t)dt} = \frac{I_m}{\sqrt{2}} = 0.707 I_m$$

Due to squaring, averaging (mean) and square-rooting operations, this is called the "RMS." value of the current and I_{RMS} is the true value of the current that we want to measure. The averaging time must be sufficiently long to allow filtering at the lowest frequencies of operation desired. Hence, in electrical terms, the AC RMS value is equivalent to the DC heating value of a particular waveform—voltage or current. For example, if a resistive heating element in an electric furnace is rated at 15 kW of heat at 220 V AC RMS, then we would get the same amount of heat if we applied 220 V of DC instead of AC.

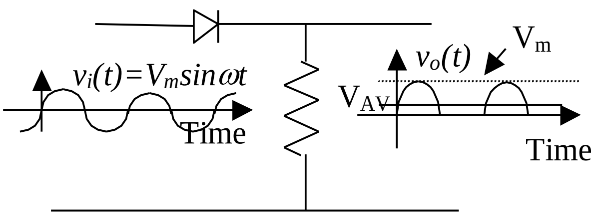


Figure 4. 22 AC to DC conversion

If the voltage is applied to the resistor through a diode as shown in Figure 4.22, the negative half cycle is chopped off since the diode can conduct current only in positive direction. This is called the half-wave rectifier. The average value of the current in the resistor becomes:

$$V_{AV} = \frac{1}{T}\int_0^{T/2} V_m \sin\omega t dt = \frac{V_m}{\pi} = 0.318 V_m$$

The Full-Wave Rectifier

The half-wave rectifier is used in some voltmeters, but the mostly adapted one uses the full wave rectifier shown in Figure 4.23. Here, a bridge-type full-wave rectifier is shown. For the + half cycle the current follows the root ABDC. For the – half cycle root CBDA is used. The current through the meter resistor R_m is the absolute value of the input current as shown in the inset. The voltage waveform on the meter resistance R_m has the same shape as the current. The average value of the voltage becomes:

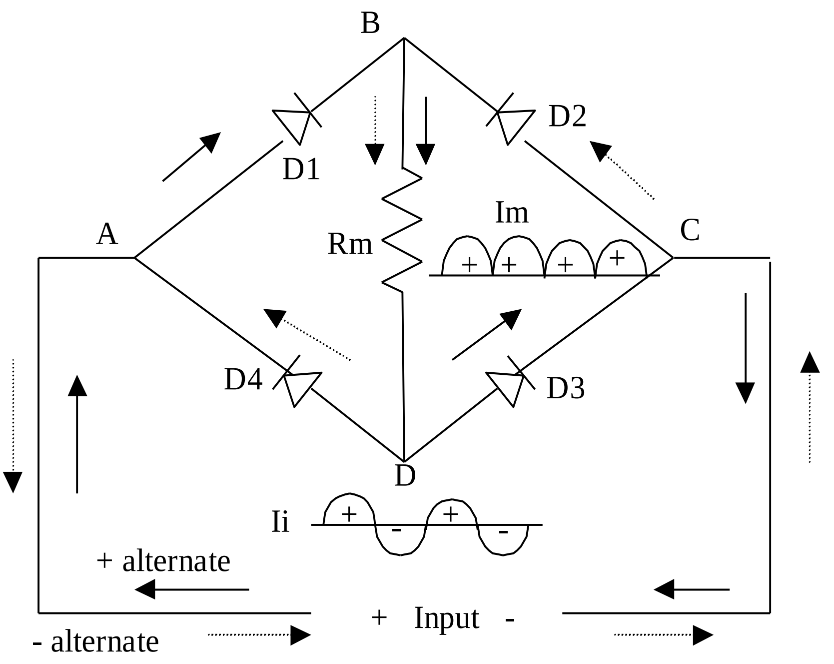


Figure 4. 23 Bridge type full-wave rectifier

$$V_{AV} = \frac{2}{T}\int_0^{T/2} V_m \sin\omega t dt = \frac{2V_m}{\pi} = 0.636V_m$$

V_{AV} is the DC component of the voltage and it is the value read by the moving coil instruments. Hence, **the inherently measured value (IM) is the average value, while the true value is the RMS value**. The voltage reading will contain reading error (unless it is corrected) as

$$\%error = (\frac{V_{indicated} - V_{true}}{V_{true}}) \times 100\% = (\frac{V_{Average} - V_{RMS}}{V_{RMS}}) \times 100\% = -10\%$$

and the indicated voltage will be 10% less than the true value.

Form Factor and Waveform Errors

For Sinusoidal Waveforms

The ratio of the true value to the measured value is called the **form factor or safe factor** *(SF)*. For sinusoidal signals the form factor is $SF = (V_{RMS}/V_{AV})$.

In AC voltmeters, the reading is corrected by a scale factor = safe factor *(SF)* = 1.11. This can be done either at the calculation of the series resistance or setting the divisions of the scale. Eventually, the error is eliminated as:

$$\%error = (\frac{V_{indicated} - V_{true}}{V_{true}}) \times 100\% = (\frac{1.11 \times V_{Average} - V_{RMS}}{V_{RMS}}) \times 100\% = 0\%$$

The value of the correction factor applied is only correct if the input signal is sinusoidal and the above formula is of course true for sinusoidal signals only. The true RMS value is actually proportional to the square-root of the average of the square of the curve, and not to the average of the absolute value of the curve. For any given waveform, the ratio of these two averages will be constant and, as most measurements are carried out on what are (nominally) sine waves, the correction factor assumes this waveform; but any distortion or offsets will lead to errors. Hence, for other (nonsinusoidal) waveforms, the error may be nonzero indicating erroneous readings.

For Triangular Waveform

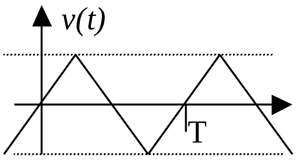


Figure 4. 24 A triangular waveform

A triangular voltage waveform *v(t)* with amplitude V_m and period *T* is shown in Figure 4.24. The negative portion is converted to positive after the full-wave rectification. Due to the symmetry of the signal, interval from 0 to *T/4* can be used for integration in finding the average (DC) and RMS values. In this interval, the signal can be expressed as $v(t) = 4V_m/T$.

Thus,

$$V_{AV} = \frac{4}{T}\int_0^{T/4} \frac{4V_m}{T}\,dt = \frac{V_m}{2} = 0.5V_m$$

This is the inherently measured (IM) value. A meter corrected for sinusoidal waveforms will indicate

V_{ind} = 1.11x0.5V_m= 0.555 V_m

The RMS value can be computed as: $$V_{RMS} = \sqrt{\frac{4}{T}\int_0^{T/4} \frac{16V_m^{\,2}}{T^2}\,dt} = \frac{V_m}{\sqrt{3}} = 0.577V_m$$

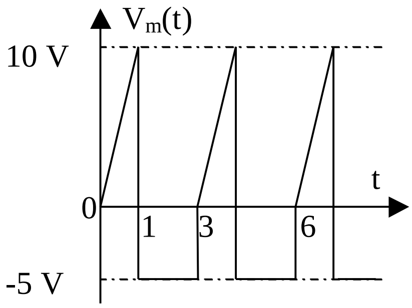
Hence, the form factor for the triangular waveform is 1.155 and 1.11$V_{average} \neq V_{RMS}$ .The percentile measurement error:

$$\%error = (\frac{V_{indicated} - V_{true}}{V_{true}}) \times 100\% = (\frac{1.11 \times V_{Average} - V_{RMS}}{V_{RMS}}) \times 100\% = \frac{0.555 - 0.577}{0.577} \times 100\% = -3.81\%$$

The Correction Factor

A correction factor (CF) is used to multiply the reading indicated by the meter to correct the measured value. The correction factor must be determined for every specific waveform individually as:

$$CF = \frac{(SF)_{waveform}}{(SF)_{\sin usoidal}} = \frac{(V_{RMS}/V_{IM})_{waveform}}{(V_{RMS}/V_{IM})_{\sin usoidal}}$$



Waveform for example 4.15

The voltage indicated for the triangular waveform using a meter adjusted for a sinusoidal waveform can be written as:

$$V_{ind} = SFx(V_{IM})_{waveform} = (\frac{V_{RMS}}{V_{AV}})_{sinusoidal} x(V_{AV})_{waveform}$$

Eventually,

$$(V_{ind})(CF) = (SF)_{wave} \bullet (V_{IM})_{wave} = (V_{RMS})_{wave} = V_{true}$$

The error without the correction: $\%error = \frac{1-CF}{CF} \times 100\%$

For the triangular wave shown in the above example $CF = \frac{0.577/0.5}{0707/0.636} = \frac{1.154}{1.11} = 1.0396$ yielding the percentile error of –3.81%, same as the one found before.

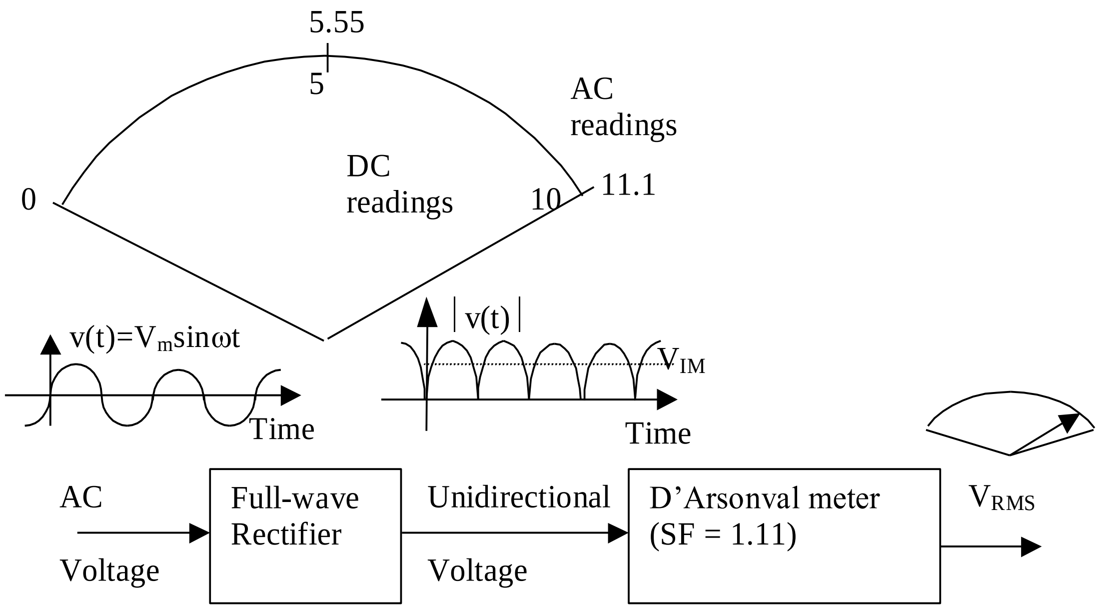


Figure 4. 25 Illustration of an AC voltmeter corrected for sinusoidal signals

Figure 4.25 shows a pictorial presentation of the scale calibrated for sinusoidal voltage waveforms, model of the AC voltmeter based on the basic D'Arsonval meter with samples of input and output waveforms.

Example 4.15

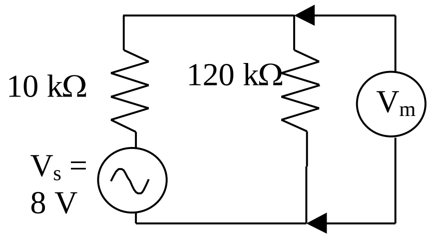
A D'Arsonval (moving coil) movement based AC voltmeter is calibrated to read correctly the RMS value of applied sinusoidal voltages. The meter resistance is 10 kΩ/V and it is used in 0 – 10 V range.

Find V_m measured by the meter and the percentile loading error.

True value of the voltage V_{true}= 8x120/130 = 7.38 V; R_m= 100 kΩ leading to R_L'= 100x120/220 = 54.5 kΩ. Therefore V_m = 8x54.5/64.5 = 6.76 V. Percentile loading error = -8.4%.

A different periodic waveform is applied and the waveform $V_m(t)$ shown appears across the meter.

Calculate V_{RMS} for this waveform; $V_{RMS}^2 = \frac{1}{3}[\int_0^1 100t^2 dt + \int_1^3 25dt = \frac{250}{9}$; V_{RMS} = 5.27V,

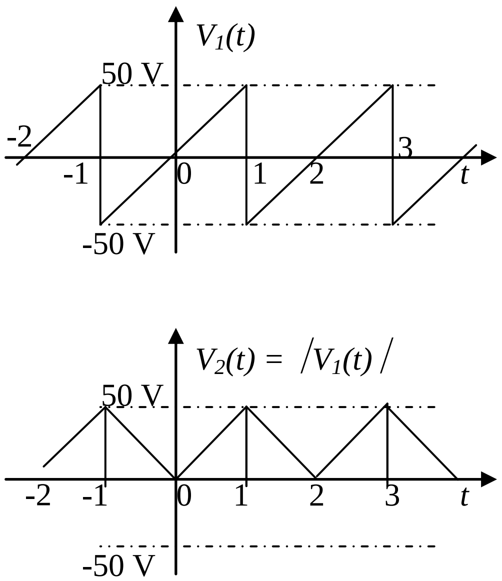


Circuit for example 4.15

How much is the voltage indicated by the meter ($V_{indicated}$)?

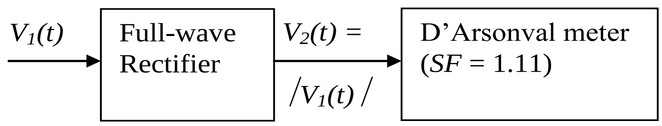
$|V|_{(AV)} = \frac{1}{3}[\int_0^1 10tdt + \int_1^3 5dt = 5V$ Therefore, V_{ind} = 1.11x5 = 5.55 V

Find the waveform error in this measurement. % waveform error = 100x(5.55 – 5.27)/5.27 = 5.3%.



Waveforms for example 4.16

Example 4.16



Model for example 4.16

The voltage waveform shown has a magnitude ± 50 V and it is applied to an AC voltmeter composed of a full-wave rectifier and a moving coil (D'Arsonval) meter. It is calibrated to measure voltages with sinusoidal waveforms correctly.

Find the average and RMS values of $V_1(t)$

$$V_{1(AV)} = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} V_1(t)dt = \frac{1}{2}\int_{-1}^{1} 50tdt = \frac{25}{2}[1-1] = 0$$

$$V_{1(RMS)} = \sqrt{\frac{1}{2}\int_{-1}^{1} 2500t^2 dt} = \sqrt{\frac{2500}{6}[1+1]} = \frac{50}{\sqrt{3}} = 28.87$$

- Sketch the waveform for $V_2(t)$
- Find the average and RMS values of $V_2(t)$.

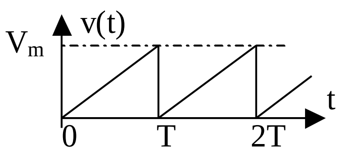
Ans. The RMS value of $V_2(t)$ is the same as that of $V_1(t)$ which is 28.87 volts. The average value can be calculated from the area of the triangle easily as 50/2 = 25 volts.

- Find the voltage indicated by the meter. Ans. 25x1.11= 27.75 volts
- Calculate the error due to the waveform and find the correction factor.

The % waveform error = 100x[27.75 – 28.87]/28.87 = -3.88%

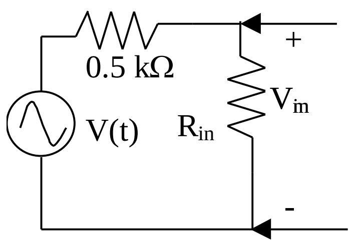
Correction factor (CF) = $(SF)_{wave}/(SF)_{sine}$= (28.87/25)/1.11 = 1.04

Example 4.17



Signal for example 4.17

A generator with 500 Ω internal resistance has a saw tooth output voltage as shown. The RMS value of this output is to be measured by a moving coil instrument whose internal resistance is 10 kΩ. The instrument has a full wave rectifier and is calibrated for sinusoidal waveforms. Calculate the error due to the waveform and also the loading error.



Circuit for example 4.17

The schematic diagram illustrates the measurement problem. For an ideal voltmeter, the meter resistance R_{in} must be very large ($R_{in} \to \infty$). Therefore, the true value of the output voltage $v_{true}(t) = v(t)$. The internal resistance is given as R_{in} = 10 kΩ yielding $v_{in}(t) = (10/10.5)v(t)$. Hence,

$$\%\ (loading)\ error = \frac{v_{in} - v_{true}}{v_{true}} x100 = \frac{\frac{10}{10.5} - 1}{1} x100 = -4.8\%$$

The voltage measured using this meter is the average of $v_{in}(t)$ which is:

$V_{AV} = \frac{10}{10.5} x \frac{1}{T}\int_0^T \frac{V_m}{T} t dt = \frac{10}{10.5} x \frac{V_m}{2}$. The reading indicated by the meter is compensated for the sinusoidal waveform and it becomes: $V_{ind} = 1.11x\frac{5V_m}{10.5} = 0.529V_m$

The true value that must be measured by the meter is the RMS value which is:

$$V_{RMS} = \frac{10}{10.5}\sqrt{\frac{1}{T}\int_0^T \frac{V_m^{\ 2}}{T^2} t^2 dt} = \frac{10}{10.5} x \frac{V_m}{\sqrt{3}} = 0.55V_m$$

Hence, the waveform error is 100x(0.529 – 0.55)/0.55 = -3.82%

If the meter would be ideal ($R_{in} \rightarrow \infty$), then $V_{true} = V_{RMS} = \frac{V_m}{\sqrt{3}} = 0.577V_m$ Having $0.529V_m$ indicated by the meter, the total measurement error (loading + waveform) becomes 100x(0.529 – 0.577)/0.577 = -8.32%

Clamp-On Meters

Clamp-on meters are used for measuring AC circuit currents in a non-invasive manner which avoids having to break the circuit being measured. The meter clamps on to a current-carrying conductor and the output reading is obtained by transformer action. Figure 4.26 illustrates the principle of operation, where the clamp-on jaws of the instrument act as a transformer core and the current-carrying conductor acts as a primary winding. Current induced in the secondary winding is rectified and applied to a moving-coil meter. Although it is a very convenient instrument to use, the clamp-on meter has low, sensitivity and the minimum current measurable is usually about 1 amp.

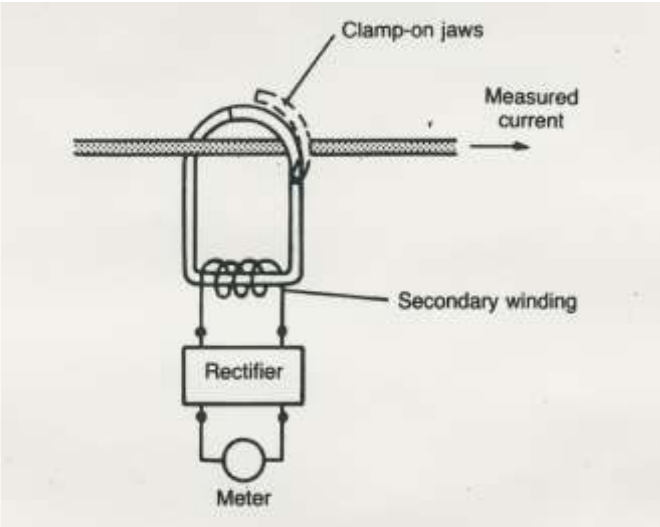


Figure 4. 26 A clamp-on meter in practice

True RMS Meters

The rectification, averaging and form factor correction approach produces adequate results in most cases. However, a correct conversion or the measurement of non sine wave values, requires a more complex and costly converter, known as a True RMS converter. The characteristics of these meters are defined in terms of the input range, bandwidth (frequency range in which the device operates successfully), accuracy and crest factor. The **crest factor** is a measurement of a waveform, calculated from the peak amplitude of the waveform divided by the RMS value of the waveform as illustrated in Figure 4.27. The power dissipated by a resistor R that is exposed to the signal is $P = I_{RMS}^2 * R = \frac{V_{RMS}^2}{R}$.

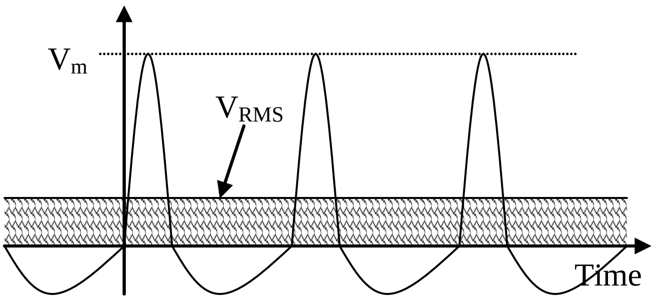


Figure 4. 27 A complex waveform with high crest factor

This principle was exploited in early thermal converters as illustrated in Figure 4.28. The AC signal would be applied to a small heating element which was twinned with a thermocouple which could be used in a DC measuring circuit. The technique is not particularly precise but it will measure any waveform at any frequency. Thermal converters have become quite rare, but as they are inherently simple and cheap they are still used by radio hams and hobbyists, who may remove the thermal element of an old unreliable instrument and incorporate it into a modern design of their own construction.

A second approach is to use analog electronic converters as illustrated in Figure 4.29. Analog electronic circuits may use:

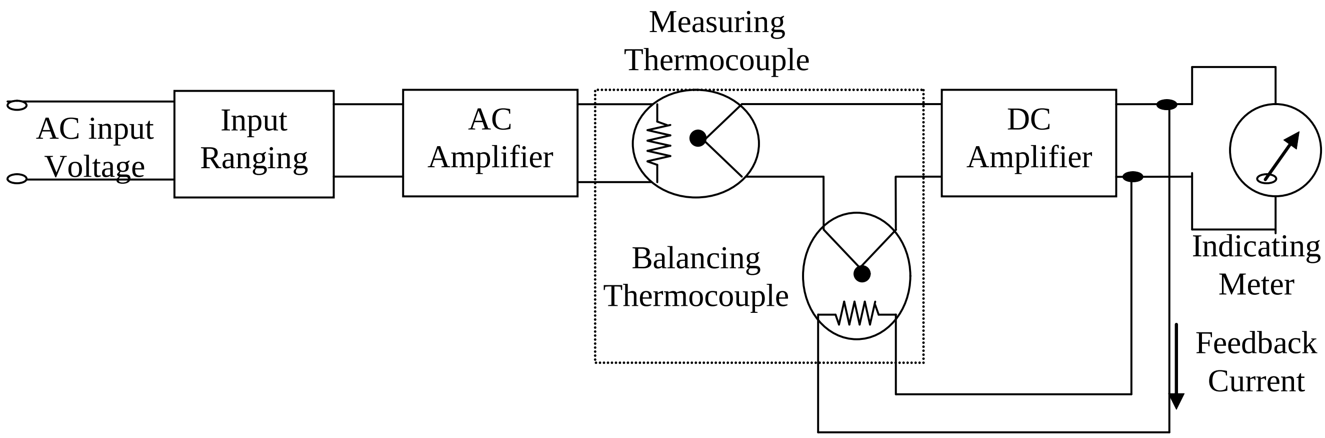


Figure 4. 28 A true RMS type AC voltmeter that uses the thermal converter principle.

- an analog multiplier in a specific configuration which multiplies the input signal by itself (squares it), averages the result with a capacitor, and then calculates the square root of the value (via a multiplier/squarer circuit in the feedback loop of an operational amplifier), or
- a full-wave precision rectifier circuit to create the absolute value of the input signal, which is fed into a operational amplifier arranged to give an exponential transfer function, then doubled in voltage and fed to a log amplifier as a means of deriving the square-law transfer function, before time-averaging and calculating the square root of the voltage, similar to above,
- or a field-effect transistor may be used to directly create the square-law transfer function, before time-averaging.

Unlike thermal converters they are subject to bandwidth limitations which makes them unsuitable for most RF work. The circuitry before time averaging is particularly crucial for high frequency performance. The slew rate limitation of the operational amplifier used to create the absolute value (especially at low input signal levels) tends to make the second method the poorest at high frequencies, while the FET method can work close to VHF. Specialist techniques are required to produce sufficiently accurate integrated circuits for complex analog calculations, and very often meters equipped with such circuits offer True RMS conversion as an optional extra with a significant price increase.

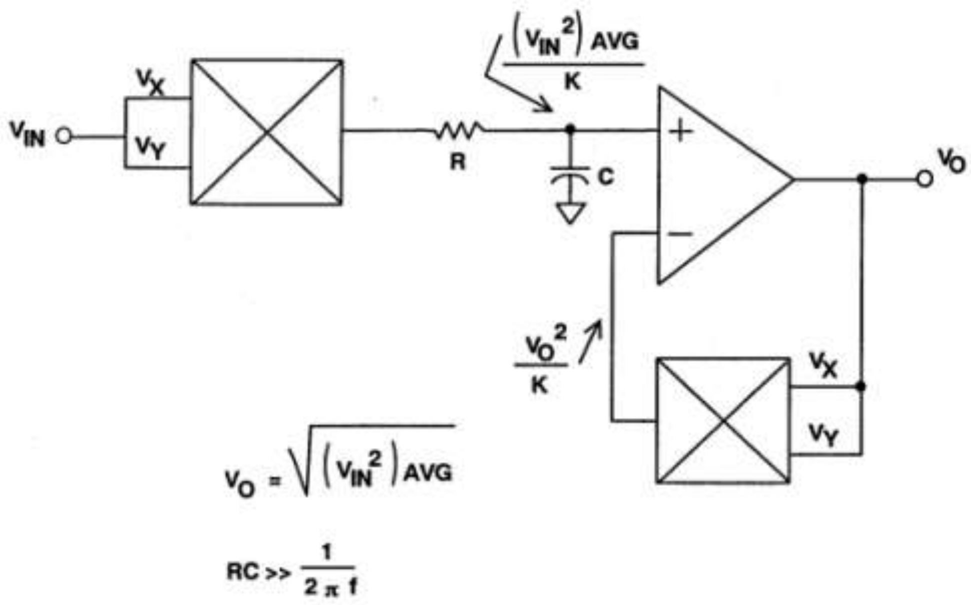


Figure 4. 29 Analog RMS to DC converter

The third approach is to use Digital RMS converters. Digital and PC-based oscilloscopes have the waveform being digitized so that the correct RMS value may be calculated directly. Obviously the precision and the bandwidth of the conversion is entirely dependent on the analog to digital conversion. In most cases, true RMS measurements are made on repetitive waveforms, and under such conditions digital oscilloscopes (and a few sophisticated sampling multimeters) are able to achieve very high bandwidths as they sample at a fraction of the signal frequency to obtain a stroboscopic effect (that will be explained later in section covering the digital storage oscilloscope).

http://www.analog.com/static/imported-files/tutorials/MT-081.pdf

ELECTRONIC COUNTERS

Oscilloscope Versus Electronic Counters and Digital Voltmeters

Commonalities Between Electronic Counters and Digital Voltmeters

Electronic counters are extensively used for measuring the frequency (number of occurrence of an event in a given time), time period of an event and time interval between two events. Most digital voltmeters generate a time-interval related to the level of the input voltage first. Then, they measure that interval and display it. They are easy to use and display the readings directly in numerical forms. Therefore, the electronic circuitries in both systems have many components in common and they will be discussed together in this chapter.

Limitations of the Oscilloscope as a Measuring Instrument

The oscilloscope is a versatile and useful device to observe waveforms. Yet, it has limitations as a measuring instrument as:

- The input impedance is 1 MΩ in all measurement ranges which may be small and cause instrument loading in some applications. Input impedances of electronic counters and digital voltmeters are much higher (in tens of MΩ) that eliminate the loading problem.
- The oscilloscope is more prone to human errors since results are obtained through calculations. In digital voltmeters and electronic counters the results are displayed directly.
- What is measured in the oscilloscope is the distance between two points on the screen. The results are limited to the reading accuracy of the observer from the screen at the first place.
- Estimates of the amplitude and time variations are made from the displacements drawn onto the screen with the help of sensitivity settings.
- The frequency can only be determined mathematically as the inverse of the period.
- The smallest possible reading error from the screen occurs when the interval to be measured covers the full 100 mm span and the starting point is aligned sharply against the first ruled vertical line. Then, the measurement error involves uncertainty only in reading the terminal point with $\pm$ 0.5 mm. Hence, the percentile error is $\pm$ 0.5% at best which can also be expressed as *one in two hundred*s. The simplest counter with a four-digit display will have an uncertainty of $\pm$ 1 digit in the last place (least significant digit) which means that the reading error can be as low as *one in ten thousand*s.

Time and Frequency Measurements

Operational Modes of Counters

Electronic counters are extensively used for measuring the frequency (number of occurrence of an event in a given time), time period of an event and time interval between two events. They display the results directly in digital forms that can be easily read by the user.

The counters work in three operational modes as:

- the frequency,
- time-period and
- time-interval.

The frequency is defined in two ways as illustrated in Figure 4.30:

- The number of occurrences of event over the time of observation (i.e. 6 events per second). All digital displays have an inherent uncertainty of ±1 digit in the last digit of the display. If the number displayed is small, this uncertainty causes large reading errors. Therefore, this mode is useful at high frequencies.

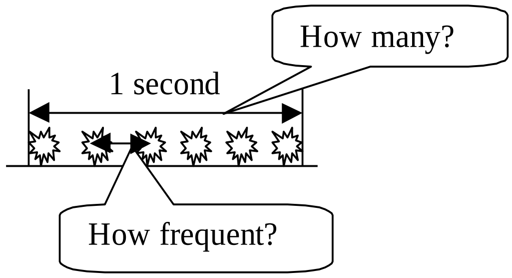


Figure 4. 30 Definitions of frequency

- The inverse of the time-period (i.e. one explosion every 100 millisecond). This is useful at low frequencies. Some counters automatically switch to this mode as the low frequency ranges are selected. The period is measured and inverted usually by digital techniques and the displayed result is the frequency. New counters contain microprocessors that perform this operation easily.

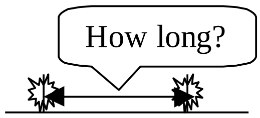


Figure 4. 31 Time-period

The time measurement is used for:

- Time-period; the time interval between two successive identical points for a periodic event as illustrated in Figure 4.31.
- The time-interval; the time interval between two events that run simultaneously as shown in Figure 4.32. This is very useful in determining the phase shift between two signals.

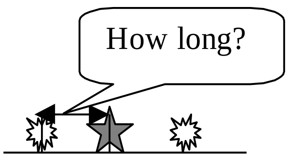


Figure 4. 32 Time-interval

Devices Commonly Used in Electronic Measuring Instruments

Amplifiers

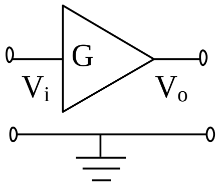


Figure 4. 33 Symbol of an amplifier

The amplifier is a device that increases the magnitude of the input voltage (voltage amplifier as in Figure 4.33), current (current amplifier) and power (power amplifier). The ratio of the output to the input (if of the same kind, i.e. both voltage) is called the **gain** if it is greater than 1 and denoted by G. For a voltage amplifier; $G=V_o/V_I$

where V_o is the output voltage and V_I is the input voltage. The gain is a unitless quantity.

Sometimes the gain is expressed in decibels (dB) as: $G_{dB} = 10log(P_o/P_I) = 20log(V_o/V_I)$

where P_o is the output power and P_I is the input power of the amplifier measured across the same resistor.

If the output is smaller than the input, this is called the **attenuation**. G_{dB} is positive for the gain and negative for the attenuation. For example, a gain of 60 dB indicates that the output is the input multiplied by 1000 while a gain of –20 dB shows that the input is reduced (attenuated) by 10 times by the system.

The Comparator

The comparator is a device that has two inputs and one output as shown in Figure 4.34. The output has two voltage levels as “high” and “low”. It detects the sign of the voltage difference and reflects it to the output level as indicated in the figure. One of the input is set to fixed voltage whose value can be set externally and it is called the “threshold”. The output shows the sign of ($V_1 – V_2$). Hence, it is in high state when the input voltage is higher than the threshold ($V_1 > V_2$) and goes to low state as the input becomes smaller than the threshold ($V_1 < V_2$).The shape of the output is reversed if the input and threshold connections are interchanged. Important parameters used in identifying a pulse are marked on Figure 4.34. In some comparators, the threshold is internally connected to the middle level (ground) and only one input is available externally. These devices are also called “zero-crossing detectors”.

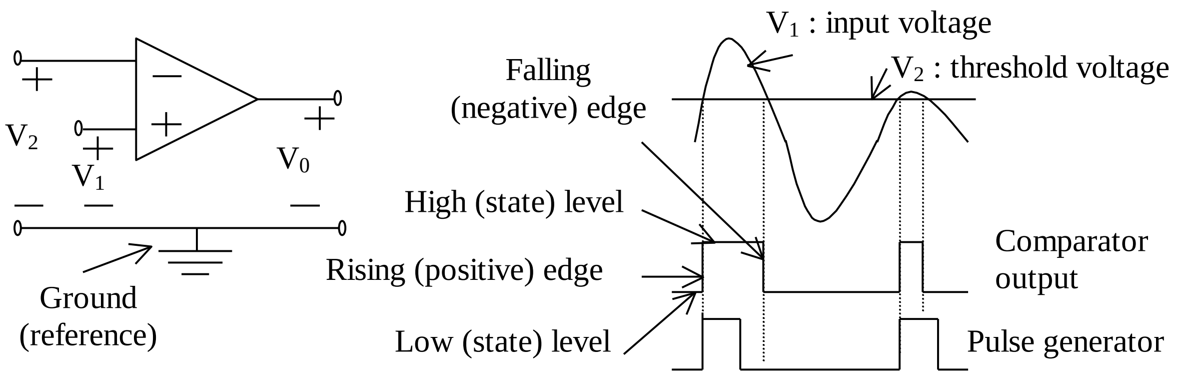


Figure 4. 34 Comparator with its timing diagram

In some comparators, the output changes its states slightly after the input walks over the point of coincidence with the threshold (in either direction) causing slight delay between the generation of the output pulse and the point of coincidence. This delay is called the “hysteresis” and it is used to avoid false detection in case of noisy input signals. Such comparators are commonly called the “Schmitt triggers”.

The Pulse Generator

The output of a comparator is a rectangular pulse-like signal whose high and low states depend upon the magnitude of the input signal as compared to a threshold voltage. The pulse generator (also known as monostable multivibrator) receives the output of the comparator or any pulse-like signal and produces a pulse with fixed duration immaterial of the duration of the input pulse. The pulse may be initiated either by the positive edge or the negative edge of the input pulse. It is set to positive-edge triggering in the example shown in Figure 4.34.

The Clock

It is a device that generates timing pulses with a very high accuracy and stability in the frequency as illustrated in Figure 4.35. Crystal-controlled oscillators are used mostly. The output is a square wave in general, but it will be represented by a sequence of short duration pulses in most applications.

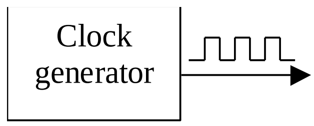


Figure 4. 35 The clock

Elements Common in All Modes of Operations of Counters

Following elements are common in all modes of counters:

- The magnitude of the input signal is not important. The periodic input signal is converted into a pulse sequence by the signal shaper, which is composed of a comparator and a pulse generator. Here, AC/DC coupling, trigger level and polarity settings are available as in the case of the oscilloscope. There is no amplitude range selection except a divide by ten (20 dB) attenuator to reduce the amplitude of the input signal to a safe level for high-amplitude inputs.

- All measurements are related to the timing information coming from an internal time-base. Therefore, a very stable time base is an essential element of the counter. Calibration of the time-base circuits may be achieved by using special frequency standards based on tuning forks, crystal oscillators or with NBS (National Broadcasting Society) standard broadcast frequencies.
- A control gate sets the duration of the counting and refresh rate (the frequency of repeating the measurement).
- They mostly use 7-segment light emitting diode (led) or liquid crystal (lcd) type displays. Depending upon the frequency range of operation, there may be six to eight digits displayed.
- Decimal counters are used to accumulate (count) incoming pulses from the pulse gate and

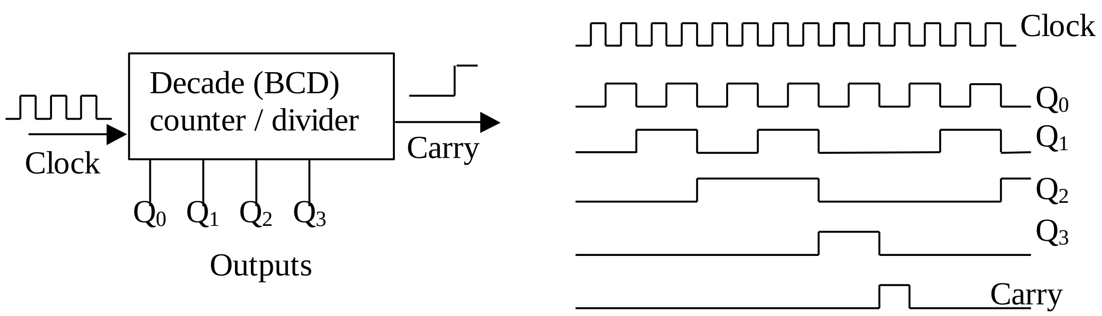


Figure 4. 36 The BCD (decade) counter and its timing diagram

generate a binary coded decimal (BCD) code at the output as illustrated in Figure 4.36. The code ranges from 0000 to 1001 corresponding to decimal "0" and "9" incrementing with every input pulse. With the 10th pulse, the code returns to 0000 and the counter provides a carry pulse to the next stage. At the end of the counting session, the code accumulated in the counters is transferred to a digital latch that holds it until the end of the next counting session. Counters are cleared automatically after the data is transferred to the latch. The user can also clear them during initialization. This code stored in the latch is applied to the display through BCD to 7-segment decoders and displayed as decimal numbers. The display also incorporates annotations for the time units (µs, ms, and s) and frequency units (Hz, kHz, and MHz). The time-base and/or gate control switches set the position of the decimal point. The unit of the measurement is highlighted.

The Counter in Frequency Mode

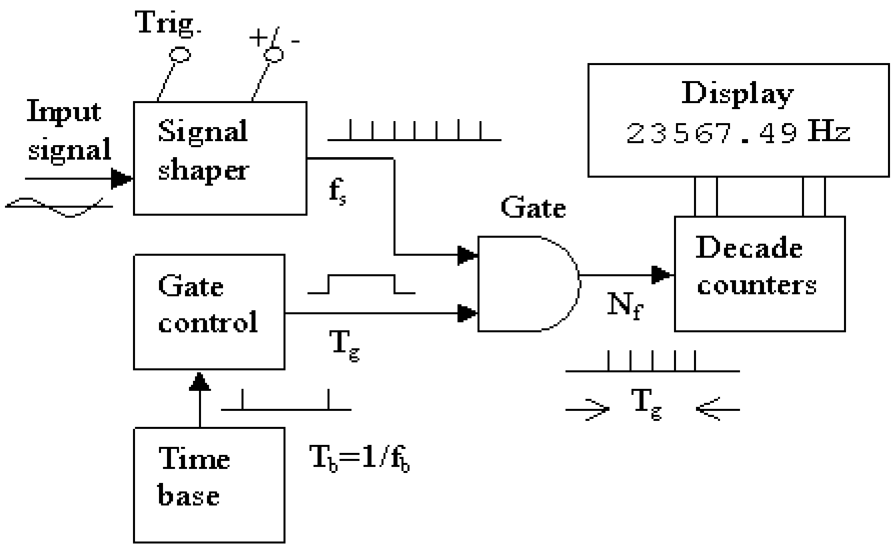


Figure 4. 37 Block diagram of the counter in frequency mode

Principle of Operation

Figure 4.37 shows the block diagram of a counter set to the frequency mode of operation. The time-base circuitry provides the start and stop pulses for the pulse gate. The pulses generated from the input signal via the signal shaper are counted. The duration of the gate signal (T_g) is equal to the period of the time base signal (T_b). Number of pulses counted

$$N_f = T_g \times f_s$$

f_s being frequency of the input signal. Commonly used values for T_b are 0.1 s, 1 s, and 10 s.

The Time Base

Accuracy of the measurement is directly affected by the uncertainty in gating. Hence, a time-base with high accuracy, precision and long-term stability is essential. This is managed via a high stability clock circuit that runs at frequency f_c shown in Figure 4.38. A series of decade counters are used (m of them in Figure 4.28) to obtain the time base signal yielding,

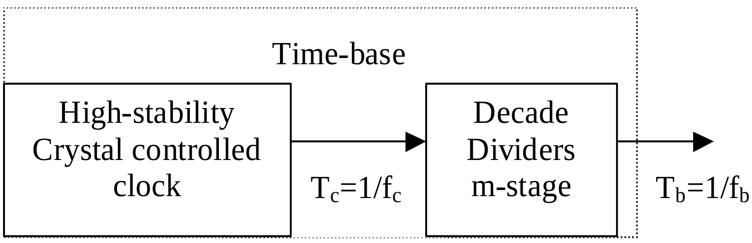


Figure 4. 38 Block diagram of the time-base

$$T_b = T_c \times 10^m$$

In some counters, the divider ratio is indicated at the time-base selector switch. Finally, the frequency of the input (f_s) is determined from the number displayed (N_f) and time-base setting (10^m) as:

$$f_s = \frac{N_f}{T_b} = \frac{N_f}{10^m} f_c$$

The decimal point automatically moves in between appropriate digits and respective frequency unit is also highlighted to ease the reading as mentioned above.

The Counter in Time-Period Mode

Principle of Operation

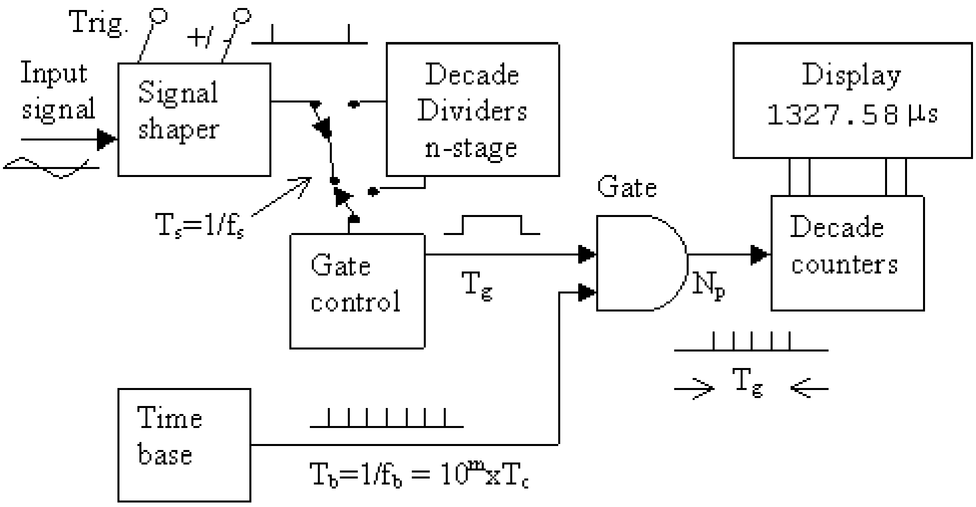


Figure 4. 39 Block diagram of the counter in time-period mode

In the period mode, the input signal provides the gating and the time-base supplies the pulses for counting as shown in Figure 4.39. The number of pulses counted:

$$N_p = f_b \times T_g$$

With T_g=T_s (the time-period of the input signal) and $f_b = {f_c}/{10^m}$, the period T_s can be expressed as:

$$T_s = \frac{N_p}{f_c} \times 10^m$$

Hence, 10^m becomes the multiplier in case of the period measurement. Period measurement is preferred to frequency measurement in determining lower frequencies. The read-out logic is designed to automatically positioned the decimal point and display the proper unit.

Averaging

The frequency measurement inherently involves accumulation of several pulses. Thus, small variations in the time-period of the input signal (jitter in the period) will cancel out each other and the resultant reading indicates the average value of the frequency of the signal rather than the instantaneous frequency. The period measurement however, uses a single period for the measurement. Furthermore, the display is normally refreshed at every three seconds or so. Fast refresh rates are not useful, since a human observer reads the display. Therefore, for a signal having time-period of fraction of a second, the system stays idle for a long time.

The accuracy of reading and reliability of the measurement may be increased by using the multiple-period average mode of operation. A series of decade dividers (n of them in the figure) are introduced between the signal shaper and gate control circuits. Hence, the measured period is averaged over 10^n cycles. The resultant equation for the period measurement becomes:

$$T_s = \frac{N_p}{f_c} \times \frac{10^m}{10^n}$$

The Counter in Time-Interval Mode

The phase-angle (shift) between two signals may be determined by measuring the time interval between similar points on the two waveforms. Figure 4.40 illustrates the principle diagram of the measuring set-up. Both inputs contain signal shapers that generate pulses corresponding to the trigger pick-off. One of the pulse controls the starting of the counting while the other one stops the counting. Trigger levels and slopes may be different for both channels. A common-separate switch (Cm / Sep) allows utilization of the same signal for both channels and with different trigger settings; the time between sections of the same waveform can be measured. This is especially important in determining the pulse duration and rise-time of the signal.

Errors in Measurements Using Counters

There are three reasons for errors in measuring frequency and time using counters as the time-base errors, trigger level error and gating errors.

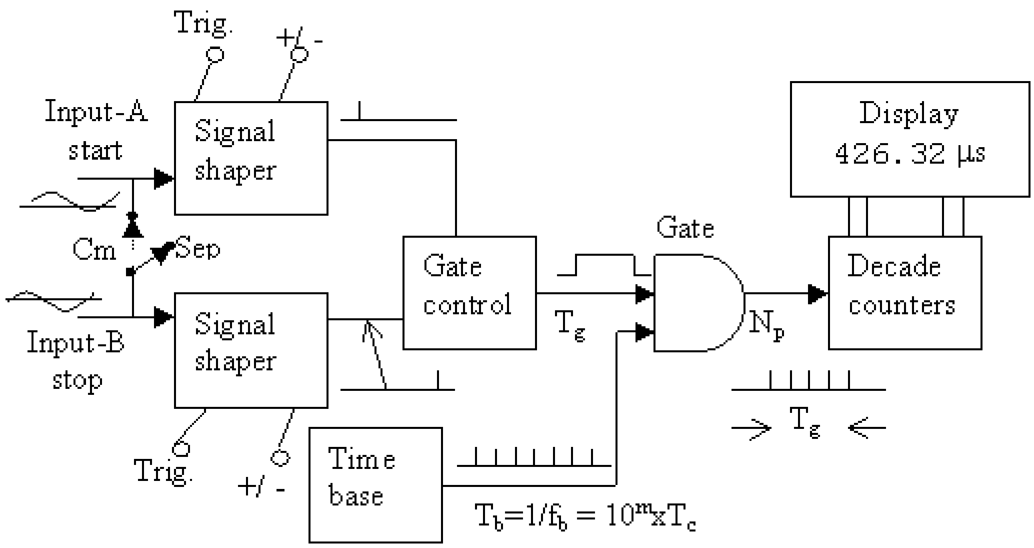


Figure 4. 40 Block diagram of the counter in time-interval mode

- **Time-base errors**: oscillator calibration errors resulting from;
- **Short-term crystal stability errors**: due to voltage transients, shock and vibration,
- **Long-term crystal stability errors**. Aging rate of the 10 MHz crystal standard is less than 3 parts in 10^7 per month for the HP 5326B counter as specified by the manufacturer.
- **Trigger-level errors** (only in time-interval and period modes). Using large signal amplitude and fast rise-time can minimize them.
- **Gating errors**: ± 1 counts of the display's last significant digit. This error is inherent to all electronic counters and is due to the lack of synchronization between the gating and the clock (counted signals).

Example 4.18: Reading Error in Frequency and Period Measurements

Two sine waves at 9.5 Hz and 200 kHz are applied to an electronic counter. Both the frequency mode and the period mode of operations are used. Time-base settings in the frequency mode and multiplier settings in the period mode are tabulated in Table 4.1 with the display readings for both signals. Signal averaging in the period mode is not used.

The clock frequency in the time-base is 10 MHz (period 0.1μs) and it is divided by an 8-stage decade counters/dividers to obtain a 10-second time-base. Interim stages are also available to obtain various time-base (T.B.) and multiplier settings. In the multiplier settings, the number of counters involved is also 8. Hence, in multiplier selection 10^8 the clock output is taken from the output of the 8^{th} stage leading to a clock frequency $10^7/10^8$ = 0.1 Hz. In 10^0=1 however, the clock output is taken directly from the clock generator. Assuming the reading error as ± 1 digit in the last place, period reading for low frequency and frequency reading for high frequency signals lead to smaller error. Relative error $\Delta T/T = \Delta f/f$ as T or f is computed from each other through the inverse relationship.

Table 4.1. Measurements of period and frequency for signals at 9.5 Hz and 200 kHz.

9.5 Hz				200 kHz			
Frequency mode		Period mode		Frequency mode		Period mode	
T. B.	Reading	Mult	Reading	T. B	Reading	Mult	Reading
10 s	0.0094 kHz	10^8	1*	10 s	191.2719 kHz	10^8	1*
1 s	0.009 kHz	10^7	1 s	1 s	191.280 kHz	10^7	1 s
0.1 s	0.00 kHz	10^6	0.2 s	0.1 s	191.27 kHz	10^6	0.1 s
10 ms	0.0 kHz	10^5	0.11 s	10 ms	191.2 kHz	10^5	0.01 s
1 ms	0.000 MHz	10^4	106 ms	1 ms	0.191 MHz	10^4	1 ms
0.1ms	0.00 MHz	10^3	106.0 ms	0.1 ms	0.19 MHz	10^3	0.1 ms
10 μs	0.0 MHz	10^2	105.97 ms	10 μs	0.1 MHz	10^2	0.01 ms
1 μs	0 MHz	10^1	105944 μs	1 μs	0 MHz	10^1	6 μs
0.1 μs	0.00 GHz	10^0=1	105951.0 μs	0.1 μs	0.00 GHz	1	5.2 μs

Example 4.19

Draw the functional block diagram of an electronic counter in frequency mode and explain the function of each block briefly. What will be the number displayed if the time-base is set to 1 msec and the frequency of the input signal is 568,321 Hz? How much is the uncertainty in the frequency reading? What would be the reading and uncertainty in reading if time-base was set to 1 sec?

Please refer to the text for the block diagram. The signal shaper converts the periodic input signal into a pulse sequence. A very stable time base provides the timing (start and stop) pulses. A control gate sets the duration of the counting and refresh rate (the frequency of repeating the measurement). AND gate allows the number of input pulses during the gate to be selected.

A display mostly using 7-segment light emitting diode (led) or liquid crystal (lcd) type displays to indicate the decimal digits. Decimal counters are used to accumulate (count) the incoming pulses from the pulse gate and generate a binary coded decimal (BCD) code at the output. The display also incorporates annotations for the time units (μs, ms, and s) and frequency units (Hz, kHz, and MHz) displayed. The decimal point automatically moves to proper place and the unit of the measurement is highlighted.

The frequency of the input signal indicates that there will be 568321 pulses in one second. With time-base set to 1 msec, only 568 pulses accumulate in the counter. Hence, the reading would be 568 kHz with uncertainty ± 1 kHz. If time-base is set to 1 sec, then reading would be 568,321 Hz with uncertainty ± 1 Hz.

Measurement of Rotative Speed

Speed of rotation of electrical motors and other rotating objects can be measured by using a shaft encoder or stroboscopic method.

The Shaft Encoder Method

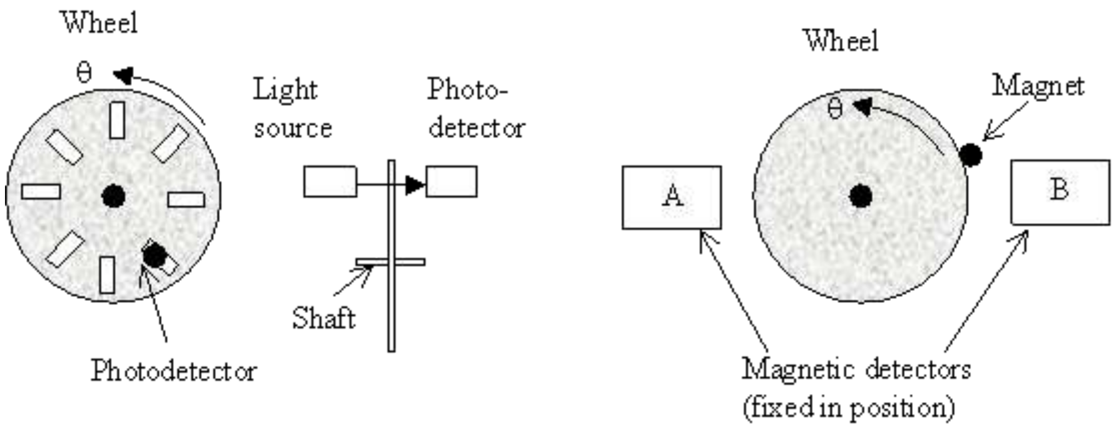


Figure 4. 41 Optical and magnetic shaft encoders to measure the rotational speed

There are two methods that are commonly used for measuring the angle of rotation and the rotational speed as illustrated in Figure 4.41. A disk is fixed on the shaft and allowed to rotate freely with it. In the optical shaft encoding, the disk is either slotted or painted with to have opaque and transparent regions. A light source illuminates one side of the disk by a thin beam of light. A light detector is facing at the opposite surface. The detector receives the beam of light only as the transparent or slotted regions fall in between the source and the detector. Then, the detector produces a pulse every time such a slot appears in front of it. Counters are used to measure the pulses and determine the speed of rotation.

The second method is the magnetic shaft encoder that has a magnet fixed on the disk and detectors are placed into fixed positions outside. The detector is made up of a simple coil that generates an electrical current pulse every time the magnet pass in front of it. These pulses are amplified and applied to a counter as in the previous case. Hence, the frequency of pulses indicates the rotational speed.

The Stroboscopic Method

Shaft encoders are very useful, but they require a disk fixed on the rotating shaft. The stroboscopic method allows computation of the rotational speed without interfering with the rotation and without necessitating fixing anything on the shaft.

A flashing light illuminates the rotating shaft. The shaft appears stationary as the ratio of the flash frequency (cycles per second) to the rotational speed (revolution per second) is expressed as the ratio of two integers (*f/v = m/n*). The rotational speed can be calculated by determining two successive flash frequencies that produce a single image as

$$v = \frac{f_1 f_2}{f_1 - f_2} \quad \text{(for } \frac{v}{f_1} = \frac{m}{n}\text{ ; } \frac{v}{f_2} = \frac{m}{n} + 1 \text{)}$$

THE DIGITAL VOLTMETER (DVM)

Use, Advantages and Operation

It is a device used for measuring the magnitude of DC voltages. AC voltages can be measured after rectification and conversion to DC forms. DC/AC currents can be measured by passing them through a known resistance (internally or externally connected) and determining the voltage developed across the resistance (*V=IxR*).

The result of the measurement is displayed on a digital readout in numeric form as in the case of the counters. Most DVMs use the principle of time period measurement. Hence, the voltage is converted into a time interval "*t*" first. No frequency division is involved. Input range selection automatically changes the position of the decimal point on the display. The unit of measure is also highlighted in most devices to simplify the reading and annotation.

The DVM has several advantages over the analog type voltmeters as:

- Input range: from $\pm$ 1.000 000 V to $\pm$ 1,000.000 V with automatic range selection.
- Absolute accuracy: as high as $\pm$ 0.005% of the reading.
- Stability
- Resolution: 1 part in 10^6 (1 μV can be read in 1 V range).
- Input impedance: $R_i \approx 10\ M\Omega$; $C_i \approx 40$ pF
- Calibration: internal standard derived from a stabilized reference voltage source.
- Output signals: measured voltage is available as a BCD (binary coded decimal) code and can be send to computers or printers.

The block diagram in Figure 4.42 illustrates the principle of operation of a digital voltmeter. It is composed of an amplifier/attenuator, an analog to digital converter, storage, display and timing circuits. There is also a power supply to provide the electrical power to run electronic components. The circuit components except the analog to digital converter circuits are similar to the ones used in electronic counters. The input range selection can be manually switched between ranges to get most accurate reading or it can be auto ranging that switches between ranges automatically for best reading.

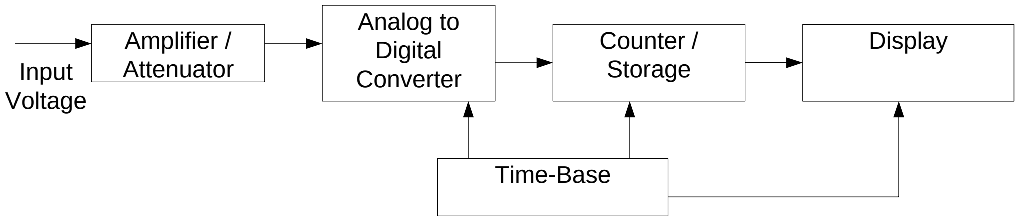


Figure 4. 42 A simplified diagram for a digital voltmeter

The Analog to Digital Converter (ADC) – Sample and Hold

The analog to digital converter contains a sample and hold circuit, and conversion circuits. The sample and hold is composed of an electronic switch and a capacitor. The switch turns on and off at regular intervals. The capacitor charges and assumes the level of the input voltage as the switch is on. It holds the charge (hence the level of the input voltage) as the switch is off. The unity-gain buffer eliminates the loading of the capacitor by proceeding analog to digital converter circuitry. Figure 4.43 shows a simplified diagram with the input and output waveforms of the circuit.

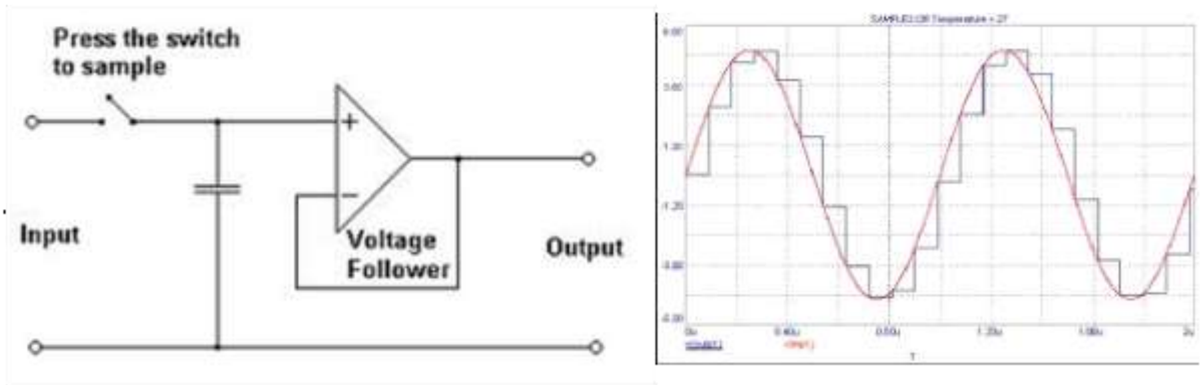


Figure 4. 43 Simplified circuit diagram with input and output waveforms of the sample and hold circuit

Digitization of Analog Signals

The input of the sample and hold circuit is a continuous time analog signal that can take any value any time. The output is a discrete time signal that can take any value but only at certain times. This signal can't be processed by a digital circuit unless it is converted into a digital code. Figure 4.44 illustrates the digitization of analog signals. The analog input signal is continuous in time and it can

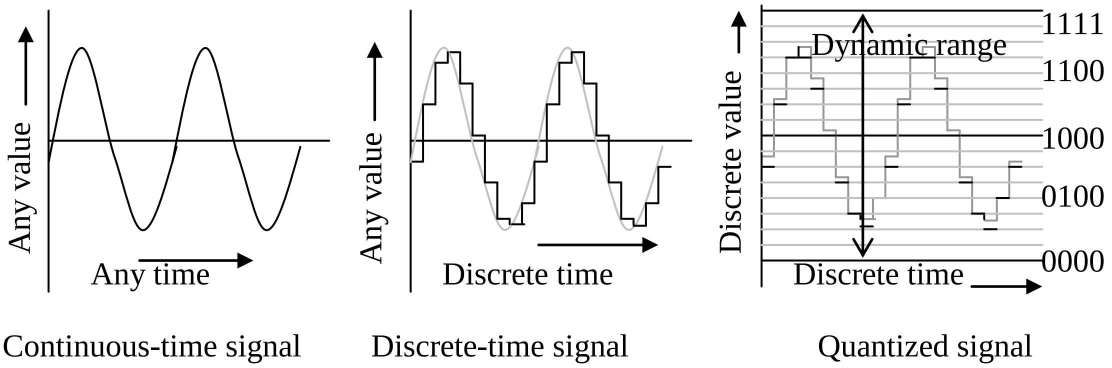


Figure 4. 44 Conversion of an analog signal into a digital signal

take any value at any time. This is converted to a discrete-time signal that can accept any value but at certain times. The next stage is to divide the amplitude range into discrete steps as well by a process called the quantization. The figure exemplifies the principles for a 4-bit converter in which the dynamic range (the maximum peak to peak amplitude that the input signal can attain) is divided into $2^4 - 1 = 15$ steps. A binary code (or binary coded decimal – BCD) is assigned for each level from 0000 to 1111 (1001 for BCD). Then,

$$V_i \approx k * Digital\ output$$

where *k* is the step size or resolution. Most digital storage oscilloscopes however, use 8-bit or 9-bit converters that divide the dynamic range into $2^8 - 1 = 255$ or $2^9 - 1 = 511$ steps.

Example 4.20

Signal from 800 – 1500 mV may be converted to 8-bit binary codes starting from 01010000_2 (80_{10}) to 10010110_2 (150_{10}). In this case, the step size *k* is equal to 10 mV. Quantization or conversion error of the ADC; $Quantization\ error = \frac{step\ size}{full\ scale} * 100 = \frac{1}{2^N - 1} * 100\%$, where *N* is the number of bit.

Several techniques are used to convert the DC analog voltage into a digital code that will be displayed. The mostly used ones are the integrating and successive approximation types. The integrating type has single-ramp, dual-ramp and digital ramp versions. The ramp type is the simplest one and it will be discussed firstly below. The single ramp type is very simple yet it has several limitations most of which are eliminated in the dual-integration type. The successive approximation type is also discussed briefly.

Integrating Type Analog to Digital Converters

The Basic Integrator

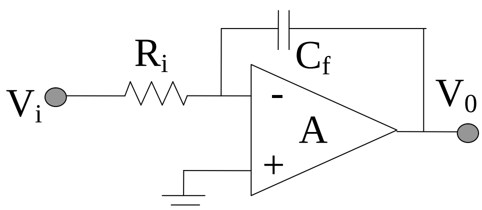


Figure 4. 45 The basic integrator circuit

This type of converters generates a time interval proportional to the input voltage. Then, this interval is measured and displayed using methods that were discussed in the counters section previously. The key circuit element is the integrator that generates an output that is related to the integral of the input. The basic integrator circuit is shown in Figure 4.45. It is similar to the inverting amplifier with the feedback resistor replaced by a capacitor. The input voltage V_i causes a current $I_i = {V_i}/{R_i}$ to flow through the capacitor C_f that generates an output voltage $V_0 = -\frac{1}{C_f}\int_0^t I_i dt - V_{co}$ since the inverting terminal of the op-amp is at virtual ground provided that the op-amp is not saturated. Hence, the output can be expressed as $V_0 = -\frac{1}{C_f R_i}\int_0^t V_i dt - V_{co}$. V_0 will decrease (or increase if V_i is negative) at a rate of $\frac{V_i}{R_i C_f}$

Functional Block Diagram of Ramp Type (Single Slope) DVM

Functional block diagram of a positive ramp type DVM is shown in Figure 4.46 The timing diagram is given in Figure 4.47. It has two major sections as the voltage to time conversion unit and time measurement unit. The conversion unit has a ramp generator that operates under the control of the sample rate oscillator, two comparators and a gate control circuitry.

The internally generated ramp voltage is applied to two comparators. The first comparator compares the ramp voltage into the input signal and produces a pulse output as the coincidence is achieved (as the ramp voltage becomes larger than the input voltage). The second comparator compares the ramp to the ground voltage (0 volt) and produces an output pulse at the coincidence. The input voltage to the first comparator must be between $\pm V_m$. The ranging and attenuation section scales the DC input voltage so that it will be within the dynamic range. The decimal point in the output display automatically positioned by the ranging circuits.

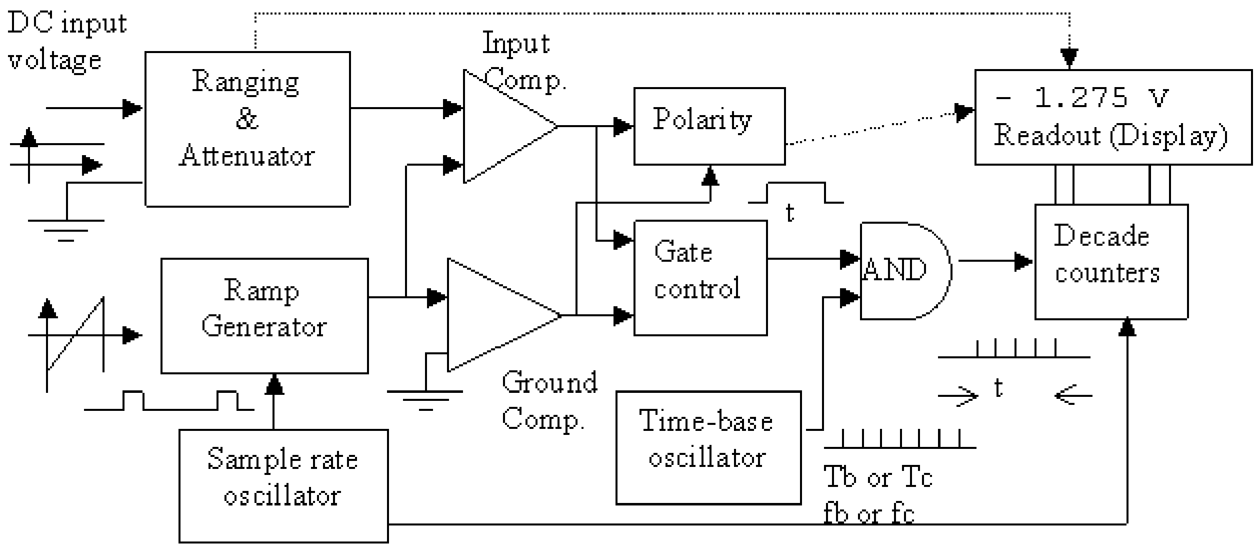


Figure 4. 46 Simplified block diagram of a single-ramp type digital voltmeter

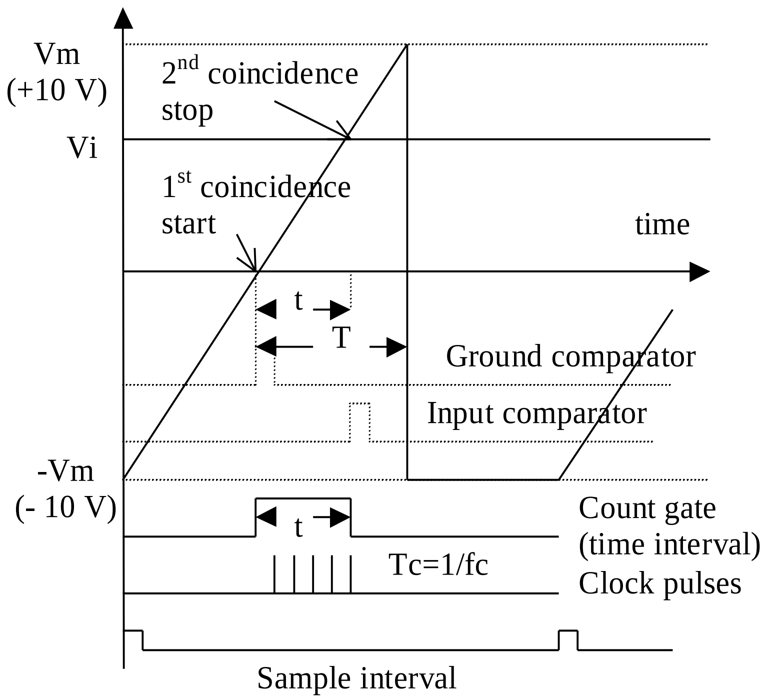


Figure 4. 47 Timing diagram for a single-ramp digital voltmeter

The outputs of the two comparators derive the gate control circuit that generates and output pulse that starts with the first coincidence pulse and ends with the second. Thus, the duration of the pulse "*t*" can be computed from the triangles as

$$\frac{V_i}{V_m} = \frac{t}{T} \Rightarrow t = \frac{T}{V_m} V_i$$

Hence, the voltage to time conversion is done yielding "*t*" to V_i with T and V_m constant. Number of time intervals (clock pulses) counted during this interval become:

$$N = t * f_c = V_i \frac{T * f_c}{V_m}$$

For the ramp voltage with fixed slope and time base that runs at fixed rate (f_c) N is directly proportional to V_i. The multiplier $T.f_c/V_m$ is set to a constant factor of 10.

The polarity of the voltage is indicated if it is "-". With no indication, it is understood that the polarity is "+". The polarity is detected by the polarity circuit with the help of comparator pulses. For positive slope ramp type voltmeter, the first coincidence of the ramp is with the ground voltage if the input is positive. With a negative input voltage however, the first coincidence will be with the input voltage.

The display stays for sometimes (around three seconds) and then it is refreshed by the sample rate oscillator. A trigger pulse is applied to the ramp generator to initiate a new ramp. Meanwhile a reset (initialize) pulse is applied to the decade counters to clear the previously stored code.

The display indicates the polarity as well as the numbers in decimal and a decimal point. The first digit contains the polarity sign and the number displayed can be only “1” or “0” for most voltmeters. Therefore, this is called “half” digit. Hence, a three and a half digit display can have up to 1999 and a four and a half digit one can go up to 19999.

Staircase (Digital) Ramp Type DVM

The ramp in the previous case has been generated by an analog integrator. It has been replaced a digitally generated one that looks like a staircase. The block diagram of the ramp generator and its output are shown in Figure 4.48. A binary counter continuously counts from a clock and its output is decoded into an analog voltage by a digital to analog converter. The input voltage is compared to the internally generated staircase ramp. It is the simplest A/D converter. The conversion is slow and the conversion time depends on the magnitude of the input signal.

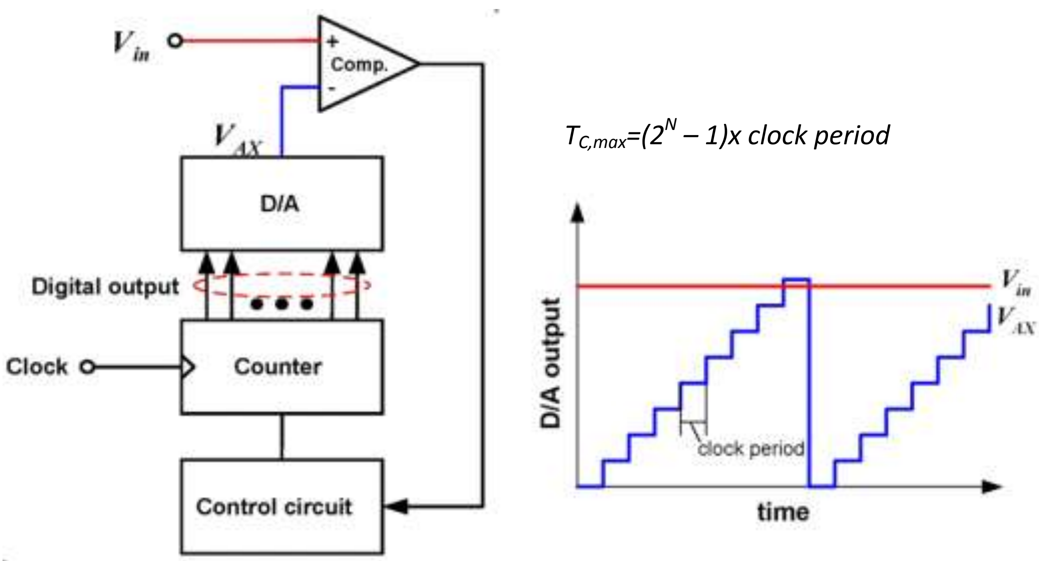


Figure 4. 48 The block diagram and output waveform of a staircase ramp generator.

Dual-Slope Integration Type DVM

The ramp type DVM (single slope) is very simple yet has several drawbacks. The major limitation is the sensitivity of the output to system components and clock. The dual slope techniques eliminate the sensitivities and hence the mostly implemented approach in DVMs. The operation of the integrator and its output waveform are shown in Figure 4.49.

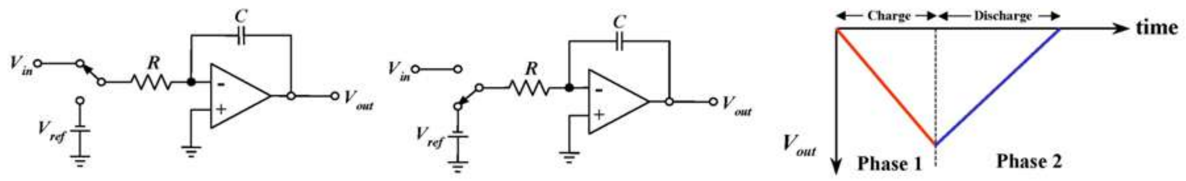


Figure 4. 49 The integrator in dual-slope type DVM and its output

The integrator works in two phases as charging and discharging. In phase-1, the switch connects the input of the integrator to the unknown input voltage (V_{in}) for a predetermined time *T* and the integrator capacitor *C* charges through the input resistor *R*. The output at the end of the charging time *T* is (assuming that $V_C(0) = 0$); $V_{out1} = \frac{V_{in}T}{RC}$. In phase-2, the switch toggles to the second position that connects the input to the reference voltage V_{ref} and the capacitor discharges until the output voltage goes to zero as; $V_{out} = \frac{V_{ref}T_x}{RC} + V_{out1}$. The value of *Tx* at which *Vout* becomes zero is; $T_x = \frac{V_{in}T}{V_{ref}}$.

The block diagram and integrator waveforms for the dual-slope DVM are shown in Figure 4.50. The figure illustrates the effects of the input voltage on charging and discharging phases of the converter. The total conversion time is the sum of the charging and discharging times. Yet, only the discharging time is used for the measurement and it is independent of the system components *R* and *C*, and the clock frequency.

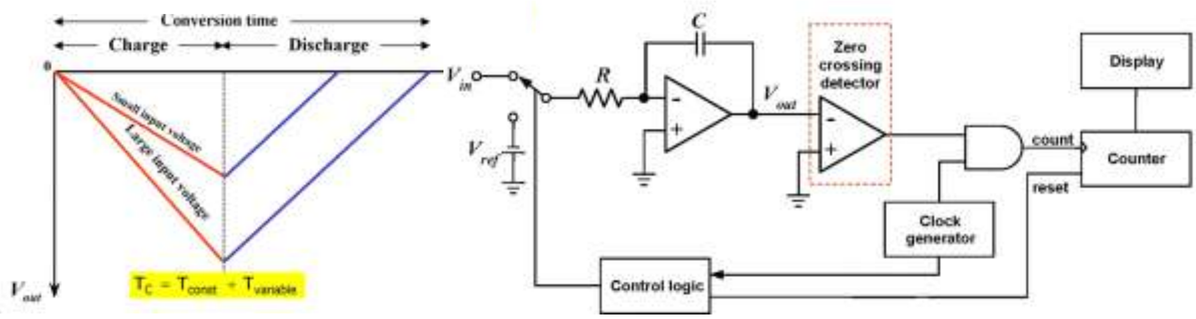


Figure 4. 50 Integrator waveform and basic block diagram of the dual-slope DVM

Example 4.21

A dual slope A/D has *R*= 100 kΩ and *C*= 0.01 µF . The reference voltage is 10 volts and the fixed integration time is 10ms. Find the conversion time for a 6.8 volt input.

$T_x = \frac{V_{in}T}{V_{ref}} = \frac{6.8V*10ms}{10V} = 6.8ms$, the total conversion time is then 10 ms + 6.8 ms = 16.8 ms

Example 4.22

A 20 V DC voltage is measured by analog and digital multimeters. The analog instrument is on its 25 V range, and its specified accuracy is ±2%. The digital meter has 3 ½ digit display and an accuracy of ±(0.6+1). Determine the measurement accuracy in each case.

Analog instrument: Voltage error = ± 2% of 25 V

= ± 0.5 V yielding;

error = ± (0.5V/20V)x100%

= ± 2.5%

Digital instrument: for 20 V displayed on a 3 ½ digit display, 1 Digit = 0.1 V

Voltage error = ± (0.6% of reading + a Digit)

= ± (0.12 V + 0.1 V)

= ± 0.22 V

Error = ± (0.22 V/20 V)x 100%

= ± 1.1%

Successive Approximation Type DVM

In this approach, the input voltage is compared to the internally generated voltage. It is the most common A/D conversion for general applications. The conversion time is fixed (not depend on the signal amplitude as in the previous cases) and relatively fast, that is; $T_C = N \times \textit{clock period}$, where N is the number of bits.

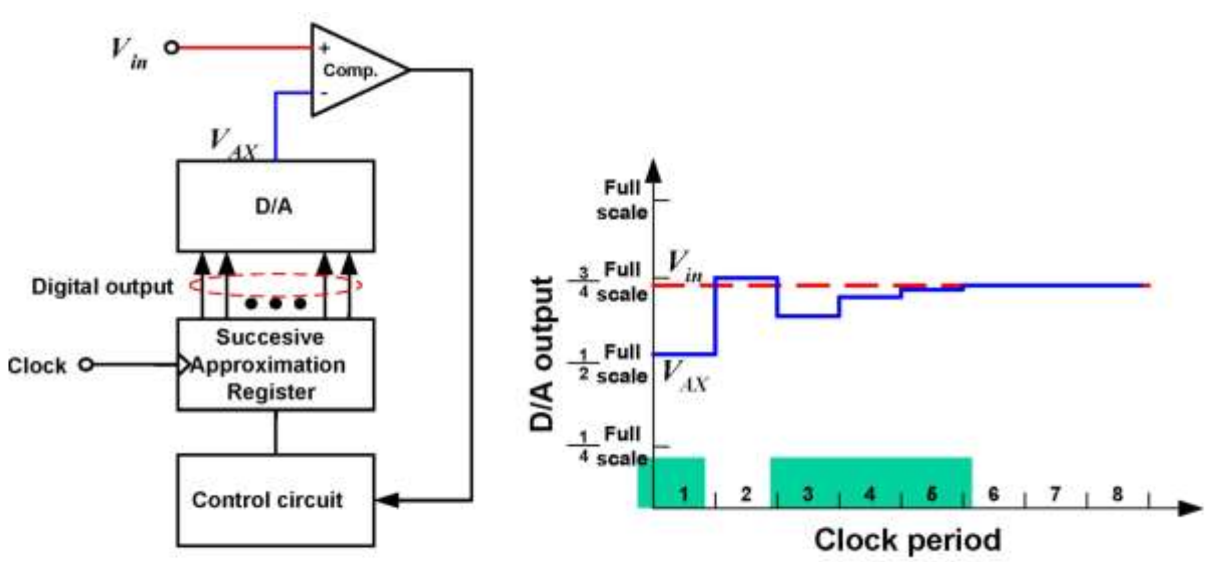


Figure 4. 51 Block diagram and output waveform of the conversion unit of the successive aproximation type DVM

The block diagram and a sample output waveform of the conversion section are shown in Figure 4.51. The block diagram looks like the one given in Figure 4.48 for the staircase type DVM except that the counter has been replaced by a successive approximation register. The register is set to the middle of the dynamic range at the beginning and the set value increases or decreases successively until the output voltage of the D/A converter approaches the input voltage with a difference smaller than the resolution of the converter. The operation of the successive approximation type D/A converter is illustrated in the following examples.

Example 4.23

Assume that we have a 9 bit binary converter. We need to determine the binary code between 0 – 511 for the input and the code to be determined is 301.

The register is set to 256 first and the output of the D/A is compared to the input. It is definitely lower than the input and the register assumes a new code that corresponds to 256 + 256/2 = 384 in step 2. This is larger than the input and the register assumes 256 + 256/4 = 320 in step 3. This is also larger and the new code in step 4 becomes 256 + 256/8 = 288 and this is smaller than the input. In step 5 the code is set to 288 + 256/16 = 304 and this is larger than the input. The code in step 6 is 288 + 256/32 = 296 and it smaller than the input. In step 7, the code is 296 + 256/64 = 300. In step 8, the code is 300 + 256/128 = 302 that is larger than the input. In step 9, which is the last step the code is 300 + 256/256 = 301 that finishes the operation.

Table 4.2 Steps of conversion for a nine bit successive approximation type A/D converter

Step	Estimate	D8 D7D6D5D4 D3D2D1D0	Result
1	256	1 0000 0000	$V_{in} > V_{AX}$
2	256 + 128 = 384	1 1000 0000	$V_{in} < V_{AX}$
3	256 + 64 = 320	1 0100 0000	$V_{in} < V_{AX}$
4	256 + 32 = 288	1 0010 0000	$V_{in} > V_{AX}$
5	288 + 16 = 304	1 0011 0000	$V_{in} < V_{AX}$
6	288 + 8 = 296	1 0010 1000	$V_{in} > V_{AX}$
7	296 + 4 = 300	1 0010 1100	$V_{in} > V_{AX}$
8	300 + 2 = 302	1 0010 1110	$V_{in} < V_{AX}$
9	302 + 1 = 301	1 0010 1101	Finished

Example 4.24

Find the successive approximation A/D output for a 4-bit converter to a 3.217 volt input if the reference is 5 volts.

(1)Set D3= 1VAX = 5/2= 2.5Volts; $V_{in} > V_{AX}$ leave D3 = 1 (1000)

(2)Set D2= 1VAX = 5/2+ 5/4= 3.75Volts; $V_{in} < V_{AX}$ reset D2 = 0 (1000)

(3)Set D1= 1VAX = 5/2+5/8= 3.125Volts; $V_{in} > V_{AX}$ leave D1 = 1 (1010)

(4)Set D0= 1VAX = 5/2+5/8+5/16= 3.4375Volts; $V_{in} < V_{AX}$ reset D0 = 0 (1010)

By this procedure, we find the output is a binary word of 1010_2.

MEASUREMENT OF ELECTRICITY

Electricity covers all aspects of our lives as the most efficient and easy way of using energy. It is the most commonly used and traded commodity in the world today. It is generated from several sources such as hydraulic, fossil fuels, sun power and nuclear fission. The nature of electrical power and energy, the ways in which it is delivered to the customers and the methods used in trade measurements are complex. The chapter provides general knowledge to electrical engineering students that they will need in their professional lives.

Utilization of Electrical Energy

Electrical Power in Resistive Loads

The rate of energy output or transfer is called the power. Capacity to do work is called the energy which is integration of power over time. The power indicates the demand for the energy. The energy is used for billing the customer for utilization of the energy.

Power is defined as $p=iv$ where v and i are the instantaneous values of the voltage and current. For constant DC, power is simply the product of the voltage and current. For AC it is not quite so simple. We can express the voltage

$$v(t) = V_{max}\cos(\omega t)$$

where ω is the radial frequency ($\omega = 2\pi f$, f is the cyclic frequency in hertz (Hz) – cycles per second). The current into a pure resistive load can be expressed as

$i(t) = I_{max}\cos(\omega t)$

The instantaneous power is

$p(t) = V_{max}I_{max}\cos^2(\omega t)$

using the trigonometric identity, $\cos^2 x = 0.5(1+\cos(2x))$

$p(t) = 0.5\ V_{max}I_{max}(1+\cos(2\omega t))$

Figure 4.52 illustrates the variation of the voltage, current and power in a resistive load during the sinusoidal cycle. On a 60 Hz single-phase system, the instantaneous power will have a waveform with a frequency of 120 Hz and varying from zero to $V_{max} * I_{max}$. (The peak voltage multiplied by the peak current.)

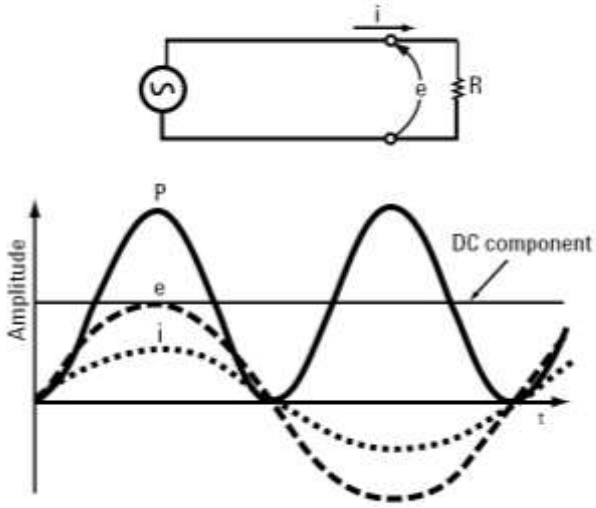


Figure 4. 52 Variation of the voltage, current and power in a resistive load during the sinusoidal cycle

The average power in one cycle of AC voltage and current applied to the load is

$$P = \left(\frac{1}{t2 - t1}\right) . \int_{t1}^{t2} v(t) . i(t) dt = I_{eff} . V_{eff}$$

Where I_{eff} is the RMS (root mean square) value of the current and V_{eff} is the RMS value of the voltage. They are defined as

$$V_{eff} = V_{rms} = \sqrt{\frac{1}{T}\int_0^T (v(t))^2} \text{ and } I_{eff} = I_{rms} = \sqrt{\frac{1}{T}\int_0^T (i(t))^2}$$

for a single frequency into a resistive load. For v(t) and i(t) expressed as in previous equations

$$V_{rms} = \frac{V_{max}}{\sqrt{2}} \text{ and } I_{rms} = \frac{I_{max}}{\sqrt{2}}$$

Electrical Power in Reactive Loads

Figure 4.53 illustrates representative voltage, current and power waveforms in a typical reactive load. The expression of the current

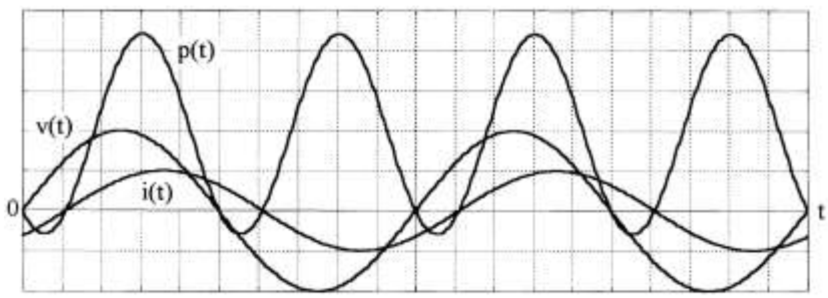


Figure 4. 53 Variation of the voltage, current and power in a reactive load during the sinusoidal cycle

becomes

$i(t) = I_{max}\cos(\omega t+\varphi)$

where φ is the angle by which the current lags (inductive) or leads (capacitive) the voltage. Figure 4.54 illustrates the phasor diagram for an inductive load. Then, the instantaneous power is

$$p=V_{max}I_{max}\cos(\omega t)\cos(\omega t+\varphi)$$

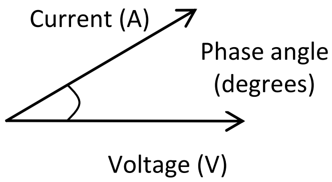


Figure 4. 54 A phasor diagram

Using trigonometric identities,

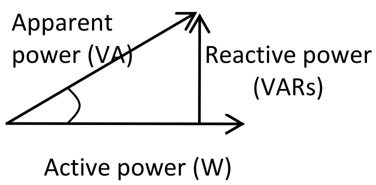
$$\cos(\omega t)\cos(\omega t+\varphi)=0.5\cos(\varphi)(1+\cos(2\omega t)) – 0.5\sin(\varphi)\sin(2\omega t))$$

We apply above equations to find the average power in one cycle yielding $P = I_{rms}V_{rms}\cos(\varphi)$

Industrial loads such as motors have both resistive and reactive components. The above equation indicates that the actual power delivered to the load can be less than the maximum possible for the effective values of the voltage and current if ϑ were zero. The maximum is called the apparent power or volt-ampere (*VA*). The ratio of real average power to apparent power is referred to as the power factor (pF).

pF = $\cos(\varphi)$ = real average power divided by apparent power, pF = P/VA

In the sinusoidal case the power factor is simply $\cos(\varphi)$ where φ is the angle by which the current leads or lags the voltage. For this reason the angle φ is often referred to as the power factor angle. A purely resistive load, one in which the voltage and current are in phase, will have a power factor of unity (1). A purely reactive load, one in which the current and voltage are out of phase with each other by ±90°, will have a power factor of zero (0).



The apparent power (VA) is the one generated and transmitted to the loads. It is expressed in Volt-Amperes (VA). It has two parts as the one converted to real work, expressed in watts and the one stored in the electromagnetic fields. The second part is called the reactive power expressed in reactive Volt-Amperes (VARs).The value of any quantity can be determined with the help of the power triangle using either the values of any other two values or any other value and the phase angle as illustrated in Figure 4.55.

Figure 4. 55 The power triangle

Figure 4.56 shows a schematic diagram to exemplify the difference between apparent power and active power. According to the ammeter and voltmeter readings, the apparent power is 464.4 VA while the power meter shows 401 Watts. This indicates that the power factor can be determined by measuring voltage , current and power. The system in the figure has a power factor of 401/464.4 = 0.86.

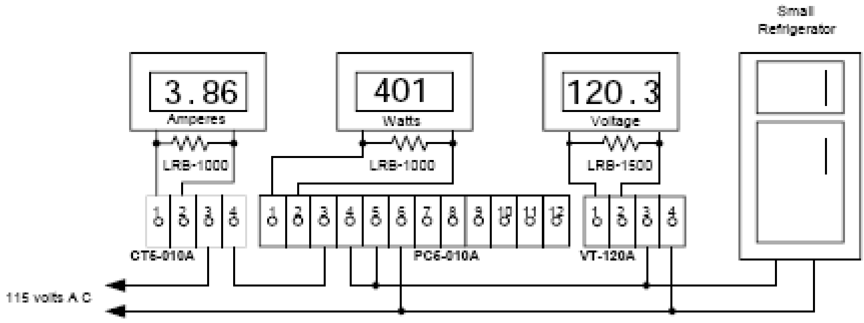


Figure 4. 56 Monitoring voltage, current, and power

Distribution of Electricity

The transmission and distribution of alternating current electricity typically ranges from 100 volts for residential consumers to 500,000 volts or greater for transmission lines. The frequency is usually 50 or 60 Hz but other frequencies (400 Hz in ships for example) are sometimes used. There are several schemes used worldwide for distribution of electricity to the customers. The power and energy measurements vary among them. The commonly used ones:

- Single phase 2-wire: a common residential service in many parts of the world which provides a single voltage, usually 100 to 240 volts.
- Single phase 3-wire: a common residential in North America which provides 2 voltages, 120 volts and 240 volts.
- Polyphase 3-wire network: common in apartment building where it provides 120 volts and 208 volts.
- Polyphase 3-wire delta: generally used in industrial operations or for a single phase motor load such as water pumping station.
- Polyphase 4-wire delta: sometimes used in supplying electricity to sparely populated rural areas. It is an economical way of providing a combination of single phase 3-wire service and a limited supply of polyphase power.
- Polyphase 4-wire wye: commonly used for industrial and commercial operations. It is widely used for electricity distribution systems, where it is transformed to other suitable service configurations.

Measuring Electric Power

Metering Electricity

Active power, reactive power and volt-ampere are commonly measured quantities. Maximum or peak power is used to determine the capacity of the generator and transmission system. Average power taken by the load in a given time interval indicates the power demand.

- Watt (W) meter: measures active electrical power, normally displayed as kW.
- Reactive Volt-Ampere (VAR) meter: measures reactive electrical power, normally displayed as kVAR.
- Volt-Ampere (VA) meter: measures apparent electrical power, normally displayed as kVA. Energy is measured by energy meters and generally used for billing the customers.
- Watt hour (Wh)meter: measures active electrical energy, integrating active electrical power with respect to time; watts x time(in hours), normally displayed as kWh.
- VAR hour (VARh) meter: measures reactive electrical energy, integrating reactive electrical power with respect to time, normally displayed as kVARh.
- VA hour (VAh) meter: measures apparent electrical energy, integrating apparent electrical power with respect to time, normally displayed as kVAh.

Electricity Metering Circuits

A power meter must sense the voltages and currents in the system to determine the power. Measurement in a single phase 2-wire system is straight forward as shown in Figure 4.57. It requires one measuring element composed of one current sensor and one voltage sensor. For polyphase systems the situation is a little bit involved. The Blondel's theorem states that in a system of N conductors, N-1 metering elements, properly connected, will measure the power or energy taken. The connection must be such that all connection coils have a common tie to the conductor in which there is no current coil. Detailed discussions of measuring circuits for various power distribution schemes is beyond the context of the present text.

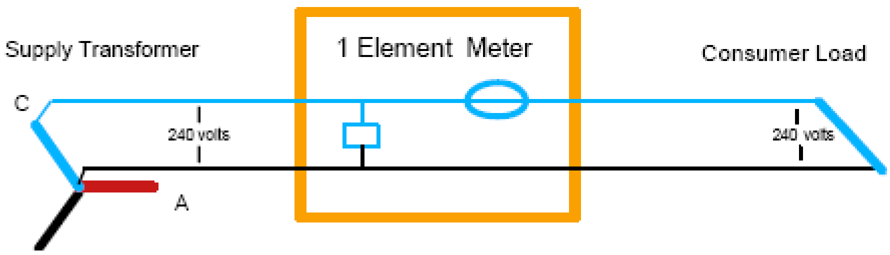


Figure 4. 57 Power meter connection for a single phase 2-wire connection

Figure 4.58 shows the wattmeter connection for a three phase, 4-wire wye service. According to Blondel's theorem (N wires -1) elements: 3 elements, each element = 1 current sensor + 1 voltage sensor provides accurate measurement.

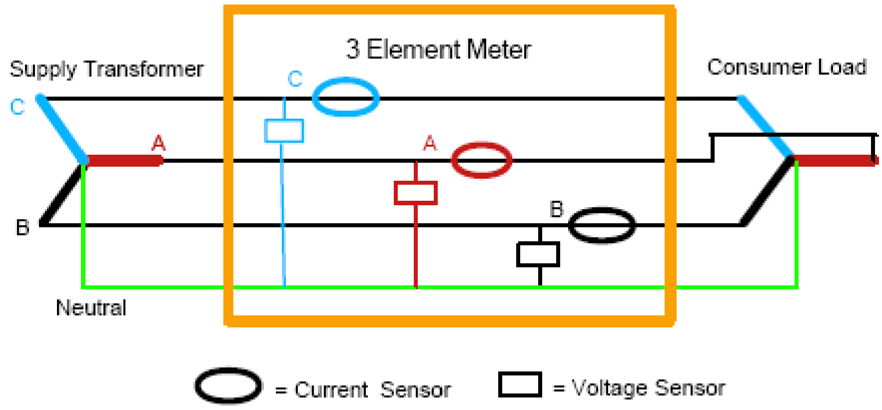


Figure 4. 58 Wattmeter connection for a three phase, 4-wire wye service

Electricity Measuring Devices

An electricity meter works on electromechanical, hybrid or electronic principles. It has four fundamental elements as sensors, multipliers, numerical conversion and registers as illustrated in Figure 4.59. Sensors provide the interface between incoming voltage and current and the metering

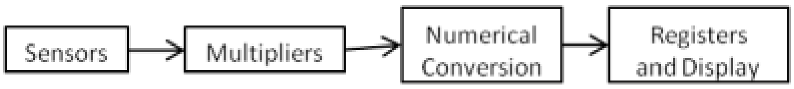


Figure 4. 59 Fundamental elements of an electricity meter

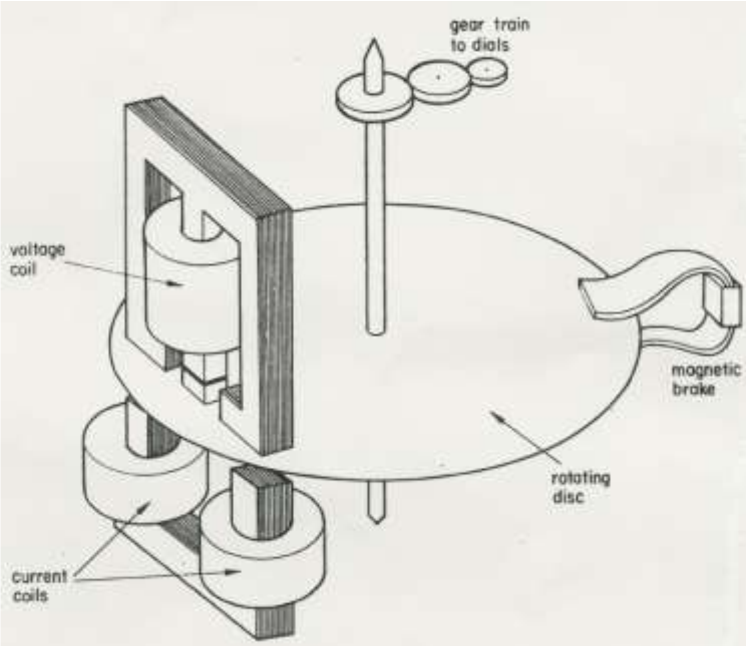
circuit. Multipliers perform the heart of the metering function by providing the product of the voltage and current the numerical conversion is the process of transforming the output of the multiplier stage into a form which can be processed by the register. And finally, registers are devices that store and display the metering quantities.

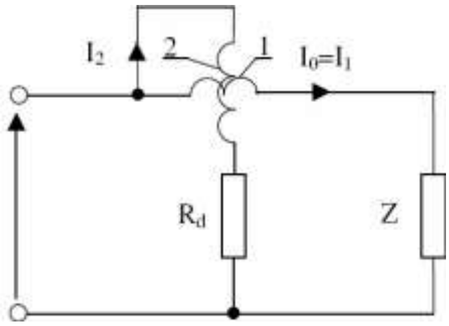
Electromechanical Meters

Mechanical means for measuring Watt-Hours are usually centered around the concept of the motor:

1. build an AC motor that spins at a rate of speed proportional to the instantaneous power in a circuit, then have that motor turn an “odometer” style counting mechanism to keep a running total of energy consumed.

2. The “motor” used in these meters has a rotor made of a thin aluminum disk, with the rotating magnetic field established by sets of coils energized by line voltage and load current so that the rotational speed of the disk is dependent on both voltage and current.





1. Current coil

Figure 4. 61 Components of an electromechanical type energy meter

ltage coil

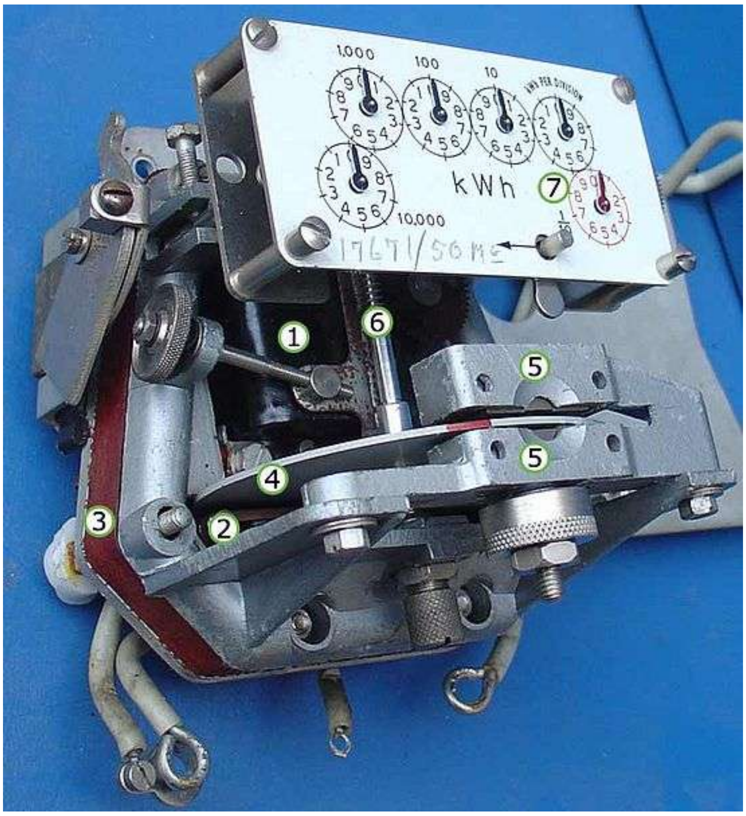


Figure 4. 60 An electromechanical energy meter

The electromechanical meter uses the induction principle that was discussed in the previous chapter. It has three main sections as the motor, braking and the gear train as illustrated in Figure 4.60 It is essentially an induction motor driving an eddy current dampening unit. The stator consists of an electromagnet and the rotor is an aluminum disc mounted on a shaft. A permanent magnet or braking system is used to keep the disc at a manageable speed. A train of gears and dials come off the disc shaft and register the energy consumed. The hybrid one is combination of electromechanical and electronic ones and used in the transition era from electromechanical to electronic technology.

Figure 4.61 illustrates an electromechanical meter. It indicates the total energy in series of dials the functions and connections are as follows:

1. Voltage coil - many turns of fine wire encased in plastic, connected in parallel with load.
2. Current coil - three turns of thick wire, connected in series with load.
3. Stator - concentrates and confines magnetic field.
4. Aluminium rotor disc.
5. rotor brake magnets.
6. spindle with worm gear.
7. display dials - note that the 1/10, 10 and 1000 dials rotate clockwise while the 1, 100 and 10000 dials rotate counter-clockwise

Electronic Meters

The electronic meter contains additional components as multiplexers, analog to digital converter, microprocessor, display/registers, communication and input/output ports, LED's and clocks as illustrated in Figure 4.62. Two basic forms of electronic metering measurement have been introduced to the industry:

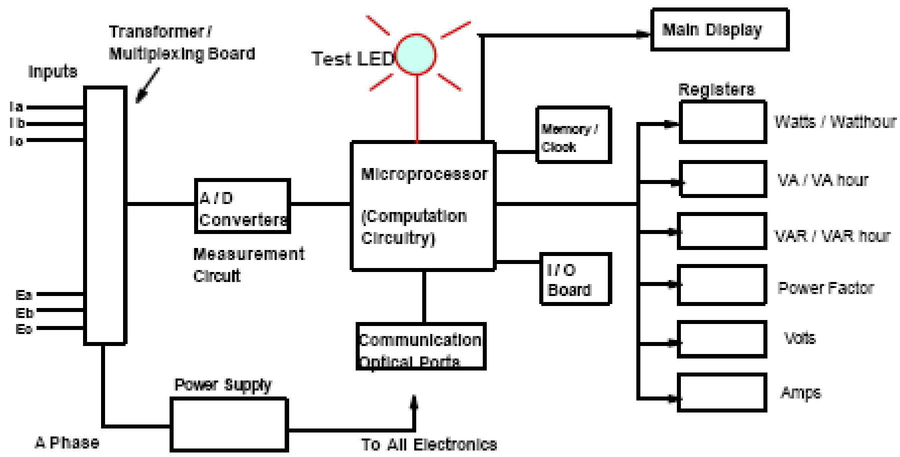


Figure 4. 62 Block diagram of a typical electronic type electricity meter

- Analog multiplying
- Digital sampling

The principle of operation of the analog multiplying type meter is illustrated in Figure 4.63. The voltage and current in a load are

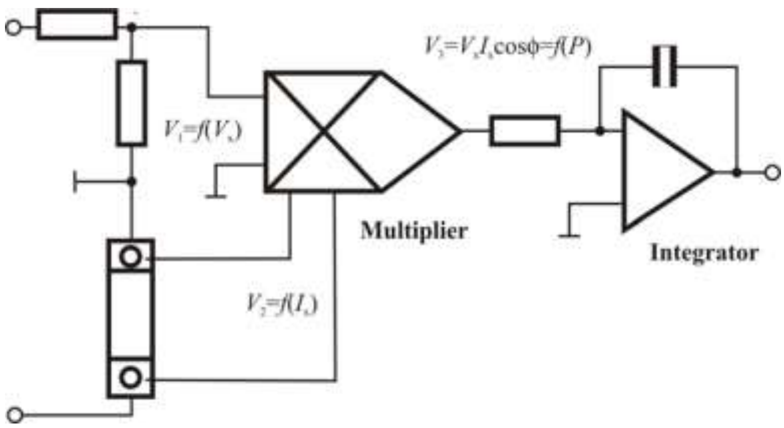


Figure 4. 63 Principle of operation of the analog multiplier type meters

sensed and the instantaneous value of the power is obtained by multiplying them via analog means. The average power is obtained by passing the instantaneous power signal through an integrator.

Analog multiplying types of meters are realized by three distinct approaches as:

- Mark-space amplitude or time division multiplexing
- Hall effect
- Transconductance

Each type will be discussed briefly in the following sections.

Time Division Multiplexing(TDM)

TDM is a well established form of electronic metering. It can be better defined as the pulse-width pulse-height multiplier. It is based on analog multiplication of instantaneous voltage and current waveforms to derive power, which is output as a series of pulses as indicated in Figure 4.64. A signal is formed with amplitude proportional to instantaneous current (*I*), and duration proportional to instantaneous volts (*V*). Average value of the waveform is equal to instantaneous power (*P*).

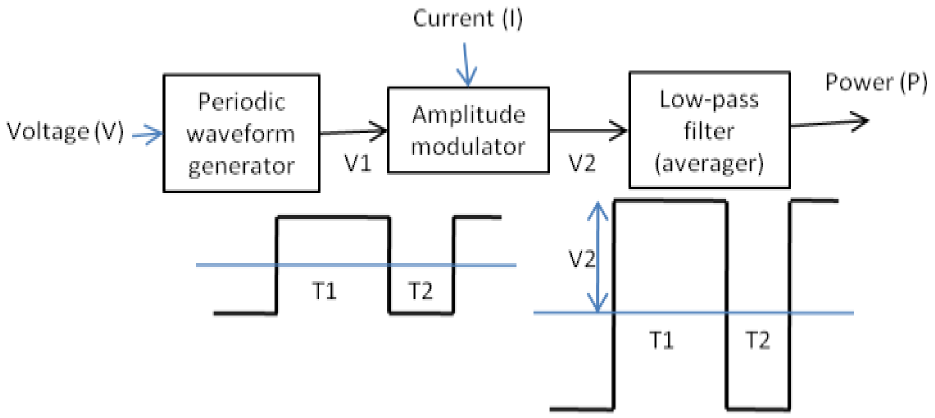


Figure 4. 64 Power to DC voltage converter using pulse-width pulse-height multiplier

$T_1 - T_2 = k_1V_1$

$V_2 = k_2I$

The average value of V_2 is

$V_{2A} = k_2I(T_1-T_2)/(T_1+T_2) = k_1k_2IV/(T_1+T_2)$

Hence, the low-pass filter / integrator yields the power as

$$P = k\frac{1}{(T_1 + T_2)}\int_0^{T_1+T_2} i(t)v(t)dt$$

Typical waveforms at various stages of the device is shown in Figure 4.65. General features of this method can be summarized as

- Good cost to accuracy ratio
- Excellent linearity and reliability
- Performance under distortion is limited
- Direct measurement limited to watts / vars
- Calibration is necessary.

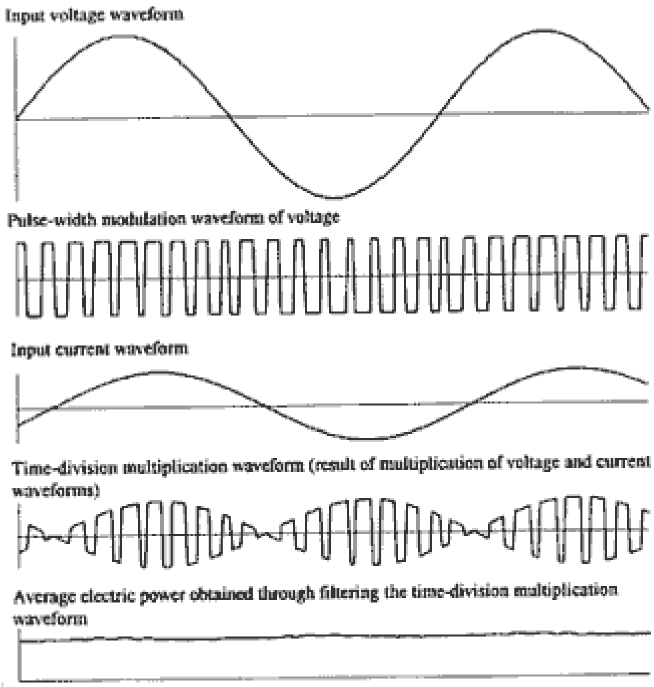


Figure 4. 65 Typical waveforms for a TDM type wattmeter

Hall Effect

A Hall probe contains a semiconductor crystal such as indium antimonide, mounted on an aluminum backing plate, and encapsulated in the probe head. If a current conducting material is placed in a magnetic field perpendicular to the direction of current flow then a voltage is developed across that material in a direction perpendicular to both the initial current direction and the magnetic field as illustrated in Figure 4.66. This voltage is called the Hall voltage.

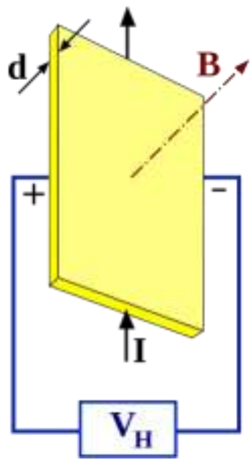


Figure 4. 66 A Hall probe

The Hall voltage arises from the deflection of the moving charge carriers from their normal path by the applied magnetic flux and its resulting transverse electric field. A voltage source with a large series resistor with the Hall cell resembles a current source that derives the cell as illustrated in Figure 4.67. The line current is used to produce a magnetic field that flows through the cell at right angles. The developed Hall voltage will be a product of the line voltage and line currents; therefore, it yields the instantaneous line power. The schematic diagram of a Hall effect type energy sensor is given in Figure 6.68.

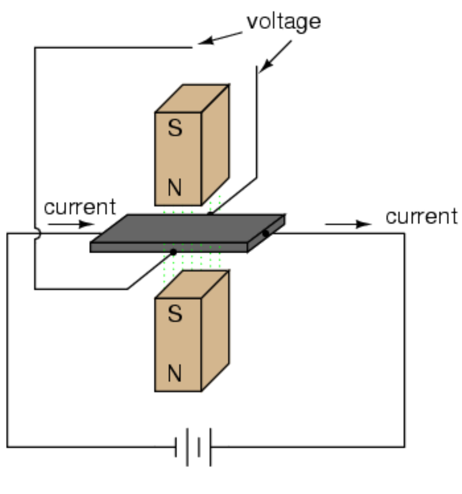


Figure 4. 67 A simplified illustration of the Hall effect sensor

$v_H(t) = R_H i(t)B(t)$, R_H = Hall constant and with $v_x(t) = ai(t)$ and $i_x(t) = bB(t)$,

$$P = \frac{1}{T}\int_0^T v_x(t)i_x(t)dt = ab\frac{1}{T}\int_0^T i(t)B(t)dt = ab\,R_H V_H$$

Analog watt transducers including Hall effect provide good accuracy even with distorted wave shapes,

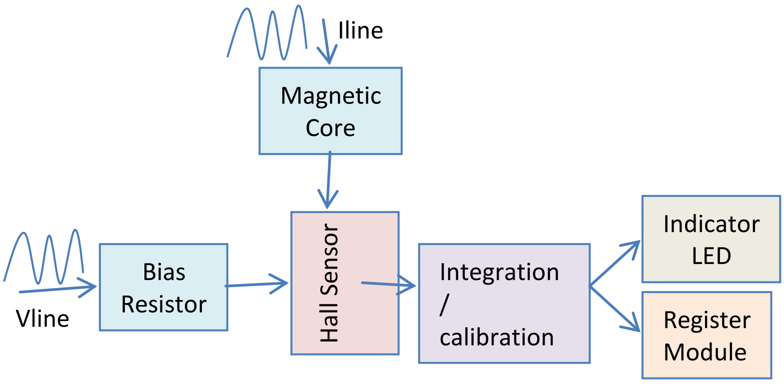


Figure 4. 68 Schematic diagram of a Hall effect type energy sensor

discontinuity, or where there is poor frequency regulation. General features of this method can be summarized as

- Very cost effective technology
- Can measure Watt /VARs, but not VA
- Linearity less than TDM technology
- Excellent response for harmonic content
- Susceptible to large temperature changes.

Transconductance

A transconductance device produces an output current (*Io*) proportional to the input voltage

$$I_0 = G_T(V_2 - V_1)$$

as illustrated in Figure 4.69. The proportionality coefficient (the transconductance - G_T) is a linear function of the bias current *Is*:

$$G_T = \beta I_s$$

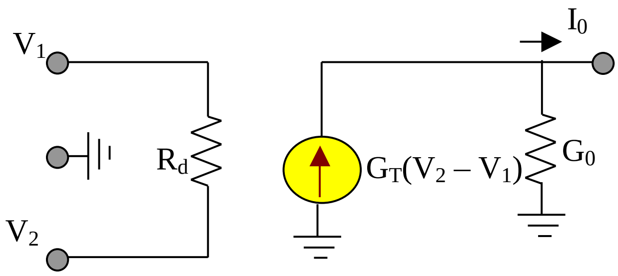


Figure 4. 69 Symbolic representation of a transconductance device

Where β is the proportionality coefficient which is constant over a wide range of the bias current. Combining previous equations and calling $V_i = (V_1 - V_2)$ yields

$$I_0 = \beta I_s V_i$$

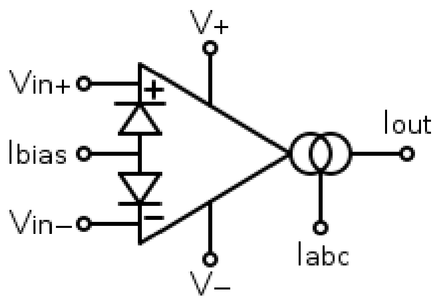


Figure 4. 70 Symbolic diagram of CE 3080 Operational Trasnconductance Amplifer (OTA)

So, the input voltage can be amplitude modulated if the modulating signal is used to vary *Is*. The bias current must flow inward all the time. The device works as a two quadrant multiplier. The output current is converted into an output voltage as it flows through a fixed resistor.The operational tansconductance amplifier (OTA) was developed originally by RCA company for the US military. Then it has been marketed for the general users as CA 3080. The symbolic diagram of the device is shown in Figure 4.70. I_{abc} is the same as I_s in the previous equation.

The transconductance is another form of metering that incorporates both TDM and Hall Effect technology by conducting analog multiplication of the line voltage and currents to produce a single voltage signal proportional to the line power. An ordinary transistor type differential amplifier can also work as a transconductance amplifier. The secondary current from the meter transformer is converted to a voltage and applied the bases of the two transistors. The line voltage is applied between the collectors and emitters of the transistors. A potential difference between the two collectors is generated. This voltage is the product of the line voltage and line currents and therefore proportional to the line power.

The transconductance type power meters possess excellent cost to accuracy ratio. However, it requires four quadrant amplifier for superior performance under varying power factors and harmonic distortions.

Digital Sampling

Digital sampling is the only technology that does not use analog values of voltage and current. In this process, the analog values of voltage and currents are converted to digital data prior to any multiplication taking place as illustrated in Figure 4.71.

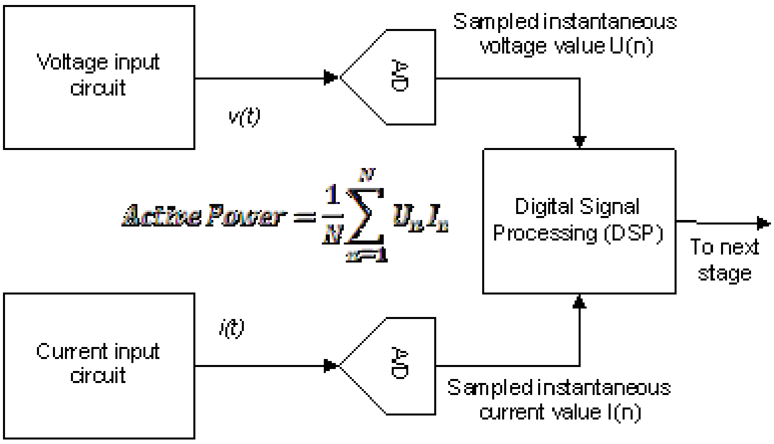


Figure 4. 71 A general view of a digital sampling based power meter

In general, the following equation is used to express the effective power, where the instantaneous power values, products of the instantaneous voltage and current values, are integrated and averaged by cycle T.

$$P = \frac{1}{T}\int_0^T u(t)i(t)\,dt \cong \frac{1}{T}\sum_{n=1}^{T/\Delta t} u(n)i(n)\Delta t$$

Where

u(t): instantaneous voltage value at time t i (t) instantaneous current value at time t

u (n): instantaneous voltage value at n^{th} sample

i (n) instantaneous current value at n^{th} sample

T : cycle (period)

From the approximation on the right above, we can tell that the effective power can be obtained by averaging the number of n (=T/ Δt) segments of the width of Δt by cycle T. A digital sampling wattmeter executes this computation almost as is. In an actual wattmeter, the waveform measurement time period is often set longer than one cycle. Δt is generally around tens of microseconds and the sampling frequency is the inverse number of the Δ t.

With N = T/Δt; $Active\ Power\ = \frac{1}{N}\sum_{n=1}^{N} U_n\, I_n$

A group of sample includes a sample of voltage and current on each of the three lines. Two consecutive cycles have samples that are 34 microseconds apart, this is called sample migration and ensures that each group of samples is not taken at an identical point during the cycling of the signal. Figure 4.72 shows typical sampled voltage, current and instantaneous power waveforms.

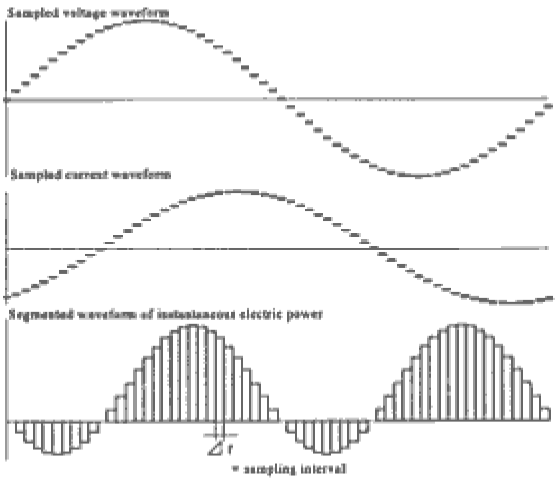


Figure 4. 72 Typical waveforms in digital sampling

Most inaccuracies can be fully compensated algorithmically eliminating the need for any physical calibration of the meter. Not very cost effective technology for single phase residential compared to TDM, Hall effect and transconductance technologies.

- Advantages:
  - Ability to handle complex billing rates
  - Increased accuracy
  - Ability to measure various quantities, one device
  - Ability to collect meter data remotely
  - Ability to program meter remotely
  - Have time saving features
  - Ability to measure all four quadrants
- Disadvantages:

- More sophisticated testing apparatus required
- More accurate reference standards required
- More advanced training is required.

When measured with digital sampling type instrumentation, the powerful micro-processors can run statistical routines to reveal computed data, oriented to particular customer requirements.

PROBLEMS ON MEASURING INSTRUMENTS

Review Questions

1. How do you measure voltage in a circuit?
2. What is an ammeter and how it is connected in a circuit to take a measurement?
3. How the electromagnetic torque is established in a moving coil?
4. What is the function of the balancing spring in a moving coil instrument?
5. Why the scale of commonly used moving coil instruments are circular?
6. What is the parallax error and how it effects the accuracy of the measurement?
7. What is a galvanometer and how it is used as a measuring instrument?
8. How do you construct a basic MC based ammeter?
9. How do you make a basic MC based basic voltmeter?
10. What is the difference between make-before-break and break-before-make type switches?
11. Why do you need a multi-range ammeter and it can be built from a basic MC meter?
12. How a MC based multi-range voltmeter can be constructed from a basic ammeter?
13. What is an ohmmeter and how it can be constructed from a simple MC based ammeter?
14. What is the loading error and how it effects the measurements?
15. What is the RMS value of a waveform and how it differs from the average value?
16. What are the ways of generating a DC signal representing an AC signal?
17. Why the full-wave rectification is preferred over the half-wave rectification in AC voltmeters?
18. What is the waveform factor?
19. What is the waveform error involved in an AC voltmeter?
20. What is the correction factor for AC and triangular waveforms?
21. How does the clamp-on ammeter work and what are the advantages over a regular ammeter?
22. What is the true RMS meter and what are the ways of realizing it?
23. What is an electronic counter and how it measures the time interval between two events?
24. What are the limitations of an oscilloscope in measuring frequency of a signal?
25. What is the role of a comparator in electronic counters?
26. What is the BCD counter and how it differs from an ordinary binary counter?

27. What is the significance of the time-base in counters?
28. What are the sources of errors in counters?
29. What is a shaft encoder and how it can be used to measure the rotational speed?
30. What are the advantages of digital voltmeters over analog counterparts and oscilloscopes as a voltage measuring device?
31. What is the sample and hold circuit as used in analog to digital converter (ADC)?
32. How does the integrating type ADC work?
33. What are the advantages of dual-slope integration over a single-slope integration in DVM?
34. What is the successive-approximation type ADC and what are its advantages in DVM?
35. How does the electrical power differ in resistive and reactive circuits?
36. What is the reactive power?
37. What is the power factor?
38. How the electricity is distributed in residential areas in Jeddah?
39. What are the techniques that use analog multiplication as used in measuring electricity?
40. How does the time division multiplexing involve in electricity meters?
41. What is the hall effect device and its function in electricity meters?
42. What is a transconductance amplifier and how it is used in electricity meters?
43. How the digital sampling is used in measuring electricity?
44. What are the advantages of digital electricity meters?

Solved Examples on Moving Coil Instruments

1. A moving coil has 100 turns, 3 cm^2 coil area, and air-gap magnetic flux density of 0.1 Tesla (Wb/m^2). The control spring exerts a torque of $3x10^{-7}$ N-m at the full-scale deflection of 100°. The potential difference across the coil terminals at the full-scale deflection is 5 mV. Using the above movement:
   - Find the full scale deflection current and coil resistance;

$I_{fsd}=T_{SP}/NBA$ = 0.1 mA, therefore $R_m= V_m / I_{fsd}$ =50 Ω

   - Design a DC ammeter with a range 0-50 mA;

R_{sh1}= 5 mV/ (50-0.1) mA = 0.1 Ω

   - Design a multi-range DC voltmeter with ranges 0-10 V and 0-200 V.

For voltmeter ranges, R_m is negligible: R_{s1} = 10V/0.1mA = 100 kΩ and R_{s2} = 2 MΩ

   - What would be the deflection angle for an input voltage of 7 V in 0-10 V range?

Since 10 V causes 100°, 7 V will cause 70° of deflection

2. A moving coil has 80 turns, 4 cm^2 coil area, and air-gap magnetic flux density of 0.1 Tesla (Wb/m^2). The control spring exerts a torque of $4x10^{-7}$ N-m at the full-scale deflection of 90°. The potential difference across the coil terminals at the full-scale deflection is 10 mV. Using the above movement:
   - Find the full scale deflection current and coil resistance;

$I_{fsd}=T_{SP}/NBA$ = 0.125 mA = 125 μA, therefore $R_m= V_m / I_{fsd}$ =80 Ω

   - Design a DC ammeter with a range 0-100 mA;

R_{sh1}= 10 mV/ (100-0.125) mA ≅ 0.1 Ω

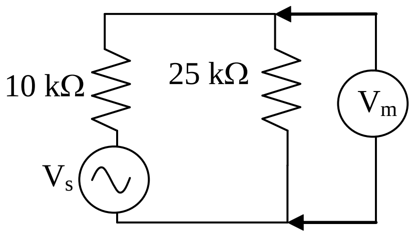
   - Design a multi-range DC voltmeter with ranges 0-100 V and 0-200 V.

For voltmeter ranges, R_m is negligible: R_{s1} =100V/0.125mA=800 kΩ and R_{s2} = 1.6 MΩ

   - What would be the deflection angle for an input voltage of 65 V in 0-100 V range?

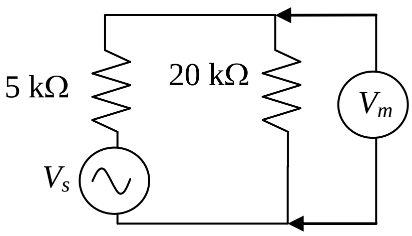
Since 100 V causes 90°, 65 V will cause 65x90/100 = 58.5° of deflection.

3. A D'Arsonval (moving coil) movement based AC voltmeter is calibrated to read correctly the RMS value of applied sinusoidal voltages. The meter resistance is 1000Ω/V, it is used in 0 – 100 V range and the scale has 50 divisions. The meter reads V_m = 50 V (RMS)



   a. Find the % error in the measured voltage due to reading error assuming that you can read down to half of the smallest scale divisions accurately. *Smallest scale division is 2 V yielding a reading error of 1V; 1*100/2 = 2%*
   b. Find V_s if it is a sinusoidal waveform with zero average. Rm = 100 kΩ, Rl' = 25*100/125 = 20 kΩ, I = 50/20 = 2.5 mA, Vs = I*(20+10)*10^3 = 75 V
   c. Find the loading error in (%). True reading with an ideal voltmeter would be = 25*75/35 = 53.57 V, error = (53.57 – 50)*100/53.57 = 6.67 %
   d. Find the total error in this measurement. *Reading + Loading = ±8.67 %*

4. A D'Arsonval (moving coil) movement based AC voltmeter is calibrated to read correctly the RMS value of applied sinusoidal voltages. The meter resistance is 4000Ω/V and it is used in 0 – 50 V range.



- Find V_s if it is sinusoidal and V_m = 36 V (RMS)

The meter resistance is 4000(Ω/V)x50V = 200 kΩ is parallel with 20 kΩ yielding R_L' = 18.18 kΩ. $I = V_m/R_L'$ = 1.98 mA. V_s = 5x1.98 + 36 = 45.9 V (rms)

The periodic waveform $v_m(t)$ shown is applied to the meter.

- **Calculate** V_{RMS} for this waveform,

$$V_{RMS}^2 = \frac{1}{3}[\int_0^1 10000xdt + \int_2^3 2500xdt]$$; V_{RMS} =70.71V,

- How much is the voltage indicated by the meter ($V_{indicated}$)?

Average value of the rectified signal = 66.67V $V_{indicated}$ = 1.11xV_{AV} = 74 V

- Find the waveform error in this measurement.

% error = 100x(74-70.71)/70.71 = 4.65%

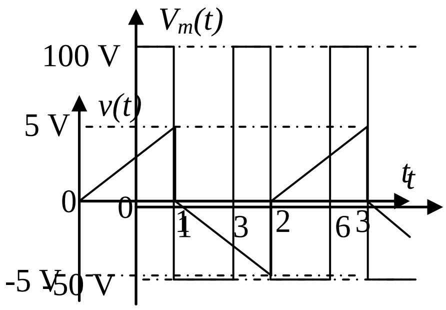
5. An AC voltmeter calibrated for sinusoidal voltages is used to measure both the input (V_1) and output (V_2) voltages. It has a scale with 100 divisions and measurement ranges: (0 – 50) mV; (0 – 100) mV; (0 – 500) mV; (0 – 1) V; (0 – 2) V; (0 – 5) V and (0 – 10) V
   - Determine the range that would yield the most accurate reading for V_1, the value indicated by the meter for V_1 and percentage reading uncertainty (assume that the reading uncertainty is $\pm$0.5 division).

The meter would indicate 1.11V_{AV} =1.11x0.636xV_{peak} = 28.27 mV. Hence, range (0 – 50) mV is the most accurate with uncertainty $\pm$0.25 mV $\rightarrow$ 0.88%

- Repeat (a) for V_2.

V_{ind} = 1.11x0.636x1.5 = 1.06 volt; range (0 – 2) V, uncertainty $\pm$ 0.01 V$\rightarrow$ 0.94%

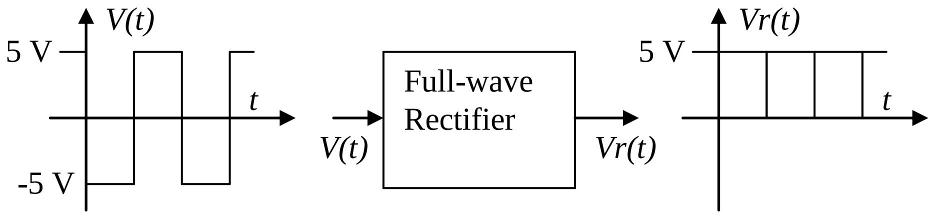
6. An average reading full-wave rectifier moving coil AC voltmeter is calibrated to read correctly the RMS value of applied sinusoidal voltages. The periodic waveform $v(t)$ shown is applied to the meter. **Calculate** V_{RMS} for this waveform, $V_{indicated}$ and the waveform error in it.



The full-wave rectifier will convert the input waveform into a saw tooth voltage waveform of question-4.3 with amplitude 5 volts and period T = 1 second. Using the equations in answer-4.3, V_{AV} = 2.5 V; $V_{RMS} = 0.577V_m$ = 2.89 V. The value indicated by the meter $V_{ind} = 1.11xV_{AV}$ = 2.775 V. Therefore, *%(waveform) error* = 100x(2.775 – 2.89)/2.89 = -4%

- **Draw** the circuit diagram and **explain** the operation of the full-wave rectifier bridge circuit used to convert D'Arsonval movement into an AC voltmeter.

Please refer to the lecture notes for the operation of the full-wave rectifier.

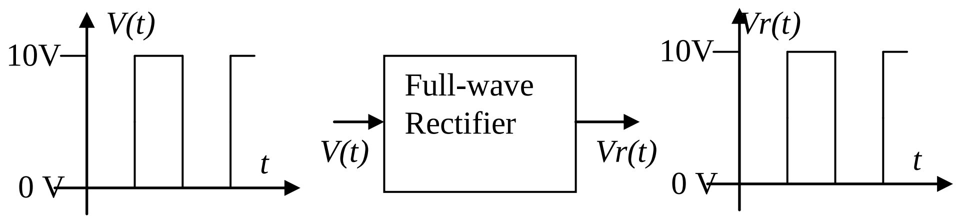


- **What** is the V_{RMS} for a zero averaged square waveform of peak to peak value = 10 V? **What** is the value indicated for it by the AC voltmeter calibrated to read applied sinusoidal voltages correctly? **What** is the percentage waveform error in that value?

The zero-averaged square wave has a magnitude ± 5 V. The magnitude becomes + 5V after the full-wave rectification for all times. $V_{RMS} = V_{AV}$ = 5 V. The meter calibrated for sinusoidal voltages will read V_{ind} = 1.11x5 = 5.55 V. Hence, *the % error* = 100x(5.55 – 5)/5 = 11 %

- **Repeat** (a) if the square wave accepts amplitude values between 0 and 10 volts.

The output of the full-wave rectifier will be the same as it's input as shown in the figure. V_{AV} = 5 V



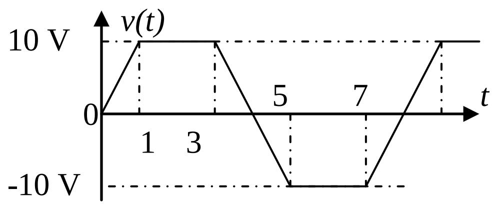
and V_{ind} = 5.55 V. The RMS voltage is different as: $V_{RMS} = \sqrt{\frac{1}{T}\int_0^{T/2} 100 \times dt} = 7.07V$.

Yielding, *the % error* = 100x(5.55 – 7.07)/7.07 = -21.5%

- **Explain** the operation of **one** circuit through which the D'Arsonval movement can be used as a meter for measuring periodic signals. What is the scale factor for calibrating such a meter?

The meter based on D'Arsonval movement inherently measures (IM) the average value of the input applied. Therefore, a zero-averaged AC input voltage would cause "V_{IM}=0" as the displayed value. The full-wave rectifier converts the AC input voltage into a waveform that is equal to the absolute value of the input. Hence, the negative half-cycle also produces a positive voltage at the output. Eventually the average of the output becomes $2V_m/\pi$, where V_m is the peak value of the voltage yielding $V_{IM} = 2V_m/\pi = 0.636V_m$ The actual value that we want to measure is the RMS value which is $V_{RMS} = 0.707V_m$ If the reading is not corrected, there will be 10% error in it. The scale factor $SF = 1.11 = V_{RMS}/V_{AV}$ is used to correct the reading and eliminate the reading error.

- What is the V_{RMS} for the waveform shown? What is the value indicated by an AC voltmeter calibrated for sinusoidal waveforms? What is the percentage waveform error in that value?

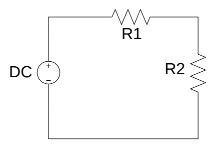


Due to symmetry, V_{RMS} can be calculated from 0 to 4 seconds as:

$V_{RMS}{}^2 = \frac{1}{4}[2x\int_0^1 100t^2dt + \int_1^3 100dt] = \frac{200}{3}$ yielding V_{RMS} = 8.16 V. The average value is computed in a similar manner as: $V_{AV} = \frac{1}{4}[2x\int_0^1 10tdt + \int_1^3 10dt] = \frac{30}{4} = 7.5V$ The voltage reading indicated by the meter is: $V_{ind} = 1.11xV_{AV}$ =8.325 V. *%error* = 100x(8.325 – 8.16)/8.16 = 2.2%

Questions with Solutions

The circuit shown has a DC voltage source driving a circuit formed by two resistors R_1 and R_2. The source voltage is 50 V, R_1 = 15 kΩ and R_2 = 10 kΩ.



1. How much is the voltage across R_2? Ans. 20 V
2. Assume that you measure the voltage across R_2 using an analog instrument on its 25 V range, meter resistance 1 kΩ/V, and its specified accuracy is ±2% of full scale. Determine the measured value and measurement accuracy.

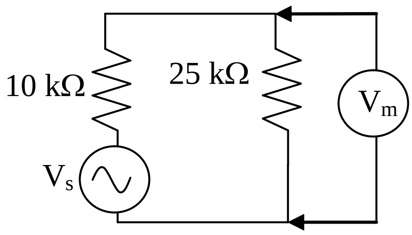
Ans. Meter resistance is 25 kΩ that comes in parallel with R_2. $R_2' = R_2//R_m$ = 7.14 kΩ, the meter can read down to ±0.5 V (2% of full scale) accurately. Hence, we can read the voltage down to the doubtful digit which is the first decimal yielding V_2' = 16.0 V. Loading error is 4 V and the total error is 4.5 V yielding ±22.5%

3. Assume that you measure the voltage across R_2 using a digital instrument with a 3 ½ digit display, meter resistance 10 MΩ and an accuracy of ±(0.5% of reading + 1digit). Determine the measured value and measurement accuracy.

$R_m >> R_2$ meaning that the loading error is negligible. The meter will display 20.0; the digit error is 0.1 V and the instrument error is 0.1 V as well. The total error is 0.2 V yielding $\pm$1%

General Questions

1. A D'Arsonval (moving coil) movement based AC voltmeter is calibrated to read correctly the RMS value of applied sinusoidal voltages. The meter resistance is 4000Ω/V, it is used in 0 – 100 V range and the scale has 50 divisions. The meter reads V_m = 50 V (RMS)



   a. Find the % error in the measured voltage due to reading error assuming that you can read down to half of the smallest scale divisions accurately.
   b. Find V_s if it is a sinusoidal waveform with zero average with and without 25 kΩ (i.e. output is open circuit).
   c. Find the loading error in (%).
   d. Find the total error in this measurement.
2. Draw the simplified functional diagram of an electronic counter for period measurement and label each block clearly. Indicate sample signals that would appear at various stages. What are the advantages of electronic counters in frequency measurement? Why we prefer measuring the period and calculating the frequency from it for low frequency signals?
3. For the digital (electronic) counter:
   a. Explain the function of the input signal shaper.
   b. Explain the function of the time-base generator.
   c. What will be the number displayed if the counter is in frequency mode, time-base is set to 1 msec and the frequency of the input signal is 985,756 Hz? How much is the uncertainty in the frequency reading?
4. Averaging is used in period measurement.
   a. What is the function of averaging used?
   b. It reduces the uncertainty in data. Prove that if N independent periods are used in averaging, each with uncertainty $\pm \Delta T$, the uncertainty in the averaged period is $\pm \frac{\Delta T}{\sqrt{N}}$
5. Assume the clock frequency is 1MHz and uncertainty is 1%. It is used to obtain a gating pulse with 1 second. How much is the percentile uncertainty in the pulse duration?
6. The electronic counter can be used for measuring the time period of periodic signals. Show that the uncertainty in the measurement can be reduced by a factor of $\frac{1}{\sqrt{N}}$ if the average of N time

periods is taken. Hint: $T_{AV} = \frac{1}{N}(T_1 + T_2 + \cdots + T_N)$ The T_i's are statistically independent, $T_i = T \pm \omega_T$,∀i

7. An electronic counter is used in period mode for measuring low frequencies
   a. Why the counter is used in the period mode?
   b. If the counter reading is T=120333.0 μs, what is the gating uncertainty in T?
   c. How much is the nominal frequency and percentage uncertainty in the frequency?
8. In the stroboscopic method of rotative speed measurement, two successive flash frequencies f_1 and f_2 that produce a single stable image are f_1=41.1Hz ± 2%, f_2=19.9Hz ± 2%.
   a. Show that the shaft speed is $v = \frac{f_1 f_2}{f_1 - f_2}$
   b. Calculating *v* from the above formula, find its nominal value and percentage uncertainty.
9. A digital voltmeter uses 3½ digit display (it can display up to 1999). It is used to measure a voltage across a standard cell whose value is 1.234 volt 5 times and following readings are obtained: 1.2202, 1.2115, 1.2456, 1.2218. Determine the accuracy, the precision and the bias of the voltmeter.
10. The digital voltmeter is of positive ramp type. The clock (time-base) runs at 1 MHz. The slope of the ramp is 1000 volt/s. The voltage applied for the measurement is 1.5 volt DC. Draw the block diagram of the digital voltmeter and sketch the diagram for voltage to time conversion. Then, determine the duration of the gate signal produced as a result of the voltage-to-time conversion and number of clock pulses applied to the counter.
11. Draw a simplified block diagram of ramp-type digital voltmeter and label each block clearly. Show sample signals at various stages. State the advantages of voltage measurement using a digital voltmeter.
12. For a ramp-type digital voltmeter:
    a. Explain the function of the time-base oscillator.
    b. Explain the voltage to time conversion.
    c. How the polarity of the voltage is identified?
    d. Assume that the number displayed is -10.025 V. How much is the uncertainty in the voltage reading?
    e. What is the significance of the sample rate?
    f. What are the factors affecting the accuracy of the measurement?
13. What are the similarities and differences between electronic counters and digital voltmeters?
14. A dual slope A/D has R= 100 kΩ and C= 0.01 μF . The reference voltage is 10 volts and the fixed integration time is 10ms. Find the conversion time for a 6.8 volt input.

15. Find the successive approximation A/D output for a 4-bit converter to a 3.217 volt input if the reference is 5 volts.
16. A 20 V dc voltage is measured by analog and digital multimeters. The analog instrument is on its 25 V range , and its specified accuracy is ±2%. The digital meter has 3 ½digit display and an accuracy of ±(0.6+1). Determine the measurement accuracy in each case.

BIBLIOGRAPHY

Further Reading

J.G Webster, *The measurement, instrumentation, and sensors handbook*, 1999, isbn=3540648305

Useful Websites

Hall effect: http://www.allaboutcircuits.com/vol_2/chpt_12/3.html

http://www.ife.p.lodz.pl/downloads/Korczynski/Electronic_Measurement%20_III_sem%20/Experiment%202%20Power%20v%2001.pdf

http://cp.literature.agilent.com/litweb/pdf/5988-9213EN.pdf

http://www.yokogawa.com/ymi/tutorial/tm-tutorial_wt_12.htm

OSCILLOGRAPHIC MEASUREMENTS AND PICTURE DISPLAYS

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Express principles of waveform displays.
2. State advantages of oscilloscope displays in measurement.
3. Discuss principles of waveform displays on an oscilloscope.
4. Draw a simplified block diagram of an oscilloscope and explain the principle of operation. Express the electrostatic deflection on an oscilloscope screen and discuss the significance of operation at sweep mode.
5. Measure voltage and time information from the oscilloscope display.
6. Explain the need for the triggered-sweep mode of operation.
7. Explain the display of high frequency signals and function of the delay line.
8. Express the operation of the oscilloscope in X-Y mode and interpret the lissajous figure.
9. Describe the advantages of multi-trace oscilloscope.
10. Explain how to obtain multiple traces from a single electron gun.
11. Express the necessity and state advantages of digital storage oscilloscopes (DSO).
12. Illustrate the principle of operation of the DSO.
13. Explain the operation and function of DSO.
14. Report current trends in DSO technology.
15. Define virtual instrumentation and its functions.
16. Identify the components of a virtual instrumentation system.
17. Point out the use of virtual instrumentation in system design.
18. Illustrate the principle of generation of a picture display.
19. Define the picture element (pixel).
20. State the standards and resolution in picture displays.
21. Discuss the CRT based picture displays.
22. Describe the principles of operation for liquid crystal displays (lcd).
23. Explain the raster scan as a mean of painting the screen.
24. Define the aspect ratio and viewable area for a display screen.
25. Compare and contrast CRT and lcd type displays.
26. Name new emerging display technologies and state their principles of operation.

WAVEFORM DISPLAY DEVICES

The signal is a physical variable (such as force, velocity, voltage, current etc) associated with a system and it is almost always a function of time. A *waveform* is a graphic representation of a wave. It is a necessity for engineers to observe waveforms for various signals in order to make certain measurements and compare them to each other. This requires conversion of the physical variable into a trace through a writing mechanism and a medium over which the information can be imprinted. What is measured is the distance between certain points on the marked trace that is a representative of the waveform for the physical variable concerned. The type and technique of display affect the quality of the measurement.

Figure 5. 1 A wave-form recording device

For inscribing the variations of a signal in time, a pen and paper can be used. In this case, the writing pen moves vertically in response to the magnitude of the signal while the writing medium (paper) moves horizontally at a constant speed as illustrated in Figure 5.1. This is called *y-t recording* since the signal is represented on vertical axis (y-axis) and the horizontal axis represents the time. In some applications the horizontal motion is controlled by another signal rather than time. This recording is called *x-y recording*. However, this technique is limited to recording low frequency applications since the mechanical parts cannot respond to high frequency signals. Oscilloscopes are used to display high frequency signals.

An oscilloscope measures voltage waves. One *cycle* of a wave is the portion of the wave that repeats. A voltage waveform shows time on the horizontal axis and voltage on the vertical axis. Oscilloscopes are electronic equipment mainly used in displaying and measuring electrical voltage signals. Other physical signals can be displayed through proper sensors. The writing pen in this equipment is the electron beam and writing medium is a special screen that glows when the electron beam strikes on it. The electron beam can be deflected from its straight path using electrical or magnetic fields, hence easily moved across the screen. Eventually a spot of light that can be placed on different locations on the screen under the control of external electrical signals becomes available. For *y-t* recording, the spot travels horizontally across the screen at a constant speed and moves vertically in response to the magnitude of the input signal. Intensity or brightness of the display is sometimes called the *z* axis as illustrated in Figure 5.2. The trajectory looks like a bouncing ball that moves across the screen and the human eye can follow it if the motion is slow. If the light ball draws the same trajectory on the screen for more than about 24 times a second, the human eye can not follow the motion and it will see it as a fixed trace on the screen. This chapter deals with measurements using oscilloscopes and principles of picture display devices.

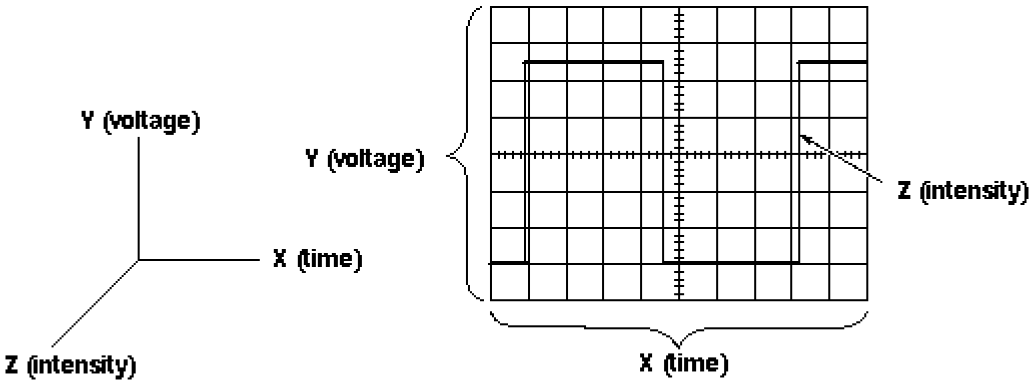


Figure 5.2 Waveform displayed on an oscilloscope

Operating Principles of an Oscilloscope

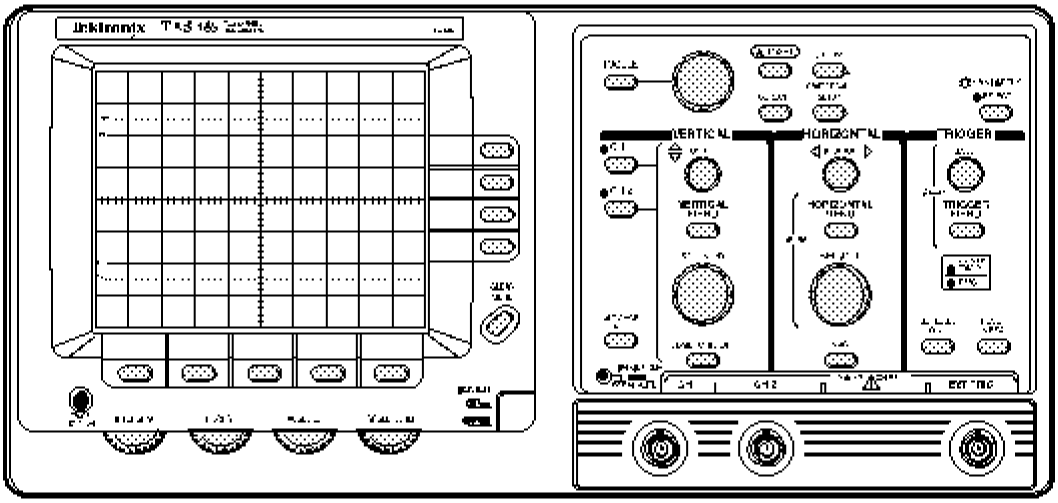


Figure 5.3 Front panel of an analog oscilloscope (Tektronix TAS 465)

Oscilloscopes can be classified as analog and digital. To better understand the oscilloscope controls, we need to know a little more about how oscilloscopes display a signal. Analog oscilloscopes work somewhat differently than digital oscilloscopes. However, several of the internal systems are similar. Analog oscilloscopes are somewhat simpler in concept and are described below. Front panel of an oscilloscope is shown in Figure 5.3. It has a display screen with a 8 cm by 10 cm grid drawn on it. The display has controls for the intensity (brightness of the trace), focus and astigmatism (sharpness of the trace). On the right hand side there are control sections for vertical, horizontal, and trigger controls and input connectors. The oscilloscope is a versatile instrument that can be used for measuring signal voltages from a few millivolts up to hundreds of volts. Depending on how we set the vertical scale (volts/div control), an *attenuator* reduces the signal voltage or an *amplifier* increases the signal voltage. One *cycle* of a wave is the portion of the wave that repeats. In general use, only a few cycles are displayed. For analog oscilloscopes, this specification indicates how fast the trace can sweep across the screen, allowing us to see fine details. The fastest sweep speed of an oscilloscope is usually given in nanoseconds/div.

Simplified Block Diagram of an Oscilloscope

The simplified block diagram of a general-purpose oscilloscope is shown in Figure 5.4. The heart of an oscilloscope is the cathode ray tube (CRT): electron gun, deflection plates, phosphorous-coated screen and an evacuated glass tube that encloses all are the main components of the CRT. The electron beam produced by the electron gun is used to produce a visual image on the screen. The CRT requires high voltages in the order of a few thousand volts for the acceleration of the electron, while a low voltage for the electron gun, which emits the electrons. Supply voltages for other circuits

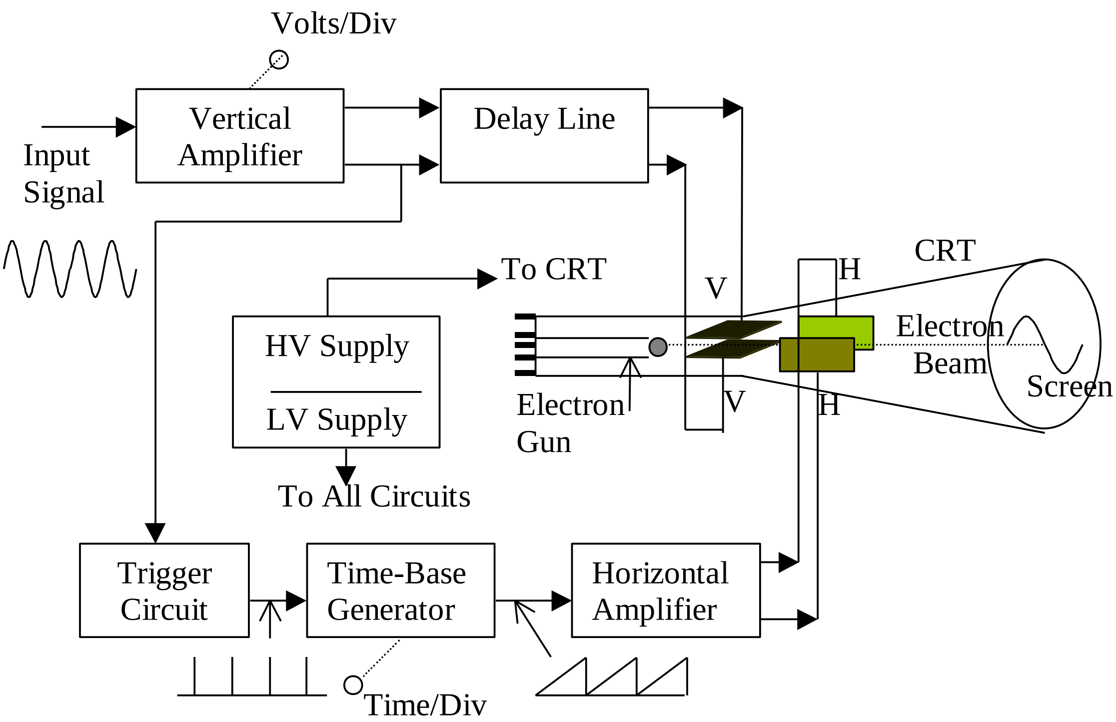


Figure 5.4 Simplified block diagram of an analog oscilloscope

are less than a few hundred volts at maximum. The power supply block provides voltages required by the CRT and the rest of the oscilloscope circuitries.

Two signals are needed to deflect the beam on the screen horizontally and vertically. The laboratory oscilloscope is generally used to display signals in time. The signal to be viewed is applied to a vertical (deflection) amplifier that increases the potential of the input signal to a level that will provide a useful deflection of the electron beam.

The time-base circuitry generates a voltage to supply the CRT to deflect the spot at a constant time-dependant rate. The voltage waveform is named commonly as the sweep signal and it has the appearance of a repetitive ramp function. A triggering circuit is used to synchronize the horizontal deflection with the vertical input, so that the horizontal deflection starts at the same point of the vertical input signal each time it runs (sweeps). Eventually, the beam moves at a constant time-dependant rate horizontally and the image generated on the screen indicates the time variation of the input signal.

Each block in a signal path causes certain time delay. Hence, the beam does not start moving horizontally immediately following the detection of the trigger point. The delay line delays the signal applied to the vertical plates by an amount equal to the time delay for the sweep signal applied to the horizontal deflection plates. Eventually, the vertical signal is displayed on the screen always starting at the trigger point.

BASIC OSCILLOSCOPE OPERATIONS

Electrostatic Deflection

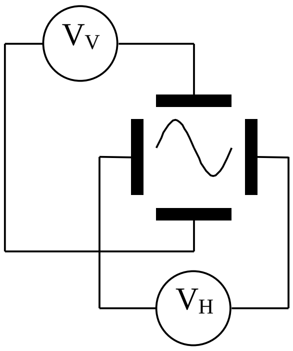


Figure 5.5 Electrostatic deflection

Two pairs of deflection plates at right angles to each other are used to provide deflection of the light spot in a Cartesian system as depicted in Figure 5.5. The amount of voltage that must be applied between a pair of deflection plates to produce a unit length of deflection of the spot depends upon the deflection factor of the CRT. Deflection factors for horizontal and vertical deflection plates are not the same.

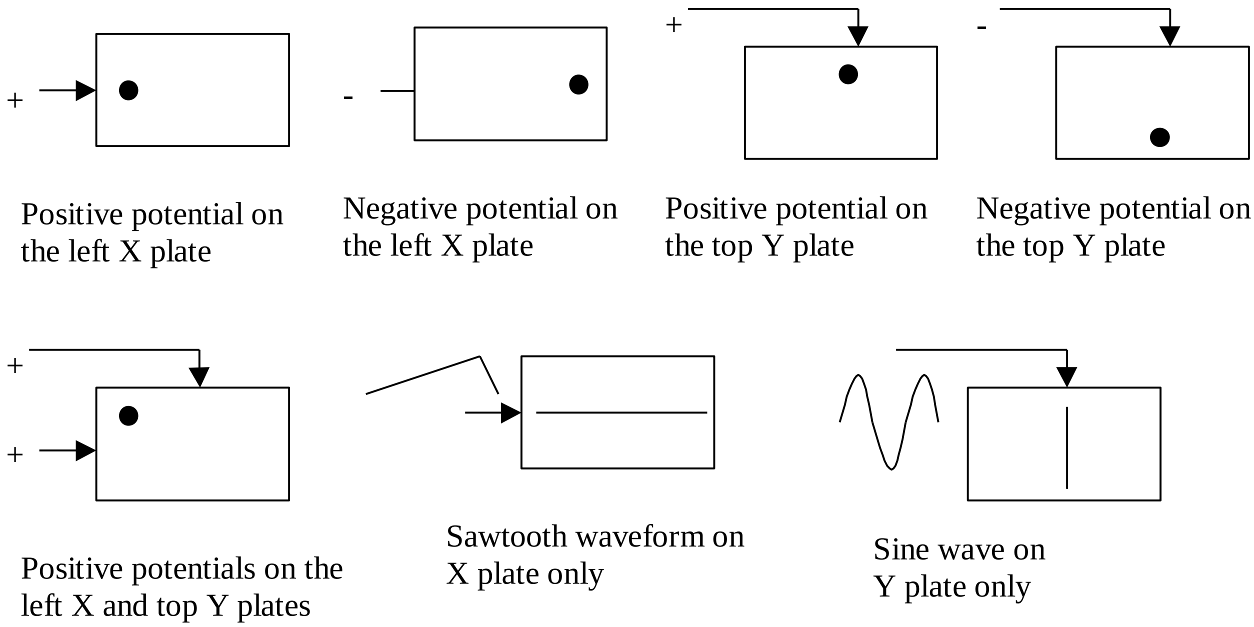


Figure 5.6 Deflection of electron beam on the CRT screen due to several combinations of voltages applied to deflection plates

Various combinations of two voltage waveforms on the screen are illustrated in Figure 5.6. A fixed spot is obtained as DC voltages are applied to both pairs of the plates. A horizontal line is drawn when a sawtooth waveform is applied to the horizontal (X) plates only. Similarly, a vertical line is drawn as a sinusoidal voltage is applied to the vertical (Y) plates only.

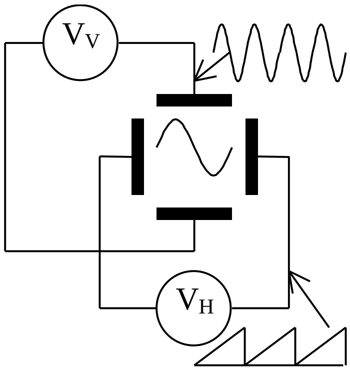


Figure 5.7 Sweep mode of operation

Operation in Sweep Mode

Principle of Operation

The CRO spot traces an image on the screen when horizontal and vertical deflection voltages are applied as shown in Figure 5.7. The voltage applied to horizontal deflection mechanism is the sawtooth that is generated by the time-base circuit. It has a fixed slope and lets the electron beam to travel horizontally at a constant speed. Meanwhile, the input signal (sinusoidal type in the figure) is amplified and applied to the vertical deflection plates.

Figure 5.8 shows a detailed illustration. The timing information for both signals is exposed in the figure. Two cycles of the input signal are displayed on the screen. The second sweep follows the first one immediately indicating that the retrace time is negligible compared to the trace time.

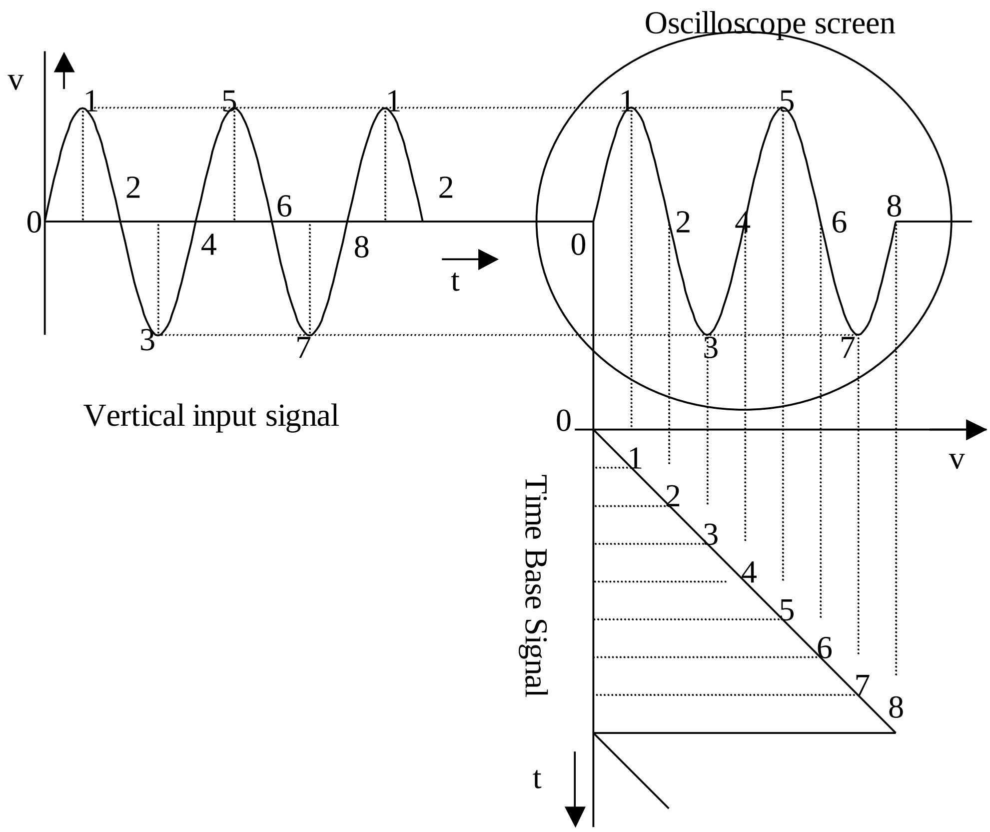


Figure 5.8 Display of signals on the oscilloscope screen in sweep mode

Measurements in Sweep Mode

Amplitude and time variations of the signals can be viewed and measured. In multichannel oscilloscopes, more than one input can be observed simultaneously and compared to each other. Figure 5.9 illustrates two signals V_1 and V_2 displayed together.

The amplitude measurement is made either as reading of the peak value or peak-to-peak value. The time measurement is done to determine the period of a periodic signal and the phase shift between two signals. The displacements in both X and Y directions are taken and multiplied by the scale factors as set at the front panel of the oscilloscope to compute the amplitude in *volt* and time in *second*.

Both measurements require a well-focused trace with gain controls at *cal* (calibrate) positions. Also, the time measurement is possible with the least error if it is done between two steep points on the trace. The steepest point of a sinusoidal signal occurs as the signal crosses the time axis. The following example illustrates basic measurements and their uncertainties.

Example 5.1

For the dual trace shown in Figure 5.9 above, the vertical settings are 0.1 V/cm and 0.2 V/cm for V_1 and V_2 respectively. The time base setting is 5 ms/cm. The trigger source is CH-1 (V_1). Assume uncertainty of ± 0.5 mm in all distances measured. Find:

- Peak and peak to peak values of V_1 and V_2 with uncertainties involved.
- Time period and frequency of V_2 and their uncertainties.
- The trigger level and slope.
- The phase shift between V_1 (CH-1) and V_2 (CH-2). Does V_1 leads or lags V_2? How much is the uncertainty in the phase shift?

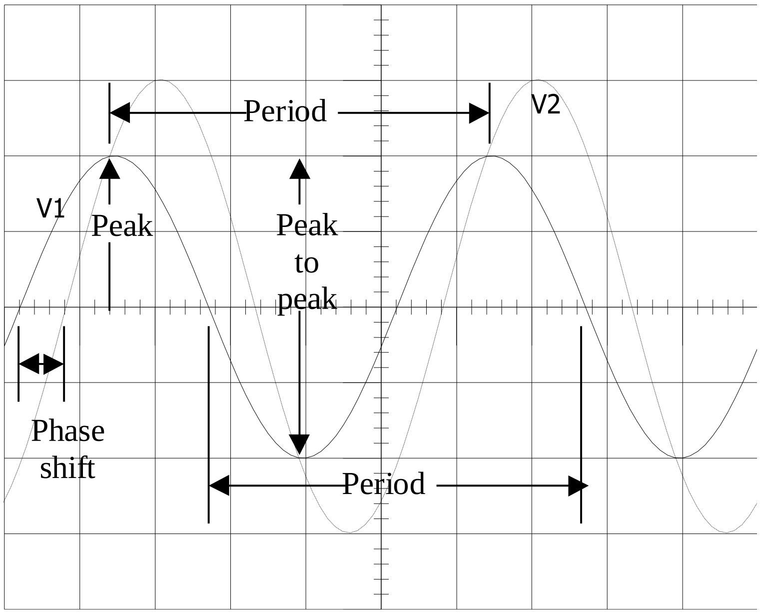


Figure 5.9 Dual channel operation in sweep mode

Solution

Peak value of $V_1 = V_{1p} = 2\text{ (cm)} \times 0.1\text{ (V/cm)} = 0.2\text{ V}$;

peak-to-peak value of V_1 = $V_{1p\text{-}p}$ = 4 (cm) x 0.1 (V/cm) = 0.4 V. Similarly,

V_{2p} = 3 (cm) x 0.2 (V/cm) = 0.6 V; $V_{2p\text{-}p}$ = 6 (cm) x 0.2 (V/cm) = 1.2 V

The uncertainty in distance is 0.5 mm yielding

V_{1p} = (2 ± 0.05) (cm) x 0.1 (V/cm) = 0.2± 0.005 V = 0.2 V± 2.5%

V_{2p} = (3 ± 0.05) (cm) x 0.2 (V/cm) = 0.6± 0.01 V = 0.6 V± 1.67%

$V_{1p\text{-}p}$ = (4 ± 2x0.05) (cm) x 0.1 (V/cm) = 0.4± 0.01 V = 0.4 V± 2.5%

$V_{2p\text{-}p}$ = (6 ± 0.1) (cm) x 0.2 (V/cm) = 1.2± 0.02 V = 1.2 V± 1.67%

Time period and frequency of V_2. T=(5± 0.05)(cm)x5(ms/cm) = 25± 0.25 ms = 25ms± 1%

f = 1/T. $\frac{\partial f}{\partial T} = -\frac{1}{T^2}$ Nominal value of the frequency; f = 40 Hz. Limiting error is the same as the expected accuracy for the frequency. $\Delta f = \left|\frac{\partial f}{\partial T}\Delta T\right| = \frac{\Delta T}{T} f$ yields the relative accuracy for the period and the frequency are the same as 1%. Hence, f = 40± 0.4 Hz = 40 Hz ± 1%

Trigger level = -0.5 cm & (+) slope.

Nominal value of the phase shift is $\theta = \frac{dx360}{T}$ = 0.6(cm)x360°/5(cm) = 43°. Among the two traces, the one that assumes its maximum first is called the leading trace. Hence, V_1 is leading V_2 (also can be said as V_2 is lagging V_1). The uncertainty in the phase: $(\Delta\theta)^2 = \left(\frac{\partial\theta}{\partial d}\right)^2 (\Delta d)^2 + \left(\frac{\partial\theta}{\partial T}\right)^2 (\Delta T)^2$;

$\frac{\partial\theta}{\partial d} = \frac{360}{T} = \frac{\theta}{d}$ and $\frac{\partial\theta}{\partial T} = -\frac{360xd}{T^2} = -\frac{\theta}{T}$ yielding

$\left(\frac{\Delta\theta}{\theta}\right)^2 = \left(\frac{\Delta d}{d}\right)^2 + \left(\frac{\Delta T}{T}\right)^2 = (0.01)^2 + (0.08)^2 = 7.04x10^{-3}$ and eventually Δθ = ± 8.4%, θ = 43°± 3.6° = 43°± 8.4%. The dominant factor in Δθ/θ is Δd/d since it is much larger than ΔT/T

Triggered-Sweep Mode

The oscilloscope is either used in storage mode or in refreshed mode. The sweep signal is applied once only in the storage mode and the traces are stored. Some cathode ray tubes have a special function that stores the trace on the screen and holds it long enough to record the readings or to take a picture. These tubes are rather expensive and the storage function is mostly replaced by a digital storage system that saves the signal in electronic circuits. The storage function is essential especially in studying transient signals that cannot be repeated.

Periodic signals are commonly used during electronic circuit design and test works. Same events are repeated over and over. Generation of the sawtooth waveform in the time-base can be synchronized so that the electron beam follows the same trajectory every time it crosses the screen. This allows utilization of refreshed type CRTs that gives real-time displays of signals. The trace will appear stationary on the screen if the repetition rate is more than 24 times a second.

The trigger circuit is used to obtain the synchronization between the input signal and the sweep signal as discussed in previous section. Its operation is summarized here if that section is skipped. The trigger circuit generates a synchronization (trigger) pulse that initiates the sawtooth waveform. It compares the input signal to a DC signal internally generated. The level of the DC signal can be controlled from the front panel of the oscilloscope. It must be set to a value between the most negative (minimum) and most positive (maximum) values of the input signal. The input signal coincides with the threshold (trigger level) two times during the cycle; first as it goes above the threshold (positive slope) and second time as it goes below the threshold (negative slope). The user can select either one of them using the buttons on the front panel.

Figure 5.10 illustrates the generation of trigger pulses and sawtooth waveforms. The top trace exemplifies the input signal. The threshold (trigger level) signal is shown on the first trace as the dashed line. A negative-slope triggering is used in the example and coincidences are marked. The

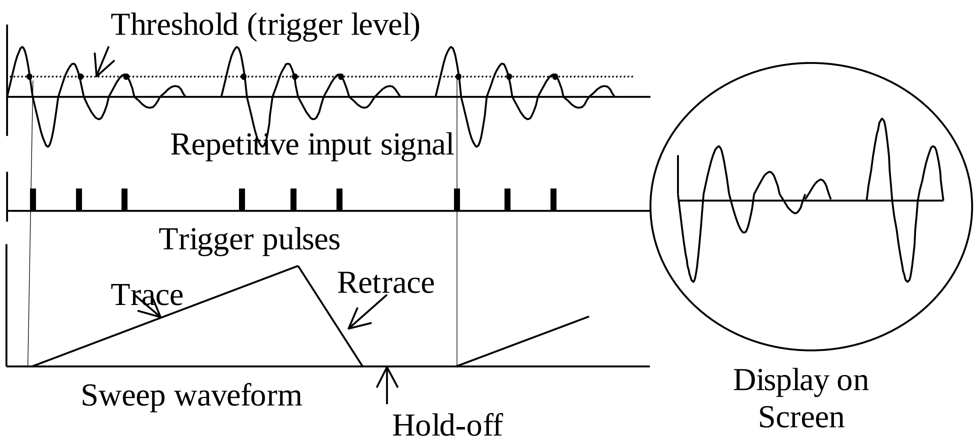


Figure 5.10 The triggered sweep mode of operation

trigger pulses generated are shown as the second trace. The sawtooth waveform is also named as the sweep signal and it is the third trace.

Trigger pulses that occur during the trace and retrace phases of the sweep are ignored. In free-running mode sweeps follow each other. The traces are drawn on the screen over each other and they do not follow the same trajectory unless the frequency of the input signal is a multiple of the frequency of the sweep signal. In triggered sweep mode, the second sweep is not generated until a new trigger pulse is received. Hence, all traces follow the same trajectory yielding a stationary display. The free-running mode is useful in determining the amplitude range of the input signal in case the trigger threshold is set beyond this range.

The sawtooth waveform drops to zero after it reaches the maximum. This drop takes certain amount of time depending upon the time-base circuit used. During this time the electron beam flies back to the left-hand side of the screen and waits (hold-off) there until the start of the next sweep. The electron gun in the CRT is turned-off (blanked) during the retrace and hold-off times to avoid the retrace appearing on the screen and a strong glowing spot on the left side of the screen. The trace, retrace and hold-off intervals are marked on the figure. The resultant oscilloscope display is shown inside the circle.

The input signal to the trigger circuit is the signal applied to the vertical deflection plates in Figure 5.10. In case of multi-trace oscilloscopes, any one of the signals displayed can be used for triggering. Line voltage at 50 Hz / 60 Hz can also be selected as the source of the trigger. This is important in applications involving the component of the 50 Hz / 60 Hz line voltage as interference on other signals. The trigger input can be applied from outside as well. The trigger source is selected using a selector switch on the front panel.

Operation at High Frequencies and Function of the Delay Line

There is an inevitable delay between the application of the input and appearance of the output in all electronic circuit elements. The amount of delay depends upon the element itself and specified by its

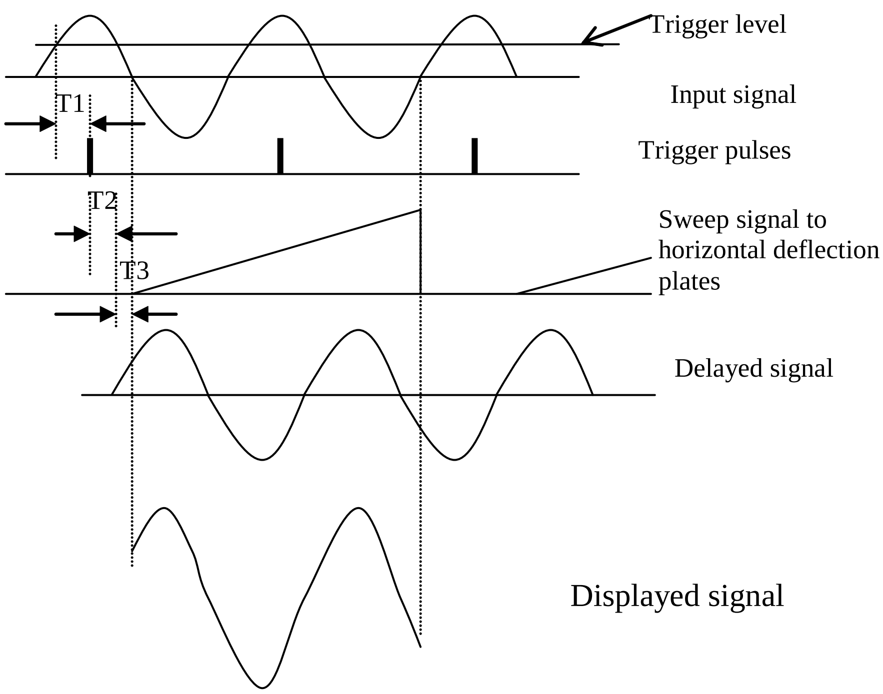


Figure 5.11 Function of the delay line

manufacturer. In the triggered mode of operation, the input signal is applied to the trigger circuit that derives the sawtooth waveform generator. Then, the resultant sawtooth waveform is applied to the horizontal deflection plates via the horizontal amplifier. Hence, there is a time delay (in the order of a few hundred nanosecond) between the coincidence of the input signal with the trigger level signal (trigger pick-off) and starting of the sweep on the display. This delay may not be objectionable at low frequency applications. For example, the period of a 1 kHz sine wave is 1 millisecond. If the delay is 100 nanosecond which is one in a ten thousand of the period. Hence, it will not be effective on displaying the signal. However, if we have the frequency as 10 MHz, the period of the signal is 100 nanosecond, which is the same as the delay. A delay line is added between the vertical amplifier and the vertical deflection plates that will delay the application of the input signal to the deflection plates by the amount of time equal to the delay comes from the time-base circuitry. Figure 5.11 illustrates the effect of the delay line. The delay in trigger circuit is T1, delay in sawtooth generator is T2 and delay in the horizontal amplifier is indicated as T3. The input signal is delayed by the same amount so that the sweeps starts displaying the input signal from the coincidence point.

Operation in X-Y Mode

The oscilloscope can be used to display two signals with respect to each other as illustrated in Figure 5.12. The time-base is switched-off. One of the inputs is applied to vertical while the other one is applied to the horizontal amplifier. This is called the X-Y mode.

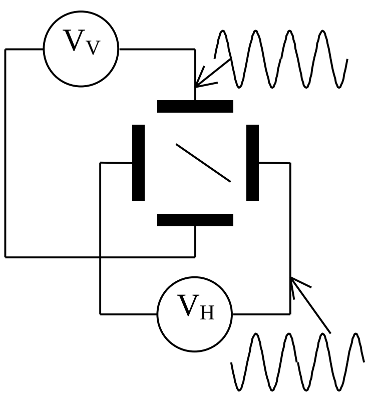
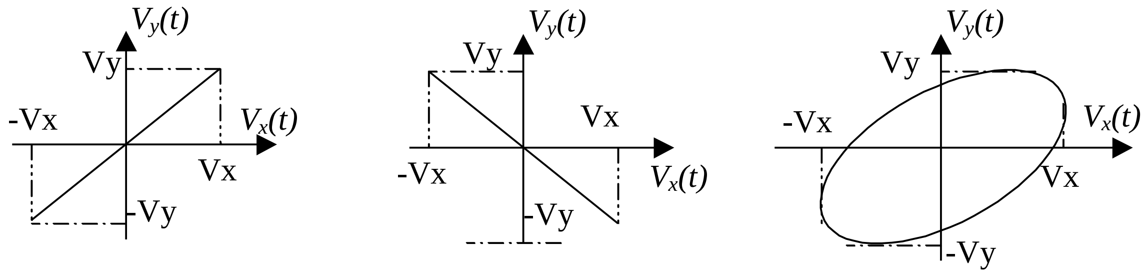


Figure 5.12 X-Y mode

If two sine waves are simultaneously applied, the resulting display in the X-Y mode is called a Lissajous pattern. The magnitudes of signals and the phase shift between them can be determined easily if both have the same frequency. In this case, the pattern is a diagonal straight line, an ellipse or circle as shown in Figure 5.13.

The operation can be studied analytically as follows: Assume that two signals $v_x(t)$ and $v_y(t)$ are applied to horizontal and vertical deflection plates respectively as shown in the figure. Both are sinusoidal signals with magnitudes V_x and V_y for $v_x(t)$ and $v_y(t)$ respectively. The plot on the screen can be expressed by the parametric equation

$$v_x(t) = V_x \sin(\omega t) \text{ and } v_y(t) = V_y \sin(\omega t)$$



$v_x(t) = V_x \sin(\omega t)$ $v_x(t) = V_x \sin(\omega t)$ $v_x(t) = V_x \sin(\omega t)$

$v_y(t) = V_y \sin(\omega t)$ $v_{y(t)} = -V_y \sin(\omega t)$ $v_y(t) = V_y \sin(\omega t+\theta)$

Figure 5.13 Displays for two sinusoidal signals of the same frequency in X-Y mode

for the first plot. This represents a straight line in the X-Y plane that can be written as

$$y = (V_y/V_x)x$$

The slope of the line is V_y/V_x . The middle plot has a negative slope due to the negative sign in the definition of $v_y(t)$. There is a phase shift between the two signals in the third case. The plot is an ellipse. If the phase shift is 90° and magnitudes are identical, then the ellipse is converted to a circle with radius $V_x = V_y$.

The magnitudes of signals can be determined from the peak values of the ellipse as shown in the previous figure. The phase shift between them is found using the magnitudes and zero crossing for $v_y(t)$. At t = 0, $v_y(t) = V_y \sin\theta$ and $v_x(t) = 0$. Hence,

$$\sin\theta = y_{intercept}/y_{max}$$

Both negative and positive angles lead to the same plot on the screen. Thus, it is not possible to tell which one of the signal is leading. Following examples illustrate the utilization of the X-Y mode.

Example 5.2 - Sketch the scope waveforms

In sweep mode for $v_1(t) = 1 \sin(4000\pi t)$, $v_2(t) = 2 \sin(4000\pi t + 45°)$ with vertical settings 0.5 V/cm for both channels, time base setting 0.1 ms/cm, screen height 8 cm, screen width 10 cm, trigger source channel-1, trigger level 0 V, and slope positive.

Assume the scope is switched to X-Y mode, $v_1(t)$ is applied to vertical (Y) and $v_2(t)$ to horizontal (X) amplifiers. The settings for are 0.5 V/cm for both inputs.

The waveforms for the sweep mode of operation are shown in Figure 5.14 at the left. The display for the X-Y mode of operation for the same signals is shown at the right in Figure 5.14.

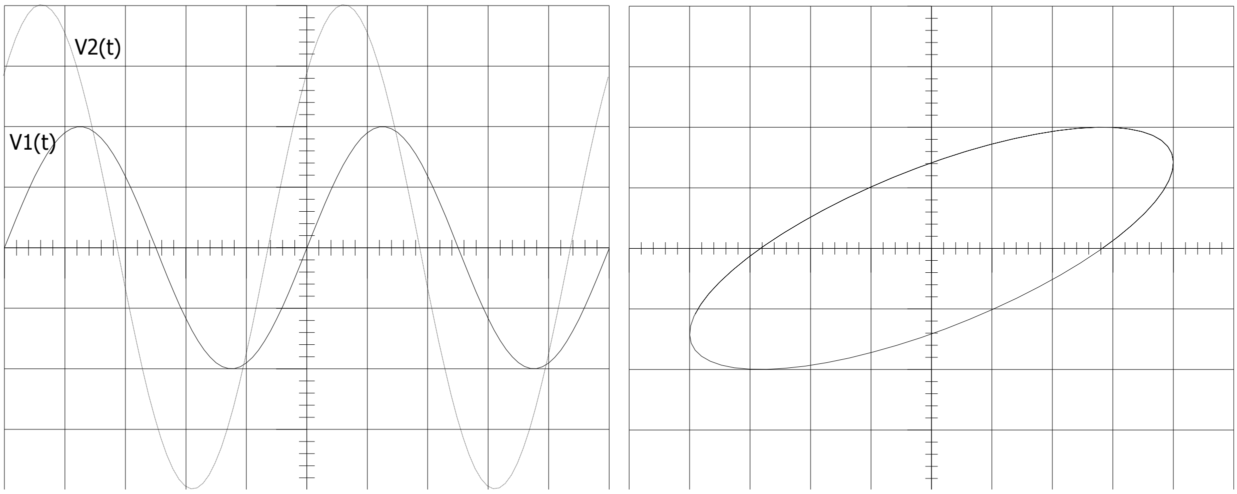


Figure 5.14 Display of signals in example 5.2 in sweep and X-Y modes

Example 5.3

The time-base is switched off and the oscilloscope is switched to XY mode of operation. V_1 is connected to the X input with sensitivity 0.1 V/cm, and V_2 is connected to the Y input with sensitivity 0.2 V/cm. The resulting ellipse is shown in Figure 5.15 with marking of distances Y_m, Y_0, X_m, and X_0.

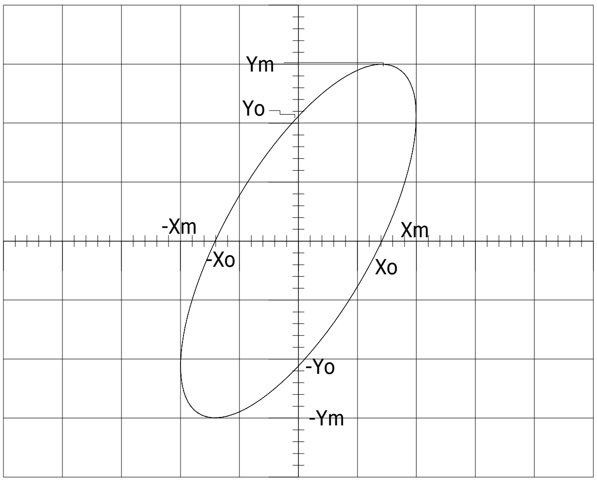


Figure 5.15 Measurements in X-Y mode

Y_m = 3 cm, Y_0 = 2.1 cm, X_m = 2 cm, and X_0 = 1.4 cm

Similarly, $-Y_m$ = -3 cm, $-Y_0$ = -2.1 cm, $-X_m$ = -2 cm, and $-X_0$ = -1.4 cm

Phase shift between V_1 and V_2 : $\sin\theta = 2.1/3 = 0.7$ yielding $\theta = 44°$

Example 5.4

The oscilloscope is switched to XY mode of operation. V_1 is connected to the X input with sensitivity 10 mV/cm, and V_2 is connected to the Y input with sensitivity 0.5 V/cm. The resulting ellipse is shown in Figure 5.16. Calculate

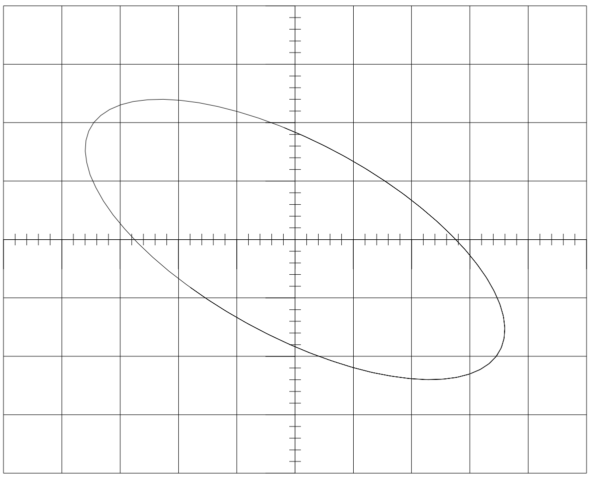


Figure 5.16 Display for example 5.4

Distances Y_m, Y_0, X_m, and X_0.

Phase shift between signal in X and signal in Y.

Peak-to-peak values of voltage for V_1 and V_2

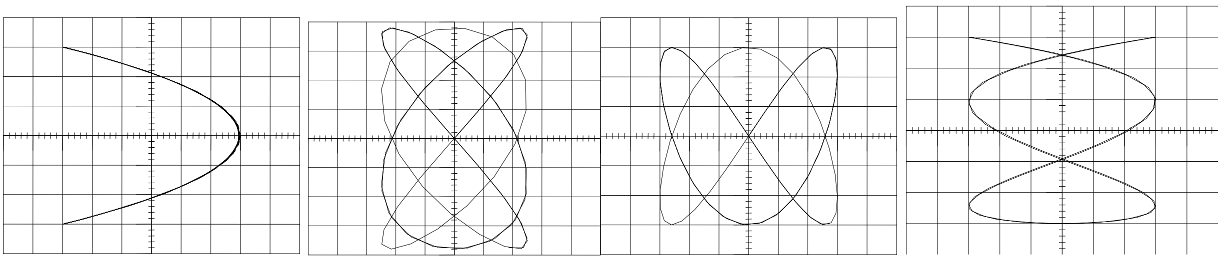
Distances Y_m, Y_0, X_m, and X_0 are 2.4 cm, 1.8 cm, 3.6 cm and 2.75cm respectively

Phase shift between signal in X and signal in Y is $\theta = 180° - \sin^{-1}(Y_0/Y_m) = 131°$

$V_{1p\text{-}p}$= $2xX_m x10$ mV/cm = 72 mV p-p; $V_{2p\text{-}p}$= 2.4 V p-p

If the two signals used in the X-Y mode have frequencies that are not identical, then the resulting Lissajous patterns are not straight lines or ellipses any more. The pattern will be stable if the frequency ratio can be expressed in a small whole number or a simple fraction. This is used in setting the frequency of an unknown oscillator using a standard oscillator. The frequency of the unknown is varied until a stable trace is obtained. Then, the ratio of frequencies can be computed easily from the horizontal and vertical tangency as

$$\frac{f_x}{f_y} = \frac{\#of\ vertical \tan gency}{\#of\ horizontal \tan gency}$$



$f_x = 2f_y$ $3f_x = 4f_y$ $3f_x = 2f_y$ $2f_x = 5f_y$

Figure 5.17 Examples of Lissajous figures

Figure 5.17 shows four examples. The shape of the plot changes with the phase shift although the ratio of frequencies is fixed. If the contact with the tangent is from one direction, then that contact is counted as a half tangent. If the contact is from two directions, this is counted as a full tangent. In the first plot at the left, the horizontal tangent has ½ tangency, while the vertical tangent has 2x1/2 if it is taken at the left and one full tangent if it taken at the right. Eventually, the ratio of horizontal to vertical frequencies is 2. In other plots the ratios are found in a similar manner as 4/3, 2/3 and 5/2.

MULTI-TRACE OSCILLOSCOPES

Most laboratory oscilloscopes display two more traces simultaneously although they have a single electron gun. Each trace can represent an independent input signal. There are identical input connector, attenuator and amplifier for each input. Outputs of vertical amplifiers are selected one-by-one by an electronic switch and applied to the driver amplifier for the vertical deflection plate assembly as illustrated in Figure 5.18. There are two modes of operation of the electronic switch as chopped and alternate. In the chopped mode, the switch runs at high frequency (around 500 kHz) and calls at each input for a fraction of the total sweep duration. Hence, traces are drawn as short

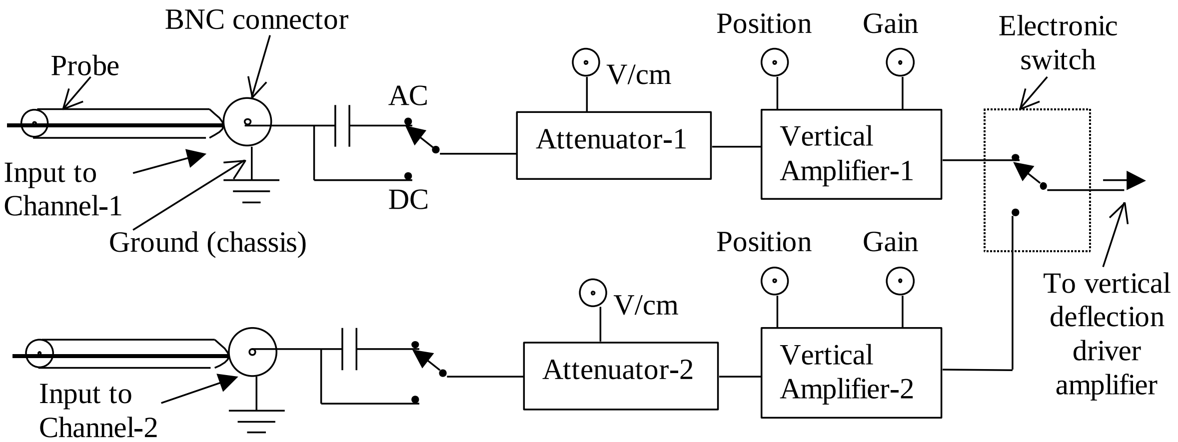


Figure 5.18 Multi-trace operation using an electronic switch

spots of light on the screen. For example, if we have two input signals each at 1 kHz and the sweep rate is 500 kHz, then there are 250 spots across one period of each trace. The illumination of the spot covers the gap between the spots. Also, the chopping is not synchronous with the sweep leading to the dots appearing at different places along the trajectory for successive sweeps. Hence, the traces are seen continuous at low frequency applications. Therefore, the chopped mode is useful at low frequencies.

In the second operational mode, the switch remains in one of the channel throughout the complete sweep duration and it picks the other one in the next sweep. Since switch displays each channel at alternate cycles of the sweep signal, the name alternate mode is used. This is useful at high frequency operations. Some laboratory oscilloscopes incorporate the selection of chopped or alternate mode in the time-base switch. Only one of the input channels is used for the trigger control in both modes. In the alternate mode if channel-1 is selected as the trigger input, it is used even while channel-2 is displayed.

DIGITAL STORAGE OSCILLOSCOPES (DSO)

Oscilloscopes also come in analog and digital types. An analog oscilloscope works by directly applying a voltage being measured to an electron beam moving across the oscilloscope screen. The voltage deflects the beam up and down proportionally, tracing the waveform on the screen. This gives an immediate picture of the waveform as described in previous sections. In contrast, a digital oscilloscope samples the waveform and uses an analog-to-digital converter (or ADC) to convert the voltage being measured into digital information. It then uses this digital information to reconstruct the waveform on the screen (Figure 5.19).

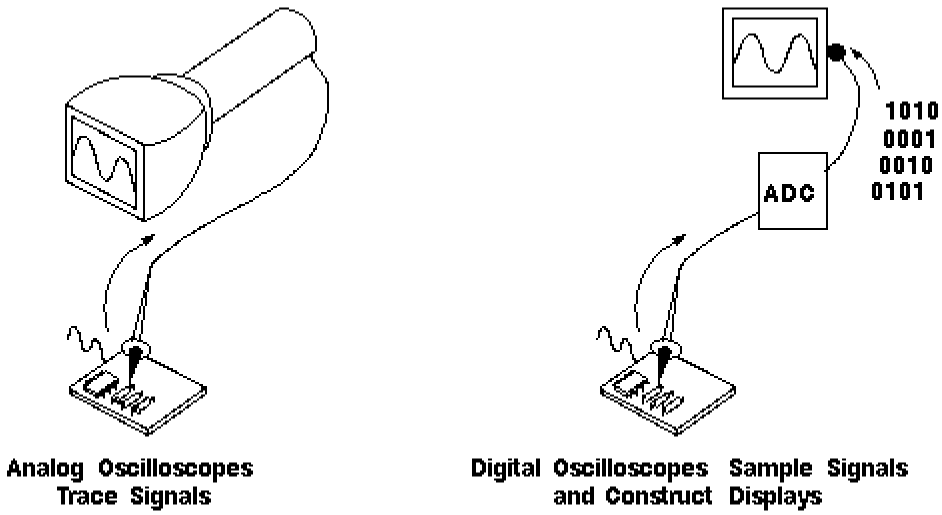


Figure 5.19 Digital and Analog Oscilloscopes Display Waveforms

For many applications either an analog or digital oscilloscope will do. However, each type does possess some unique characteristics making it more or less suitable for specific tasks. People often prefer analog oscilloscopes when it is important to display rapidly varying signals in "real time" (or as they occur). Digital oscilloscopes allow us to capture and view events that may happen only once. They can process the digital waveform data or send the data to a computer for processing. Also, they can store the digital waveform data for later viewing and printing.

Necessity for DSO and Its Advantages

If an object passes in front of our eyes more than about 24 times a second over the same trajectory, we cannot follow the trace of the object and we will see the trajectory as a continuous line of action. Hence, the trajectory is stored in our physiological system. This principle is used in obtaining a stationary trace needed to study waveforms in conventional oscilloscopes. This is however, is not possible for slowly varying signals and transients that occur once and then disappear. Storage oscilloscopes have been developed for this purpose.

Digital storage oscilloscopes came to existence in 1971 and developed a lot since then. They provide a superior method of trace storage. The waveform to be stored is digitized, stored in a digital memory, and retrieved for displayed on the storage oscilloscope. The stored waveform is continuously displayed by repeatedly scanning the stored waveform. The digitized waveform can be further analyzed by either the oscilloscope or by loading the content of the memory into a computer. They can present waveforms before, during and after trigger. They provide markers, called the cursors, to help the user in measurements in annotation (detailing) of the measured values.

Principles of Operation

Principle Diagram Representing Operation of the DSO

A simplified block diagram of a digital storage oscilloscope is shown in Figure 5.20. The input circuitry of the DSO and probes used for the measurement are the same as the conventional oscilloscopes. The input is attenuated and amplified with the input amplifiers as in any oscilloscope. This is done to scale the input signal so that the dynamic range of the A/D converter can be utilized maximally. Many DSOs can also operate in a conventional mode, bypassing the digitizing and storing features. The output of the input amplifier drives the trigger circuit that provides signal to the control logic. It is also sampled under the control of the control logic. The sample and hold circuit takes the sample and stores it as a charge on a capacitor. Hence, the value of the signal is kept constant during the analog to digital conversion. The analog to digital converter (A/D) generates a binary code related to the magnitude of the sampled signal. The speed of the A/D converter is important and “flash” converters

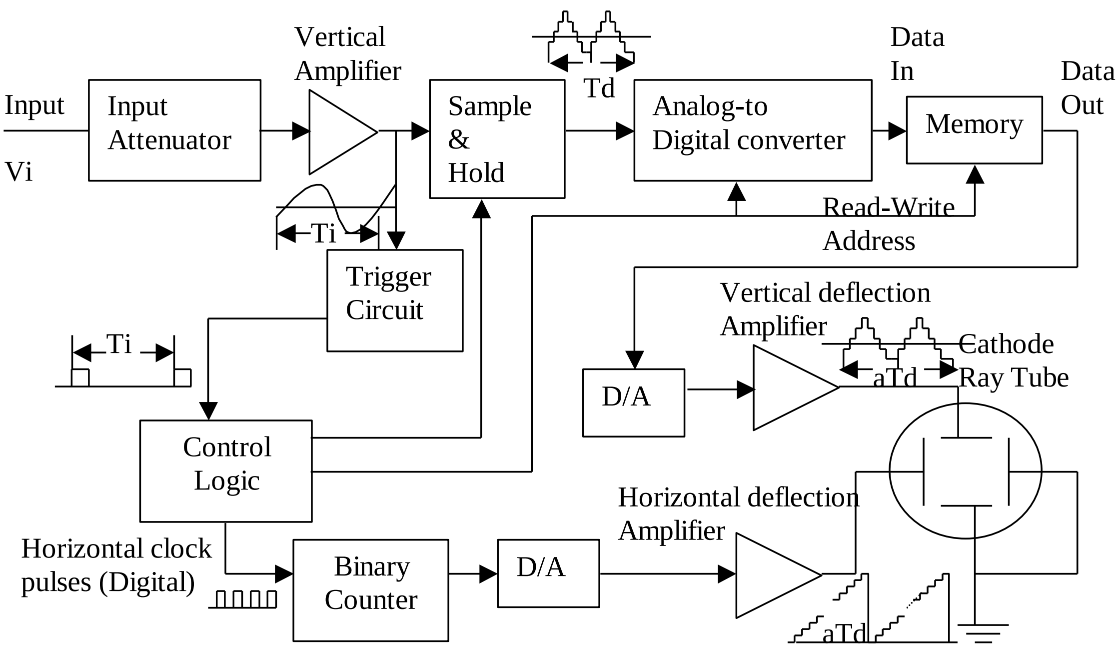


Figure 5.20 Simplified block diagram of a digital storage oscilloscope (DSO)

are mostly used. The binary code from the A/D converter is stored in the memory. The memory consists of a bank of random access memory (RAM) integrated circuits (ICs).

The Time-Base Circuit

The control logic generates a clock signal applied to the binary counter. The counter accumulates pulses and produces a binary output code that delivered to a digital to analog (D/A) converter to generate the ramp signal applied to the horizontal deflection amplifier. The horizontal deflection plates are supplied with this ramp signal to let the electron to travel across the screen horizontally at a constant speed. The speed of the transition of electron depends upon the slope of the ramp that is controlled by the clock rate. The capacity of the counter is taken to have the maximum number accumulated corresponding to the rightmost position on the screen. With the next clock pulse, the binary output of the counter drops to all zeros yielding the termination of the ramp.

The Displayed Signal

Meanwhile, the data currently in the store is read out sequentially and the samples pass to the second D/A converter. There they are reconstructed into a series of discrete voltage levels forming a stepwise approximation of the original waveform. This is fed to the vertical deflection plates via the vertical deflection amplifier. For a multi-trace oscilloscope, each channel has the same circuitry and outputs of the D/A converters are combined in the vertical deflection amplifier.

The delay line used in conventional oscilloscopes for synchronization is not needed in digital storage oscilloscopes since this function can be easily handled by the control logic. The read out and display of samples constituting the stored waveform need not occur at the same sample rate that was used to acquire the waveform in the first place. It is sufficient to use a display sample rate adequate to ensure that each and every trace displayed is rewritten fifty or more times a second to prevent the flicker of the display. Eventually, the time interval of the signal on the display is not *Td* of the input signal. Assume that we have a sampling rate of 1000 samples per second and we use 1000 samples for the display. The time referred to the input signal is *Td* = 1 second and it takes 1 second for the DSO to store the information into the memory. Writing to the memory and reading from the memory are independent activities. Once the information is stored, it can be read at any rate. Assume the memory is scanned using a clock signal of 50 kHz. Then, it takes (1/50) second to scan 1000 memory cells and *aTd* which is the duration of the signal that actually appears on the screen becomes 20 millisecond.

Current Trends

The DSOs can work at low sweep rates allowing utilization of cheaper CRTs with wider screen and deflection yoke (coils that provide magnetic field instead of electrical field produced by the deflection plates). In some current DSOs, even liquid crystal displays (LCDs) are used with television like scanning techniques. This allows the development of hand-held and battery operated instruments. Some of these techniques will be dealt with in the section for display technologies.

VIRTUAL INSTRUMENTATION

Definition

A virtual instrumentation system is computer software that a user would employ to develop a computerized test and measurement system, for controlling from a computer desktop an external measurement hardware device, and for displaying test or measurement data collected by the external device on instrument-like panels on a computer screen as illustrated in Figure 5.21. The virtual instrument is a system that uses customizable software and modular measurement hardware to create user-defined measurement systems as opposed to traditional hardware instrumentation systems such as digital multimeters and oscilloscopes that are made up of pre-defined hardware components.

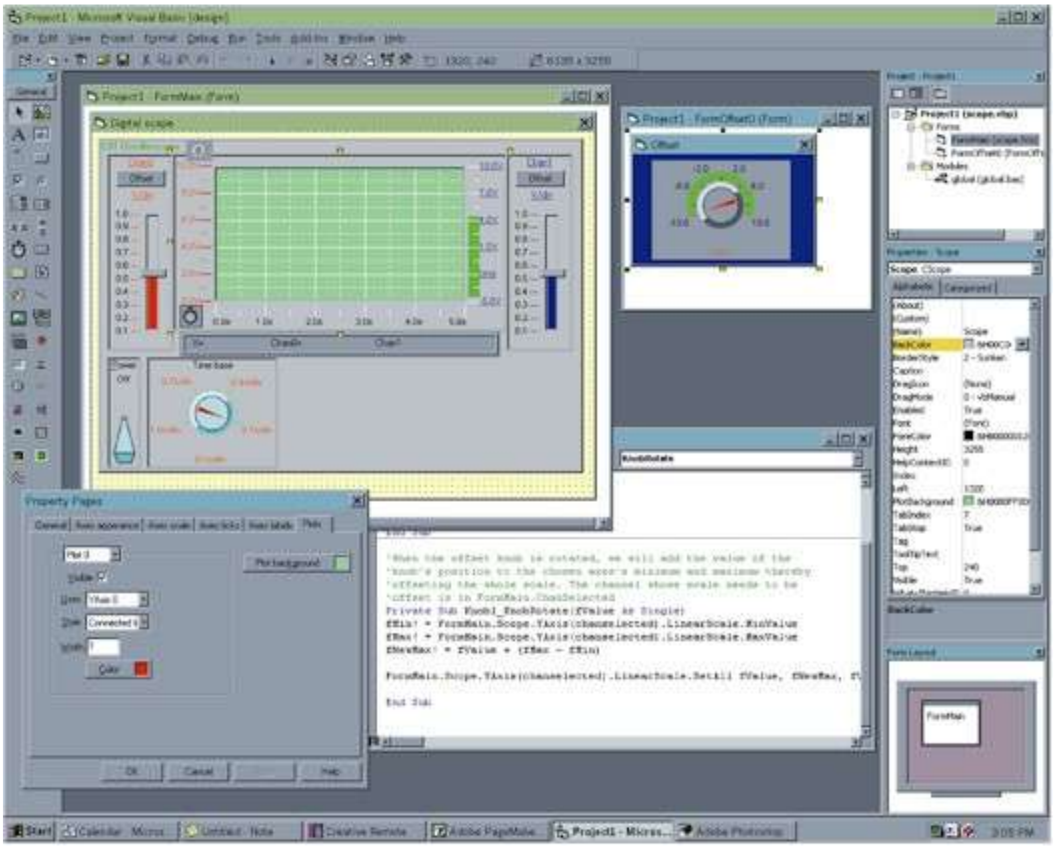


Figure 5.21 A display panel for a virtual instrumentation system

The traditional systems are completely specific to their stimulus, analysis, or measurement function and because of their hard-coded function, these systems are more limited in their versatility than virtual instrumentation systems. Hence, the primary difference between hardware instrumentation and virtual instrumentation is that software is used to replace a large amount of hardware.

The software enables complex and expensive hardware to be replaced by already purchased computer hardware. Virtual instrumentation extends also to computerized systems for controlling processes based on data collected and processed by a computerized instrumentation system. The vision of virtual instrumentation revolutionized the way engineers and scientists work, delivering solutions with faster development time, lower costs, and greater flexibility.

Components of Virtual Instrumentation

Virtual instrumentation thus refers to the use of general purpose computers and workstations, in combination with data collection hardware devices, and virtual instrumentation software, to construct an integrated instrumentation system; in such a system the data collection hardware devices, which incorporate sensing elements for detecting changes in the conditions of test subjects, are intimately coupled to the computer, whereby the operations of the sensors are controlled by the computer software, and the output of the data collection devices is displayed on the computer screen, in a manner designed in software to be particularly useful to the user, for example by the use of displays simulating in appearance the physical dials, meters and other data visualization devices of traditional instruments.

Virtual instrumentation is combination of a productive software, modular input/output (I/O), and scalable platform as shown in Figure 5.22. The heart of any virtual instrument is the flexible software that allows an innovative engineer or scientist to develop a user-defined instrument specific to the application needs. With such software, engineers and scientists can interface with real-world signals; analyze data for meaningful information, and share results and applications.

Figure 5. 22 Virtual instrumentation combines productive software, modular I/O, and scalable platforms

The second virtual instrumentation component is the modular I/O for measurements that require higher performance, resolution, or speeds. In combination with powerful software, engineers can create custom-defined measurements and sophisticated analysis routines.

The third virtual instrumentation element is - popular and commercially available computing platform (PC or Server) to run the software and connect to I/O module, often enhanced with accurate synchronization - ensures that virtual instrumentation takes advantage of the very latest computer capabilities and data transfer technologies. This element delivers virtual instrumentation on a long-term technology base that scales with the high investments made in processors, buses, and more. Together, these components empower engineers and scientists world over to create their own solutions with virtual instrumentation.

Virtual Instrumentation for Design

The same design engineers that use a wide variety of software design tools must use hardware to test prototypes as illustrated in Figure 5.23. Commonly, there is no good interface between the design phase and testing/validation phase, which means that, often the issues discovered in the testing phase require a design-phase reiteration.

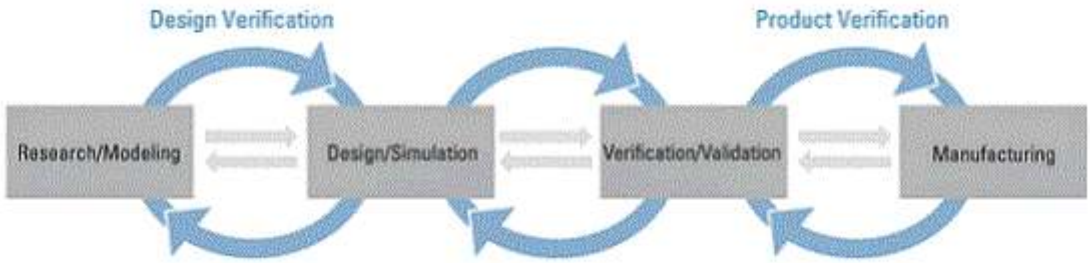


Figure 5.23 Test plays a critical role in the design and manufacture of today's electronic devices

In reality, the development process has two very distinct and separate stages – design and test are two individual entities as illustrated in Figures 5.24 and 5. 25 respectively. On the design side, EDA tool vendors undergo tremendous pressure to interoperate from the increasing semiconductor design and manufacturing group complexity requirements. Engineers and scientists are demanding the capability to reuse designs from one tool in other tools as products go from schematic design to simulation to physical layout. Similarly, test system development is evolving toward a modular approach. The gap between these two worlds has traditionally been neglected, first noticeable in the new product prototype stage.

Systems with intrinsic-integration properties are easily extensible and adapt to increasing product functionality. When new tests are required, engineers simply add new modules to the platform to make the measurements. Virtual instrumentation software flexibility and virtual instrumentation hardware modularity make virtual instruments a necessity to accelerate the development cycle.

Virtual instrumentation has gradually increased addressable applications through continuous software innovation and hundreds of measurement hardware devices. Having influenced millions of test and automation professionals, today it is winning over experts in the control and design domains. Virtual Instrumentation is rapidly revolutionizing the functions of control design, distributed control, data logging, design verification, prototyping, simulation and more.

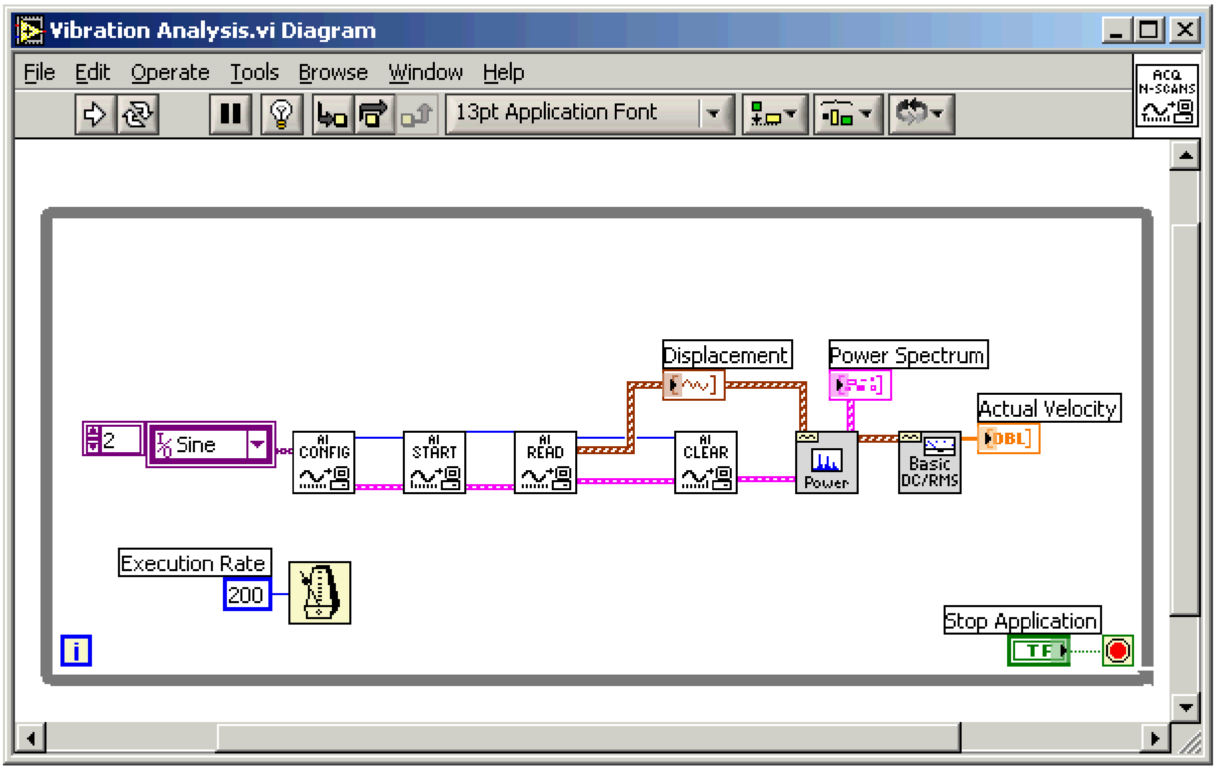


Figure 5.24 An example design screen for the virtual instrumentation in LabView (National Instruments)

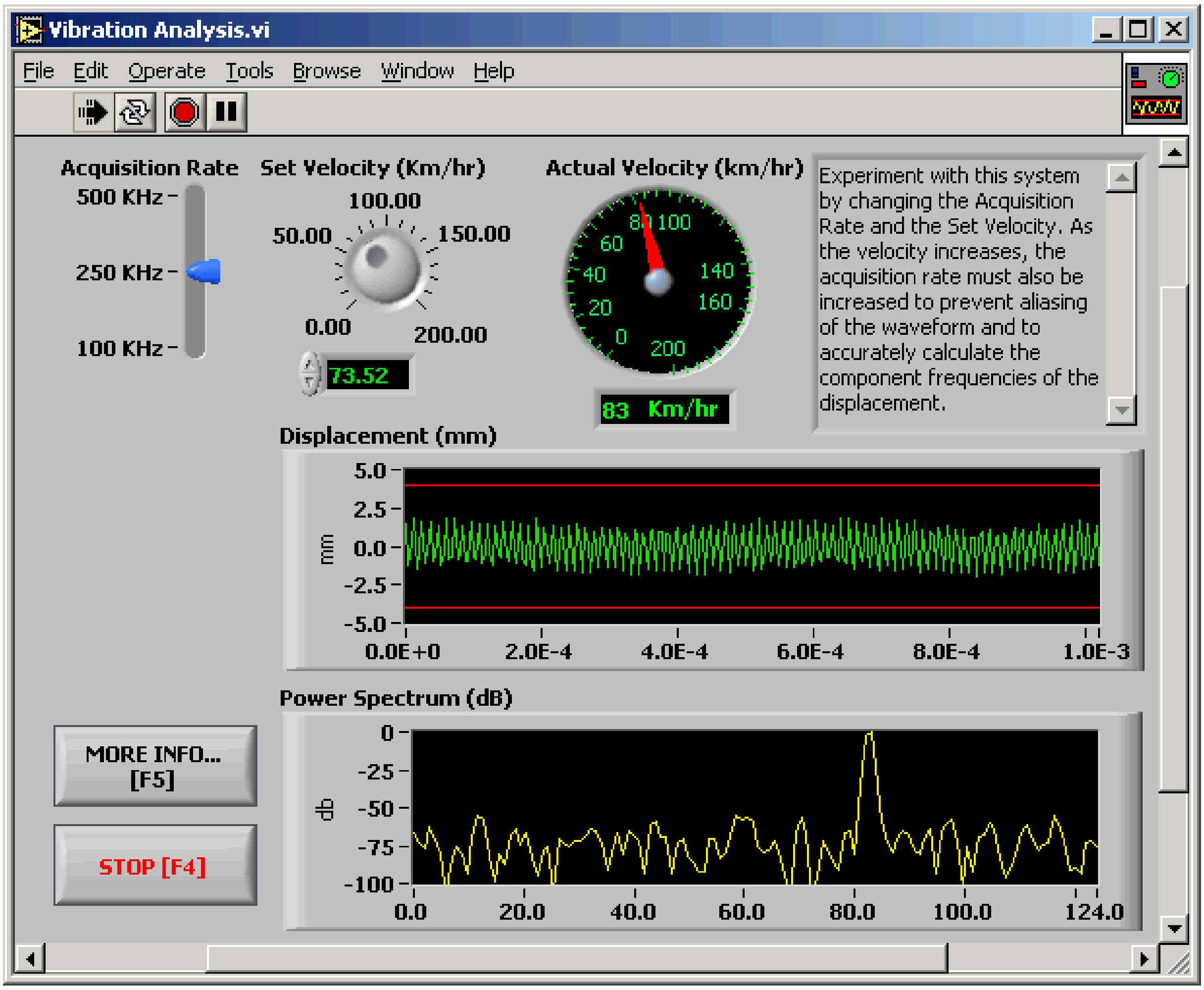


Figure 5.25 An example of a test and analysis screen for virtual instrumentation in LabView (National Instruments)

http://www.eeherald.com/section/design-guide/dgni100003.html

http://www.datatranslation.com/docs/whitepapers/Evolution-of-Virtual-Instrumentation.pdf

PICTURE DISPLAY

Generation and Presentation of Picture

A visual display is the most-used output device of computers, entertainment instruments and scientific equipment. It is often referred to as a monitor when packaged in a separate case. It provides instant feedback by showing the text and graphic images as we work or play. Light is the energy that carries the information to our eyes. It is either internally generated or supplied via an external source. This section will introduce principle of operation of major displays and details will be given in Appendix-C.

Moving Scene from Still Pictures

If we divide a still image into a collection of small colored dots, our brain will reassemble the dots into a meaningful image. Both televisions and computer screens (as well as newspaper and magazine photos) rely on this fusion-of-small-colored-dots capability in the human brain to chop pictures up into thousands of individual elements. On a TV or computer screen, the dots are called **pixels** as shown in Figure 5.26. The resolution of our computer's screen might be 800x600 pixels, or maybe 1024x768 pixels.

Figure 5.26

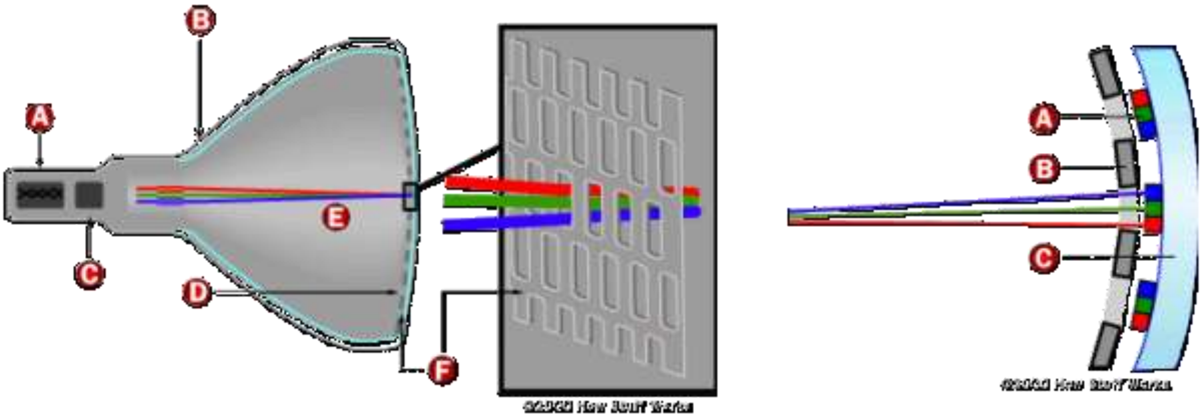
If we divide a moving scene into a **sequence of still pictures** and show the still images in **rapid succession**, then the brain will **reassemble** the still images into a **single, moving scene**. By putting together 15 or more subtly different frames per second, the brain integrates them into a moving scene. Fifteen per second is about the minimum possible -- any fewer than that and it looks jerky.

Display Technologies

Often referred to as a **monitor** when packaged in a separate case, the display is the most-used output device on a computer. The display provides instant feedback by showing you text and graphic images as you work or play. Most desktop displays use liquid crystal display (LCD) or cathode ray tube (CRT) technology, while nearly all portable computing devices such as laptops incorporate LCD technology. Because of their slimmer design and lower energy consumption, monitors using LCD technology (also called **flat panel** or **flat screen** displays) are replacing the venerable CRT on most desktops. There are emerging display technologies as well in addition to the classical CRT and LCD displays. Important ones among them are the plasma displays, Organic Light-Emitting Diode (OLED) Surface-Conduction Electron Emitter Displays (SED) and Field Emission Displays (FED).

The Cathode Ray Tube (CRT)

The CRT for a picture display is very similar to that is found in an oscilloscope as shown in Figure 5.27. The major difference is that it has three cathodes, a shadow mask and the screen for three colors. It is composed of dots and the distance between neighboring dots is called the dot pitch. The beams are rooted on the phosphors for individual colors using a special guiding technique that contains either an aperture grill or shadow mask as illustrated in Figure 5.28.



The CRT Monitor - A: Cathode; B: Conductive coating; C: Anode; D: Phosphor coated screen; E: Electron beam; F: Shadow mask

The CRT screen – A: Phosphors, B: Shadow mask and C: The glass

Figure 5.27 The CRT type display

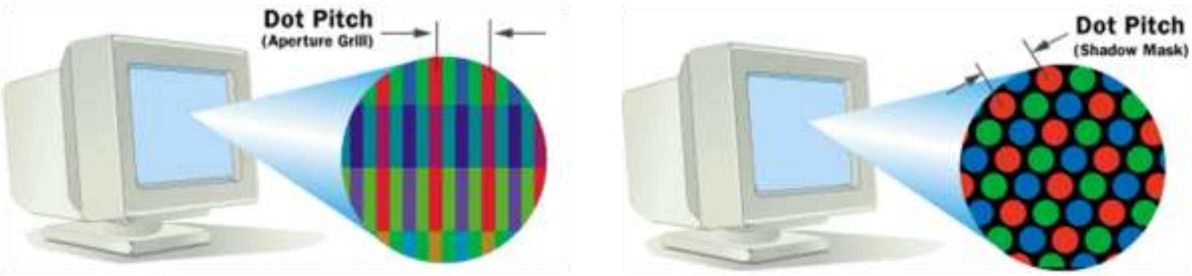


Figure 5.28 Pixels and dot pitch in a color CRT monitor

The CRT technology is a classical and well established one. It is still advantages as compared to other emerging technologies in terms of price, color representation, responsiveness to fast changes and ruggedness.

Color Depth

The combination of the display modes supported by your graphics adapter and the color capability of your monitor determine how many colors it displays. For example, a display that operates in SuperVGA (SVGA) mode can display up to 16,777,216 (usually rounded to 16.8 million) colors because it can process a 24-bit-long description of a pixel. The number of bits used to describe a pixel is known as its bit depth. With a 24-bit bit depth, eight bits are dedicated to each of the three additive primary colors -- red, green and blue. **Color bit depth** refers to the number of bits used to describe the color of a single pixel. The bit depth determines the number of colors that can be displayed at one time.

Standards and Resolution

Resolution refers to the number of individual dots of color, known as **pixels**, contained on a display. It is expressed by identifying the number of pixels on the **horizontal axis** (rows) and the number on the **vertical axis** (columns), such as 800x600. The resolution is affected by a number of factors, including the size of the screen.

Steering Coils (Deflection Yoke)

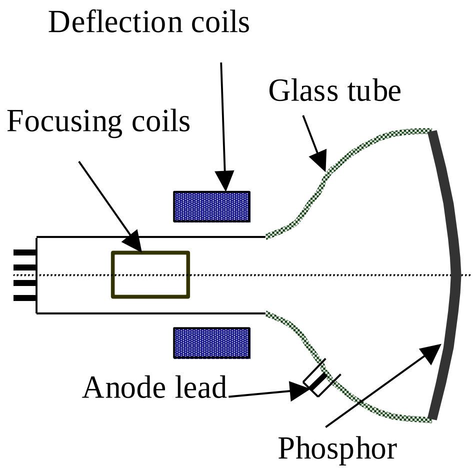


Figure 5.29 Magnetic deflection

Electron beam can be deflected from its path if it is subjected to a magnetic field as well. In this case, the force acting on the electron is perpendicular to both the direction of electron flow and the magnetic field itself. Two sets of coils are placed perpendicular to each other over the neck of the CRT outside the glass envelope as shown in Figure 5.29. The current in these coils provide the two magnetic fields in X and Y directions. As the electron comes in Z direction, it is deflected in Y and X directions respectively. The mechanism of coils is called the *deflection yoke*.

The neck of the CRT is considerably shorter and thinner than the case of electrostatic deflection. There is also no geometric limitation on the deflection angle resulting in larger display area. There are two basic limitations in application of the electromagnetic deflection. Firstly, the inductance and distributed capacitance of the coil require higher voltages to be applied for a given current as the frequency of the deflection current increases. Practical tubes are limited to frequencies up to 20 - 25 kHz. The minimum deflection frequency in the cheapest laboratory oscilloscope is 20 MHz. Eventually, almost all high frequency laboratory oscilloscopes use electrostatic deflection mechanisms. The second limitation comes from the increased screen size. The trajectory of the spot covers varying lengths as it travels along the screen. This requires a more complicated focusing circuitry. The magnetic deflection is used in television and computer displays and most of the digital storage oscilloscopes that have CRT screens.

Liquid Crystals

The Basics of LCD

Liquid crystal display technology works by blocking light. Specifically, an LCD is made of two pieces of polarized glass (also called substrate) that contain a liquid crystal material between them. A backlight produces light that passes through the first substrate. At the same time, electrical currents cause the liquid crystal molecules to align to allow varying levels of light to pass through to the second substrate and generate the colors and images that we see.

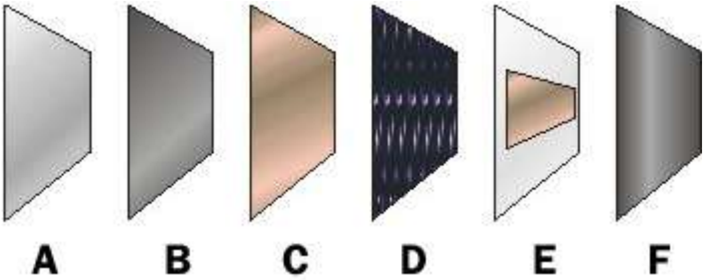


Figure 5.30 Layer of the liquid crystal display

The LCD needed to do this job is very basic and it has six layers as illustrated in Figure 5.30.

- It has a mirror (**A**) in back, which makes it reflective.
- Then, we add a piece of glass (**B**) with a polarizing film on the bottom side,
- And a common electrode plane (**C**) made of indium-tin oxide on top. A common electrode plane covers the entire area of the LCD.
- Above that is the layer of liquid crystal substance (**D**).
- Next comes another piece of glass (**E**) with an electrode in the shape of the rectangle on the bottom and,
- On top, another polarizing film (**F**), at a right angle to the first one.

The electrode is hooked up to a power source like a battery. When there is no current, light entering through the front of the LCD will simply hit the mirror and bounce right back out. But when the battery supplies current to the electrodes, the liquid crystals between the common-plane electrode and the electrode shaped like a rectangle untwist and block the light in that region from passing through. That makes the LCD show the rectangle as a black area.

The LCD in a calculator display requires an **external light source**. Liquid crystal materials emit no light of their own. Rather, small electrodes charge the liquid crystals and make the layers untwist so that light is not transmitting through the polarized film. Small and inexpensive LCDs are often **reflective**, which means to display anything they must reflect light from external light sources. Most computer displays are lit with built-in **fluorescent tubes** above, beside and sometimes behind the LCD. A white diffusion panel behind the LCD redirects and scatters the light evenly to ensure a uniform display.

Display Types

There are two basic types of LCD as the passive matrix and active matrix. Passive matrix LCDs use a simple grid to supply the charge to a particular pixel on the display. The rows or columns are connected to **integrated circuits** that control when a charge is sent down a particular column or row. To turn on a pixel, the integrated circuit sends a charge down the correct column of one substrate and a ground activated on the correct row of the other. The row and column **intersect** at the designated pixel, and that delivers the voltage to untwist the liquid crystals at that pixel.

Active-matrix LCDs depend on **thin film transistors** (TFT). Basically, TFTs are tiny switching transistors and capacitors. They are arranged in a matrix on a glass substrate. To address a particular pixel, the proper row is switched on, and then a charge is sent down the correct column. Since all of the other rows that the column intersects are turned off, only the capacitor at the designated pixel receives a charge. The capacitor is able to hold the charge until the next refresh cycle. And if we carefully control the amount of voltage supplied to a crystal, we can make it untwist only enough to allow some light through. By doing this in very exact, very small increments, LCDs can create a **gray scale**. Most displays today offer 256 levels of brightness per pixel.

An LCD that can show colors must have **three subpixels** with red, green and blue color filters to create each color pixel. Through the careful control and variation of the voltage applied, the intensity of each subpixel can range over **256 shades**. Combining the subpixels produces a possible palette of **16.8 million colors** (256 shades of red x 256 shades of green x 256 shades of blue).

The LCDs are used as alternative to CRT screens in monitors and text display applications due to their power meagerness, lightness in weight and adaptability into specific applications as briefed below. Yet, they have lagged behind plasma displays in size because they are harder to make. An LCD's polarized light is highly directional, making it harder to view from the side than a cathode-ray tube (CRT) or plasma display. And the speed at which picture frames are refreshed is slower than a plasma display, causing blurring in some fast action scenes.

Painting the Screen

To "paint" the entire screen, electronic circuits inside the monitor use the magnetic coils shown in Figure 5.29 to move the electron beam in a "**raster scan**" pattern across and down the screen. The beam paints one line across the screen from left to right. It then quickly flies back to the left side, moves down slightly and paints another horizontal line, and so on down the screen. The electron beam is "on" when the beam is "painting,", and it is "off" when flying back, hence it does not leave a trail on the screen. As the beam paints each line from left to right, the intensity of the beam is changed to create different shades of the colors across the screen. Because the lines are spaced very closely together, your brain integrates them into a single image.

In monitors based on CRT technology, the refresh rate is the number of times that the image on the display is drawn each second. If your CRT monitor has a refresh rate of 72 Hertz (Hz), then it cycles through all the pixels from top to bottom 72 times a second. Refresh rates are very important because they control flicker, and you want the refresh rate as high as possible. Too few cycles per second and you will notice a **flickering**, which can lead to headaches and eye strain.

Because your monitor's refresh rate depends on the number of rows it has to scan, it limits the maximum possible resolution. Most monitors support multiple refresh rates. Keep in mind that there is a tradeoff between flicker and resolution, and then pick what works best for you. This is especially important with larger monitors where flicker is more noticeable. Recommendations for refresh rate and resolution include 1280x1024 at 85 Hertz or 1600x1200 at 75 Hertz. A CRT supports the resolution that matches its physical dot (pixel) size as well as several lesser resolutions. For example, a display with a physical grid of 1280 rows by 1024 columns can obviously support a maximum resolution of 1280x1024 pixels. It also supports lower resolutions such as 1024x768, 800x600, and 640x480. An LCD monitor works well only at its native resolution.

Two measures describe the size of your display: the **aspect ratio** and the **screen size**. Historically, computer displays, like most televisions, have had an aspect ratio of 4:3. This means that the ratio of the width of the display screen to the height is 4 to 3. For widescreen LCD monitors, the aspect ratio is 16:9 (or sometimes 16:10 or 15:9). Widescreen LCD displays are useful for viewing DVD movies in widescreen format, playing games and displaying multiple windows side by side. High definition television (HDTV) also uses a widescreen aspect ratio.

Screen sizes are normally measured in inches from one corner to the corner diagonally across from it. This diagonal measuring system actually came about because the early television manufacturers wanted to make the screen size of their TVs sound more impressive. Interestingly, the way in which the screen size is measured for CRT and LCD monitors is different. For CRT monitors, screen size is measured diagonally from outside edges of the display casing. In other words, the exterior casing is included in the measurement. For LCD monitors, screen size is measured diagonally from the inside of the beveled edge, hence the measurement does not include the casing. Because of the differences in how CRT and LCD monitors are measured, a 17-inch LCD display is comparable to a 19-inch CRT display.

Popular screen sizes are 15, 17, 19 and 21 inches. Notebook screen sizes are smaller, typically ranging from 12 to 17 inches. As technologies improve in both desktop and notebook displays, even larger screen sizes are becoming available. For professional applications, such as medical imaging or public information displays, some LCD monitors are 40 inches or larger! Obviously, the size of the display directly affects resolution. The same pixel resolution is sharper on a smaller monitor and fuzzier on a larger monitor because the same number of pixels is spread out over a larger number of inches. An image on a 21-inch monitor with an 800x600 resolution will not appear nearly as sharp as it would on a 15-inch display at 800x600.

Emerging Display Technologies

Among the important monitor technologies, we can count the touch screen monitors and wireless monitors. Each type will be briefed below and details will be left to the reader who may refer to the references for further information.

Touch-Screen and Wireless Monitors

Displays with touch-screen technology let you input information or navigate applications by touching the surface of the display. The technology can be implemented through a variety of methods, including infrared sensors, pressure-sensitive resistors or electronic capacitors. Quantum Tunneling Composite (QTC) is a new class of electrically conductive material that has been developed to advance the capability of switching and sensing systems. QTC is a pressure switching and sensing material technology and it will be briefly explained later in relation to mechanical pressure sensors.

Wireless monitors looks like tablet PC. They use technology such as 802.11b/g to connect to your computer without a cable. Most include buttons and controls for mousing and web surfing, and some also include keyboards. The displays are battery-powered and relatively lightweight. Most also include touch-screen capabilities.

Plasma Panels

Plasma is generated in a gas made up of free-flowing ions and electrons. In a plasma with an electrical current running through it, negatively charged particles are rushing toward the positively charged area of the plasma, and positively charged particles are rushing toward the negatively charged area. In this mad rush, particles are constantly bumping into each other. These collisions excite the gas atoms in the plasma, causing them to release photons of energy. Xenon and neon atoms, the atoms used in plasma screens, release light photons when they are excited. Mostly, these atoms release ultraviolet light photons, which are invisible to the human eye. But ultraviolet photons have higher energy than the visible light photons and they can be used to excite visible light photons.

A plasma panel display is made up of millions of phosphor-coated gas-filled pixel cells. Each pixel is made up of three fluorescent lights: a red light, a green light and a blue light. Just like a CRT television, the plasma display varies the intensities of the different lights to produce a full range of colors. When excited by a voltage, the gas emits UV light that makes the cells red, green or blue phosphor coating emit visible light. Having each pixel lit individually makes the image very bright and looks good from almost every angle.

Plasma displays have wide screens, comparable to the largest CRT sets, but they are only about 15 cm thick. The biggest drawback of this technology has been the price. However, falling prices and advances in technology mean that the plasma display may soon replace the old CRT sets. Proponents say that the plasma technology produces more natural colors and a softer picture than the stark brightness of a uniformly backlit LCD making viewing easier for tired eyes. However, PDP screens have a shorter lifetime than an LCD and consume more power.

Organic Light-Emitting Diode (OLED)

Organic light emitting diodes (OLEDs) are thin-film LED (Light-Emitting Diode) displays that don't require a backlight to function. OLEDs consist of stacks of organic layers (thickness about 100 nm), which are inserted between a cathode and an anode. The material emits light when stimulated by an electrical current, which is known as **electroluminescence**.

Key advantages of the organic luminescence are the chemical variability of the organic light-emitting diodes, allowing virtually any color including white, and the thin film system, allowing large-area and low-cost deposition. The possibility to use thin and even flexible substrates allow us to realize a novel class of lighting and display solutions not possible for other technologies. Advantages also include lower power requirements, a less-expensive manufacturing process, improvements in contrast and color, and the ability to bend. In the years ahead OLEDs will see applications in personal computers, cell phones, televisions, general wide area lighting, signs, billboards, communications and any of a number of information appliances.

Surface-Conduction Electron Emitter (SED) and Field Emission (FED) Displays

SED is a display technology which is currently developing various flat panel displays by a number of companies as an electronic visual displays. SEDs use nanoscopic-scale electron emitters to energize colored phosphors and produce an image. In a general sense, a SED consists of a matrix of tiny cathode ray tubes, each "tube" forming a single sub-pixel on the screen, grouped in threes to form red-green-blue (RGB) pixels.

After considerable time and effort in the early and mid-2000s, SED efforts started winding down in 2009 as LCD became the dominant technology. In August 2010, Canon announced they were shutting down their joint effort to develop SEDs commercially, signaling the end of development efforts. SEDs are closely related to another developing display technology, the field emission display, or FED, differing primarily in the details of the electron emitters. Sony, the main backer of FED, has similarly backed off from their development efforts. In a general sense, a FED consists of a matrix of cathode ray tubes, each tube producing a single sub-pixel, grouped in threes to form red-green-blue (RGB) pixels.

SEDs and their young cousins FEDs combine the advantages of CRTs, namely their high contrast levels and very fast response times, with the packaging advantages of LCD and other flat panel technologies. They also offer the possibility of requiring less power, about half that of an LCD system.

PROBLEMS

Review Questions

1. What is a waveform and how it can be displayed?
2. Why oscilloscopes are mostly used for displaying waveforms?
3. How a waveform can be displayed on an oscilloscope screen?
4. What are the fundamental components of an oscilloscope and how they work?
5. What is the basic function of the sweep signal in an oscilloscope?
6. How can you measure the frequency of a periodic signal using an oscilloscope?
7. How can you measure the magnitude of a waveform using an oscilloscope?
8. How do you estimate the measurement errors in oscilloscope displays?
9. What is the triggered sweep and how it helps in measurements?
10. What is the delay line and its function?
11. What is the difference between X-Y mode of operation and sweep mode?
12. What are the applications of X-Y mode of operation?
13. Why do we need multi-trace oscilloscopes?
14. What are the ways for obtaining multiple traces from a single electron gun?
15. What is an oscilloscope probe and how it differs from an ordinary connection wire?
16. Why we need high impedance probes?
17. What is a digital storage oscilloscope and how it differs from the analog ones?
18. What are the advantages of digital storage oscilloscopes?
19. What are the fundamental components of a digital storage oscilloscope?
20. How the time-base circuit operates in a digital storage oscilloscope?
21. What are the current trends in digital oscilloscope technology?
22. What is a virtual instrument?
23. What are the advantages of virtual instruments over the conventional measuring instruments?
24. What are the basic components of a virtual instrumentation and how they function?
25. How can virtual instrumentation be used in system design?
26. How can you obtain moving images from still pictures?
27. What are the commonly used technologies for picture display?
28. What is a picture element (pixel) and dot pitch?
29. How does a CRT based color display screen work?
30. What are the standards and resolution in picture displays?
31. What is the basic difference between the CRT tubes used for oscilloscope and picture displays?

32. How does a liquid crystal display work?
33. How does the reflective and backlit type lcd differ from each other?
34. What do we mean by active matrix and passive matrix type lcd displays?
35. What is the raster scan and how it is used as a mean of painting the screen?
36. What are the aspect ratio and viewable area for a display screen?
37. How can you compare and contrast CRT and lcd type displays?
38. What are new emerging display technologies and their principles of operation?
39. How can you obtain a touch-screen type display?
40. What is the organic led and how it is used in the display technology?
41. What are the similarities and difference between surface conduction electron emitter (SED) and field emission (FED) type displays?

Solved Examples

1. Sketch the scope waveforms for $v_1(t)$ = 1 sin(2000πt), $v_2(t)$ = 0.5 sin(2000πt-30°) with vertical settings 0.5 V/cm and 0.2 V/cm for channel 1 and 2 respectively, time base setting 0.2 ms/cm, screen height 8 cm, screen width 10 cm, trigger source channel-1, trigger level 0 V, and slope negative.

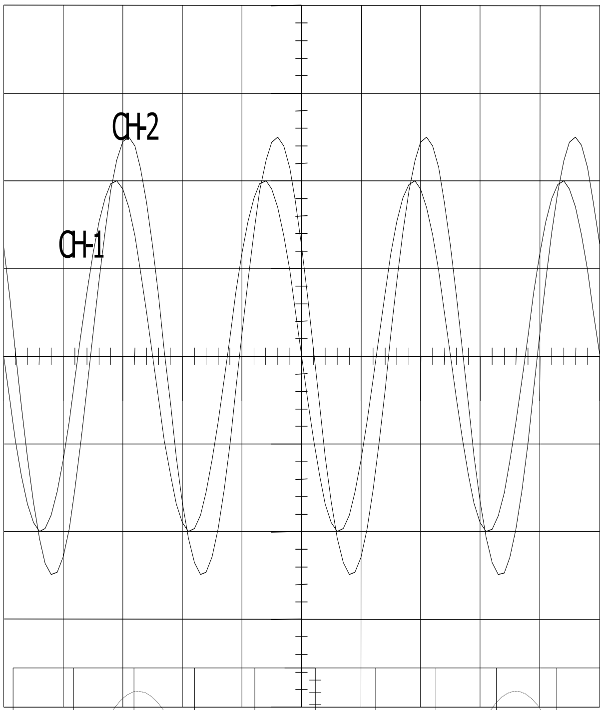


Figure for solved examples 1.

2. The input and output to an amplifier are two sinusoidal voltages V_1 and V_2 respectively. These voltages are applied to an oscilloscope in dual-trace operation as V_1 to CH-1 and V_2 to CH-2. The vertical settings for CH-1 = 20 mV/cm and CH-2 = 0.5 V/cm, time-base setting = 1 ms/cm. Assume an uncertainty of $\pm$ 0.5 mm in all distances measured. CH-2 is used for triggering. Determine:
   - Trigger level and trigger slope, and phase shift. Does V_1 leads or lags V_2?
   - The period and frequency of the signals.
   - Values of voltages V_1 and V_2 and their uncertainties.

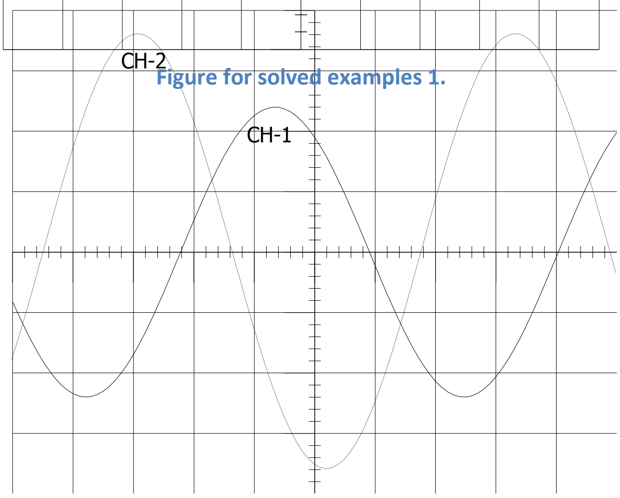


Figure for solved examples 2.

- Assume that the output is applied to a resistor 1 k$\Omega \pm$ 10%. Determine the RMS value of the power delivered to the resistor and its uncertainty.

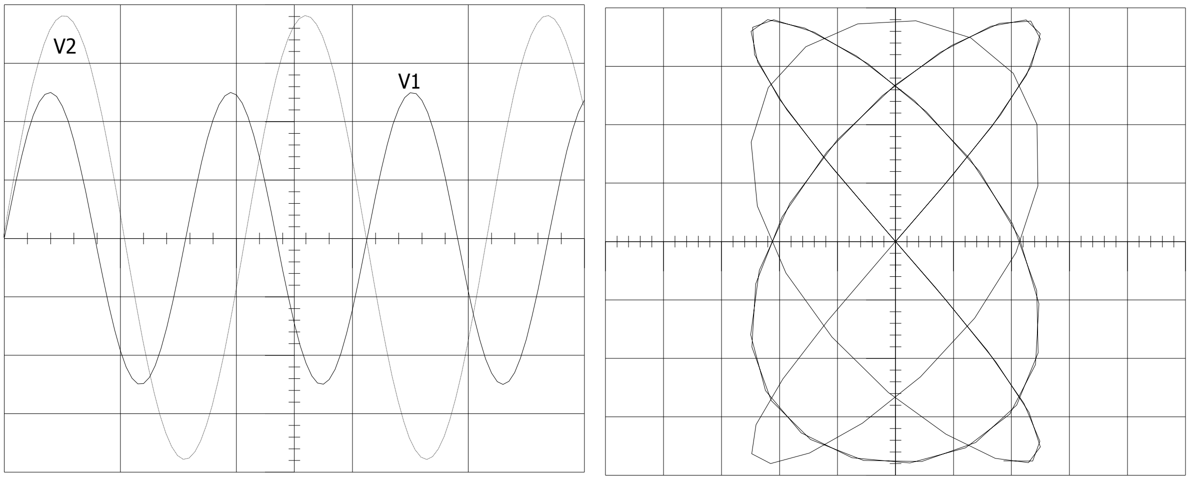
Answer

Trigger level is -0.75V; trigger slope (+), V_1 lags V_2. T=6.2 cm and d=2.3 cm leading to $\theta=2\pi x2.3/6.2$ = 2.33 rad = 133.5°

The period T = 6.2 ms and frequency f= 161 Hz.

$V_1 = (2.4 \pm 0.05)x20x10^{-3} = (48 \pm 1)$mV (peak); and $V_2 = (3.6 \pm 0.05)x0.5 = (1.8 \pm 0.025)$V (peak).

$P = \frac{V_{RMS}^2}{R}; \quad V_{RMS}^2 = \frac{V_{peak}^2}{2}; \quad P = \frac{V_{peak}^2}{2R}$ Assume peak value of the voltage is used. Then, $\frac{\partial P}{\partial V} = \frac{V}{R} = 2\frac{P}{V}; \quad \frac{\partial P}{\partial R} = -\frac{V^2}{2R^2} = -\frac{P}{R}; \quad (\Delta P)^2 = (\frac{\partial P}{\partial V})^2(\Delta V)^2 + (\frac{\partial P}{\partial R})^2(\Delta R)^2$ leading to $(\frac{\Delta P}{P})^2 = 4(\frac{\Delta V}{V})^2 + (\frac{\Delta R}{R})^2 = 0.0108$ and uncertainty in P = 10.4%. So, P= 1.62 mW $\pm$ 10.4%.



Figures for solved examples 3.

3. Two sinusoidal voltages are applied to an oscilloscope in dual-trace operation and X-Y mode of operation as shown in the figures. The sensitivities are 0.1 V/cm and 0.5 V/cm for V_1 and V_2 respectively. The time base sensitivity is 1 ms/cm. The trigger source is V_1. In the X-Y mode, V_1 is applied to X-input and V_2 is applied to the Y-input. Using both plots, calculate
   - Peak-to-peak values for both signals

V_{1p-p} = 5 cm x 0.1 V/cm = 0.5 V p-p; V_{2p-p} = 7.6 cm x 0.5 V/cm = 3.8 V p-p

- The frequency and time period of both signals

T_1 = 3.1 cm x 1 ms/cm = 3.1 ms , T_2 = 4.15 ms ; $f_1 = 1/T_1$ = 323 Hz, f_2 = 241 Hz

The ratio of frequencies. $f_1/f_2 = 4/3$ as obtained from the tangents in X-Y mode

General Questions

1. Draw a diagram showing all major blocks of the oscilloscope, and shortly describe what does each do. Show the input and output signals in blocks related to the time-base circuitry.
2. For a cathode ray tube:
   a. What are the major components? What are the factors effecting the brightness of the trace?
   b. Referring to Appendix-C: The accelerating voltage is 2,000 V, the length of deflection plates is 4 cm and separation between plates is 1 cm.
   c. What is the velocity of the electron as it enters the deflection plates?
   d. How much is the maximum deflection angle possible?
   e. What is the minimum distance required between the center of the plates and the screen if the maximum deflection on the screen is 4 cm?
   f. How much is the voltage is required across two deflection plates to full scale deflection?
   g. What is the deflection sensitivity? What is the deflection factor?
3. The oscilloscope has a screen size of 8 cm vertically and 10 cm horizontally. Sketch the scope waveforms for $v_1(t) = 1.5 \sin(300\pi t)$, $v_2(t) = 0.5 \sin(300\pi t\text{-}30°)$ on a graph paper. Available vertical settings (V/cm): 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, and 10; horizontal settings (s/cm): 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1. Select vertical and horizontal settings to obtain minimum possible measurement errors for the amplitude and time readings. Indicate your selections. Assume trigger source is channel-1, trigger level is 0 V, and trigger slope is negative.
4. For an oscilloscope:
   a. What are the parameters affected by the following knobs
      i. Intensity
      ii. Volts/cm
      iii. Time/cm
      iv. Trigger level
      v. Focus?
   b. What are the functions of the vertical deflection system?
   c. How do you obtain a multi-trace display using a single electron gun?
5. An oscilloscope is used for the measurement of phase shift θ between two signals V_1 and V_2 of the same frequency. The following results were obtained:
   a. For the ellipse method [$\theta = \sin^{-1}(y_0/y_m)$], $y_0 = (3.5 \pm 0.05)$ cm, $y_m = (5 \pm 0.05)$ cm

b. For the dual trace method [$\theta = 2\pi d/D$ rad]; d = (1 ± 0.05) cm, D = (8 ± 0.05) cm.

In both cases determine the phase shift θ and its uncertainty. Can any one of the methods (a) or (b) be used to determine if V_1 leads or lags V_2?

6. The input and output to an amplifier are two sinusoidal voltages V_1 and V_2 respectively. These voltages are applied to an oscilloscope in dual-trace operation as V_1 to CH-1 and V_2 to CH-2. The vertical settings for CH-1 = 20 mV/cm and CH-2 = 0.5 V/cm, time-base setting = 1 ms/cm. Assume an uncertainty of ± 0.5 mm in all distances measured. CH-2 is used for triggering. Determine:

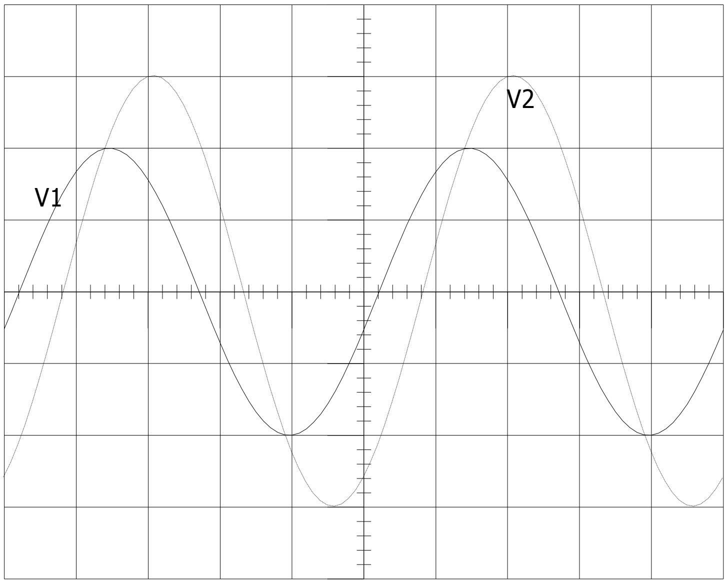


Figure p6.

   a. Trigger level and trigger slope, and phase shift. Does V_1 leads or lags V_2?
   b. The period and frequency of the signals.
   c. Values of voltages V_1 and V_2 and their uncertainties.
   d. Determine the gain of the amplifier ($G=V_2/V_1$) and its uncertainty.

7. Now the time-base is switched off and the oscilloscope is set to XY mode of operation. V_1 is connected to the X input with sensitivity 20 mV/cm, and V_2 is connected to the Y input with sensitivity 0.5 V/cm. Draw the resulting ellipse to the space at the right. Mark carefully the values of distances Y_m, Y_0, X_m, and X_0.

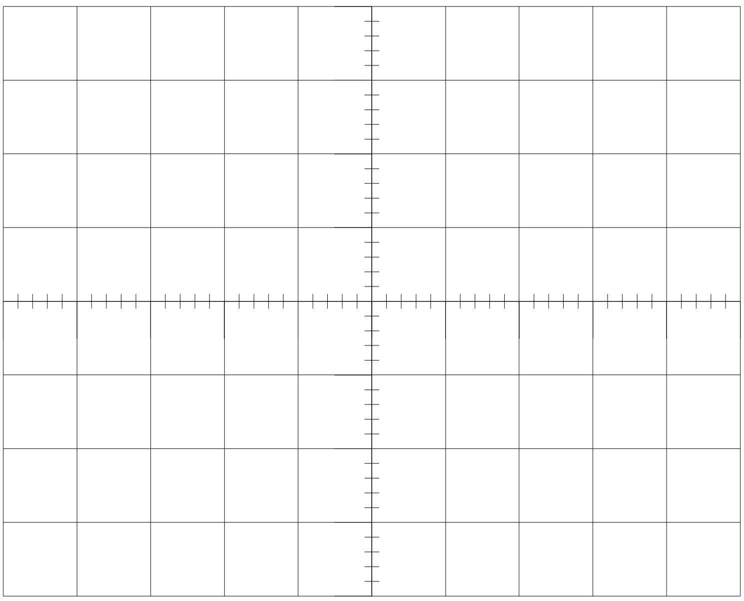


Figure p7.

8. Draw the block diagram related to the trigger and time-base

circuitry of the oscilloscope including all input/output and control connections. Describe the function of each block shortly. Show the input and output signals that would appear in the blocks.

9. For the oscilloscope, explain:
    a. Functions of the intensity control and the focus adjustments;
    b. Generation of the sweep signal that drives the x deflection plates of the CRT in a **digital storage oscilloscope** (DSO).
    c. Explain how to obtain a dual-trace display using a single electron gun;
    d. The input and output to an amplifier are two sinusoidal voltages V_1 and V_2 respectively. These voltages are applied to an oscilloscope in dual-trace operation as V_1 to CH-1 and V_2 to CH-2. The vertical settings for CH-1 = 20 mV/cm and CH-2 = 0.5 V/cm, time-base setting = 2 ms/cm. Assume an uncertainty of $\pm$ 0.5 mm in all distances measured. Determine:
        i. Value of voltages (with errors) if the peak vertical deflections on the screen are 3.5 cm and 2.8 cm for CH-1 and CH-2 respectively;
        ii. The gain of the amplifier and it's uncertainty ($G=V_2/V_1$)
        iii. The frequency of the signal with percentile error if one signal period is 6.2 cm.
10. For an oscilloscope:
    a. What are the major components of a cathode ray tube (CRT)? What are the factors effecting the brightness of the trace? Explain conversion of voltage into displacement in the CRT. What is the deflection sensitivity?
    b. For the digital storage oscilloscope, explain the function of sample and hold circuit by showing typical input and output signals.
    c. The triangular voltage waveform shown is applied to the oscilloscope in trigger mode. Trigger level is "0" and slope "-". Time-base setting = 0.5ms/div. and vertical setting = 0.2 V/div. 1div. = 1cm. Draw the waveform carefully to the CRO screen given.
    d. A sinusoidal voltage waveform with frequency 1.0 kHz is applied to the X-input and other sinusoidal voltage waveforms with unknown frequencies are applied to the Y-input one-by-one. Following stable figures are obtained.

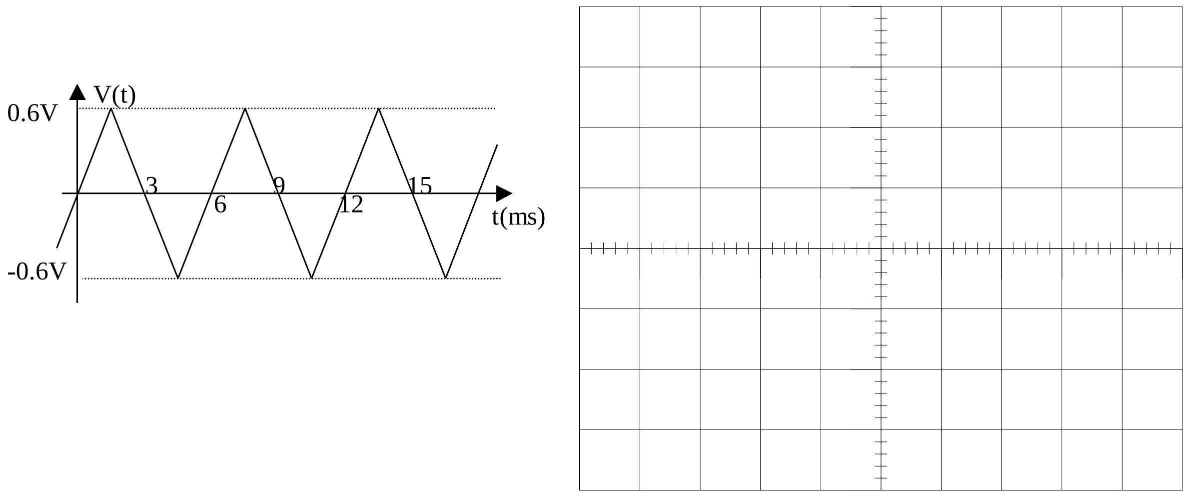


Figure p10-c.

Determine the frequency for each case.

11. Compute the phase shifts between X and Y for (b), (c) and (d) and write them down below the figure.

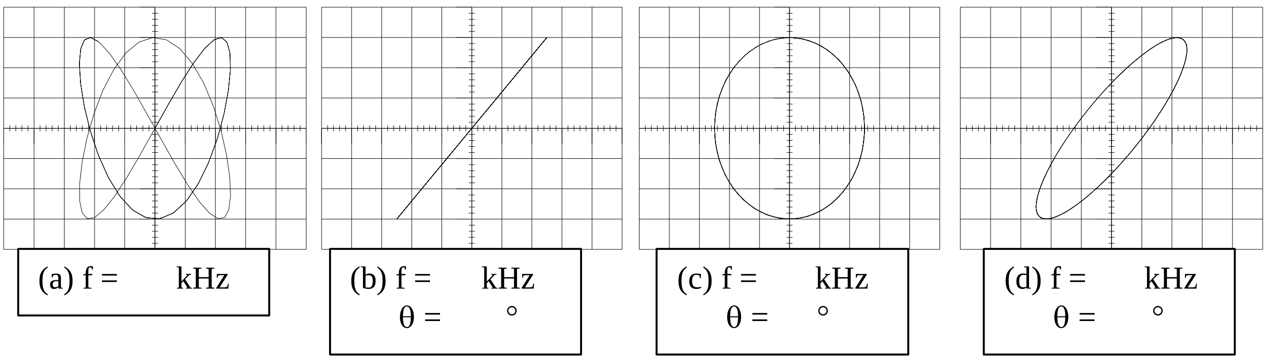


Figure p11.

12. In the sketch shown, the vertical settings are given as 0.2 V/div and 0.5 V/div for CH-1 and CH-2

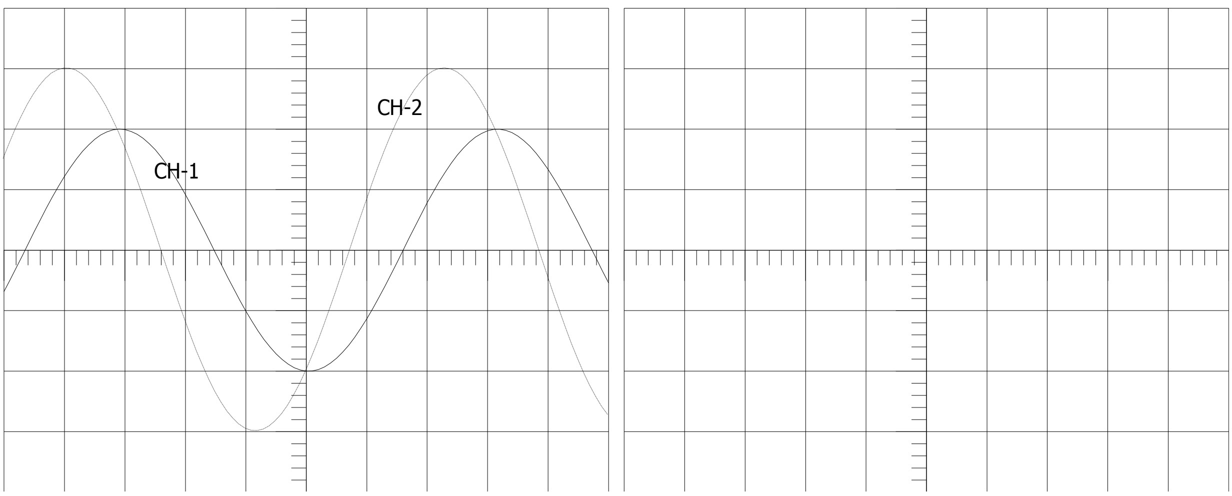


Figure p12.

respectively. The time base setting is 1 ms/div. CH-2 is used for triggering. The uncertainty is ± 0.5 mm in all distances measured. Make necessary measurements and fill in the blank spaces:

a. Peak to peak amplitude for CH-1: ___________V; % error = ________ %

b. Peak to peak amplitude for CH-2: __________ V; % error = ________ %

c. Trigger level: ____V and slope: __

d. Frequency of the signal in CH-1: _____ kHz with % error = ________ %

e. The phase shift is: _____ degrees with % error = ________ % CH- ___ is the leading one. Sketch the X-Y plot for the waveforms displayed. Use CH-1 for the X-input and CH-2 for the Y-input.

13. Two sinusoidal voltages are applied to an oscilloscope in dual-trace operation. The vertical settings are 10 mV/cm and 0.5 V/cm for CH-1 and CH-2 respectively. The time base setting is 10

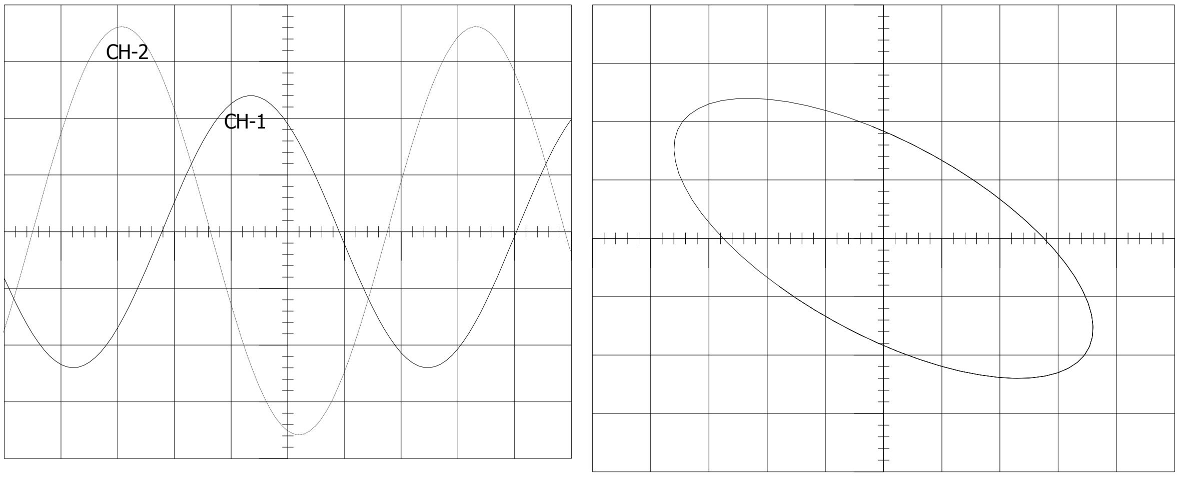


Figure p13.

ms/cm. The trigger source is CH-2. For the dual trace shown, find

a. The peak to peak values for voltages in CH-1 and CH-2

b. The time period and frequency of both signals.

c. The trigger level and trigger slope.

d. The phase shift between V_1 (CH-1) and V_2 (CH-2). Does V_1 leads or lags V_2?

14. Now the oscilloscope is switched to XY mode of operation. V_1 is connected to the X input with setting 10 mV/cm, and V_2 is connected to the Y input with setting 0.5 V/cm. Draw the resulting ellipse, and calculate

a. Distances Y_m, Y_0, X_m, and X_0.

b. Phase shift between signal in X and signal in Y.

c. Peak-to-peak values of voltage for V_1 and V_2

15. Using the X-Y plot given, calculate the required parameters and write down the corresponding

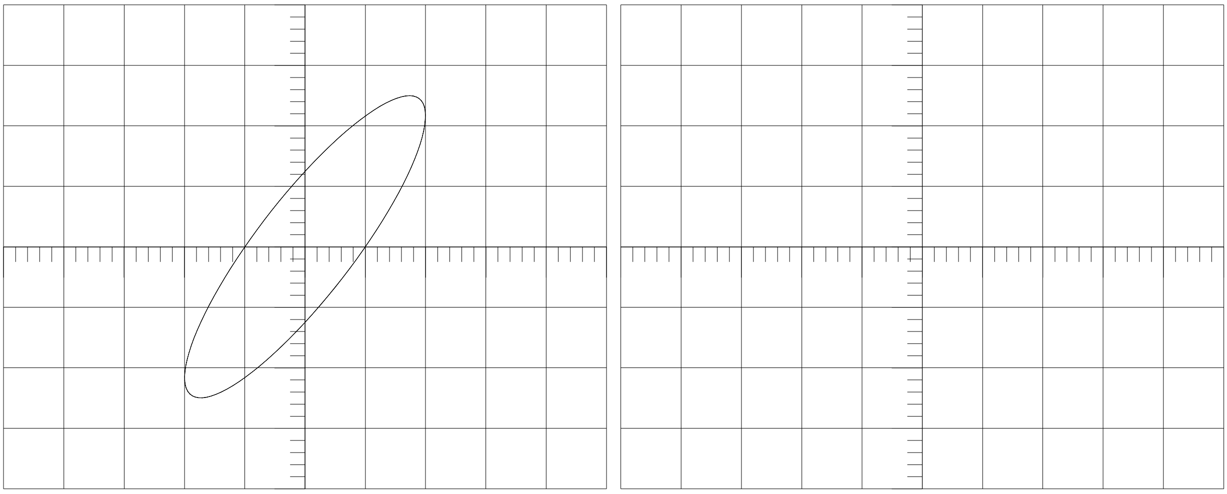
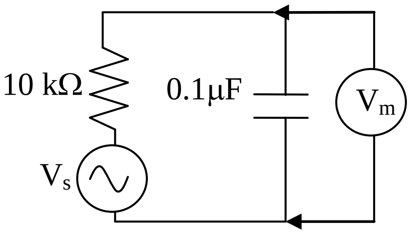


Figure p15.

values into the fill in the blank spaces provided. The settings for both channels is 1 V/div.

- Maximum value of the signal in X: ____ V
- Maximum value of the signal in Y: ____ V
- The phase shift between X and Y : _____ degrees.

16. Sketch the sweep mode display for figure p15 that you would see on the oscilloscope screen if X is applied to CH-1 and Y is applied to CH-2. Assume the frequency is 1 kHz, time-base setting is 0.2 ms/div. Trigger source is CH-1 with 0 level and positive slope.
17. Explain the following terms related to the digital storage oscilloscope:
    a. Sampling;
    b. Quantization
    c. Control logic
    d. Digital to analog converter.
18. For the following RC circuit



    a. Determine the time constant of the circuit
    b. Draw the input and output waveforms for a square wave input with magnitude ± 1 V and frequency
        i. 100 Hz
        ii. 1 kHz
        iii. 10 kHz
    c. Draw the input and output waveforms for $v_s(t) = 5\cos(2000\pi t)$.
19. Set the circuit in the previous problem and experimentally verify the correctness of your solutions and determine the time constant of the circuit.

BIBLIOGRAPHY

Further Reading

G. Held, Introduction to Light Emitting Diode Technology and Applications, CRC Press, 2009.

Useful Websites

SOURCES OF ELECTRICAL ENERGY

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Express the need for a power source and define a power supply.
2. Explain the power supply terms such as ripple factor and load regulation.
3. State power supply types.
4. Draw the block diagram representation of a linear regulated power supply.
5. Discuss the need for AC line components for an unregulated power supply and briefly explain the function of each component.
6. Describe rectifier diodes and bridges and select the proper type for a given application.
7. Describe types of smoothing filters and compute the requirements for a given application.
8. Discuss the need for a regulator.
9. Explain development of linear (dissipative)regulators and select an IC regulator for a given application.
10. Explain devices used for protection of circuits in case of regulator failure.
11. Compare and contrast linear and switching type power supplies.
12. Describe the general layout and principle of operation of switching power supplies.
13. Describe rectifiers and filters of a switching power supply.
14. State various switching regulator configurations.
15. Discuss briefly advantages and disadvantages of switching supplies.
16. Use key formulas available in solving power supply problems.
17. Illustrate principles of operation of batteries.
18. Differentiate between primary and secondary batteries.
19. List categories and types of commonly used batteries.
20. Express the battery capacity.
21. Describe techniques for care and maintenance of batteries.
22. Define the scope and purpose of electrical safety.
23. Define the electrical shock and describe how it occurs.
24. Discuss methods for preventing the electrical shock.
25. Describe faults that commonly occur in offices and electrical safety measures to prevent the electric shock.

LINEAR REGULATED POWER SUPPLIES

Definitions

A power supply is a device that supplies electrical energy to one or more electric loads. A regulated power supply is one that controls the output voltage or current to a specific value; the controlled value is held nearly constant despite variations in either load current or the voltage supplied by the power supply's energy source. The power supply obtains the energy that it supplies to its load, as well as any energy it consumes while performing that task, from an energy source. Depending on its design, a power supply may obtain energy from:

- Electrical energy transmission systems. Common examples of this include power supplies that convert AC line voltage to DC voltage as in the case of the laboratory power supply.
- Energy storage devices such as batteries and fuel cells.
- Electromechanical systems such as generators and alternators.
- Solar power.

A power supply may be implemented as a discrete, stand-alone device or as an integral device that is hardwired to its load. In the latter case, for example, low voltage DC power supplies are commonly integrated with their loads in devices such as computers and household electronics. Whatever the type and application might be, constraints that commonly affect power supplies include:

- The amount of voltage and current they can supply.
- How long they can supply energy without needing some kind of refueling or recharging (applies to power supplies that employ portable energy sources).
- How stable their output voltage or current is under varying load conditions.
- Whether they provide continuous or pulsed energy.

The laboratory power supply converts alternating current to DC current to meet the power requirements of solid-state electronic circuits as illustrated in Figure 6.1. DC voltages from 3 to 24 volts are used with ±5 volts, ±6 volts and ±12 volts being most popular. The ideal power supply can provide the output DC current from 0 ampere (no load) to the maximum (full load) without any change in the output voltage. The closeness of a practical power supply is determined by two parameters as the *ripple factor (r)* and *load regulation*.

$$ripple\,factor(r) = \frac{Effective(rms)\,value\,of\ alternating\ component\,of\ the\,wave}{Average(dc)\,value\,of\ the\,wave}$$

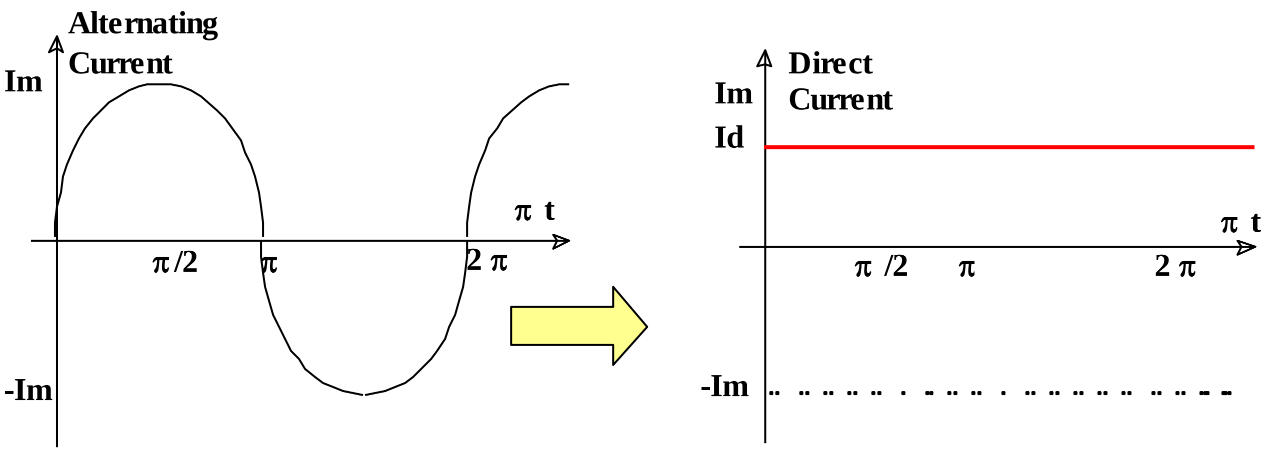


Figure 6.1 Power supply converts alternating current (AC) to direct current (DC)

$$Load\ regulation(\%) = \frac{V_{ML} - V_{FL}}{V_{\otimes}} 100 = \frac{R_0}{R_L} x\, 100$$

where V_{ML} and V_{FL} represent the output voltage for minimal load (or with open circuit) and full (maximal) load respectively. V_O is the nominal (reference) output voltage and it is generally taken as V_{FL}.

There are other factors like the efficiency, power dissipation, cost, complexity, weight etc. related to the power supply performance. Two of them are the *input regulation* that represents the capability of the power supply to adjust its output under varying input conditions and the *efficiency (η)*.

$$Input\ regulation(\% / V_{IN}) = \frac{\Delta V_O}{\Delta V_{IN} \bullet V_O} x\, 100$$

$$Efficiency(\eta) = \frac{P_{OUT}}{P_{IN}} = \frac{V_O \, x \, I_L}{V_{IN} \, x \, I_{IN}}$$

where ΔV_O is the change that takes place at the output voltage in response to the change at the input voltage (ΔV_{IN}).

Power Supply Types

Power supplies for electronic devices can be broadly divided into linear and switching power supplies. The linear supply is usually a relatively simple design, but it becomes increasingly bulky and heavy for high-current equipment due to the need for large mains-frequency transformers and heat-sinked electronic regulation circuitry. Linear voltage regulators produce regulated output voltage by means of an active voltage divider that consumes energy, thus making efficiency low. A switched-mode supply of the same rating as a linear supply will be smaller, is usually more efficient, but will be more complex. Power supplies used with small electronic equipment, either embedded or provided externally utilize linear regulation schemes while those used in computers, printers and other electronic equipment that require large currents use switching regulation strategies.

General Outline of a Linear Power Supply

The voltage produced by an unregulated power supply will vary depending on the load and on variations in the AC supply voltage. For critical electronics applications a linear regulator may be used to set the voltage to a precise value, stabilized against fluctuations in input voltage and load. The regulator also greatly reduces the ripple and noise in the output direct current. Linear regulators often provide current limiting, protecting the power supply and attached circuit from overcurrent.

Adjustable linear power supplies are common laboratory and service shop test equipment, allowing the output voltage to be adjusted over a range. For example, a bench power supply used by circuit designers may be adjustable up to 30 volts and up to 5 amperes output. Some can be driven by an external signal, for example, for applications requiring a pulsed output.

Figure 6.2 shows a general block diagram of a linear regulated power supply. The DC voltage

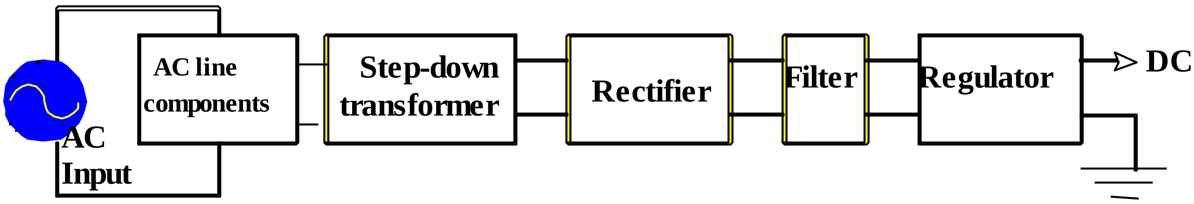


Figure 6.2 Block diagram of a linear-regulated power supply

is obtained from the line (mains) voltage. The first step is to drop the line voltage down to the level needed. This is carried out by a step-down transformer. Then, conversion of AC to DC takes place at the stage of rectifier. The filter reduces the ripple factor and the regulator diminishes the ripple factor and improves the regulation. Each block will be presented proceeding sections below.

AC Line Components for An Unregulated Power Supply

The portion of the power supply that contains AC line components (optional for linear power supplies, but compulsory for switching supplies), input step-down transformer (in linear power supplies), rectifier and filter. The output contains ripple and varies with the load and input variations. Hence, it is an unregulated DC voltage.

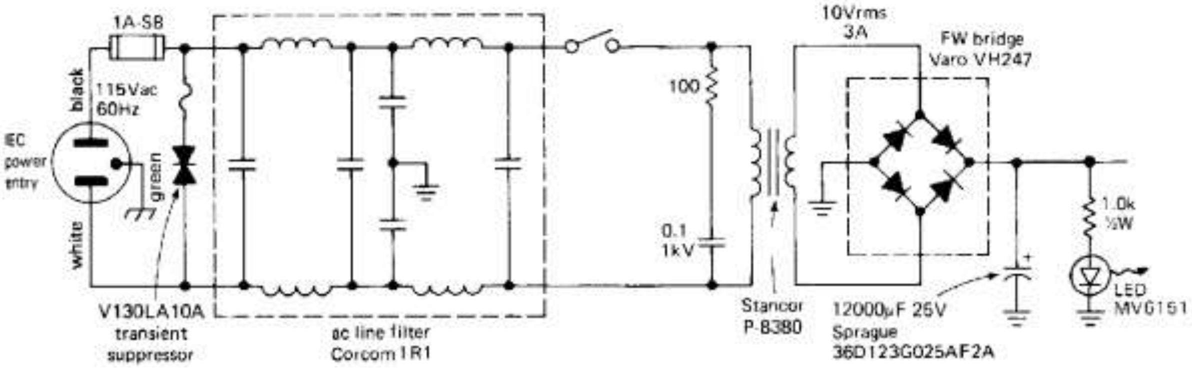


Figure 6.3 An unregulated power supply with transient suppressors and line filters

Figure 6.3 shows a schematic diagram of an unregulated supply with AC line components. They include:

1. The input socket and wiring,
2. A fuse,
3. A transient suppressor,
4. An AC line filter,
5. A power on/off switch,
6. An RC snubber.

Some of the components are optional but highly recommended.

Input Socket and Wiring

It is **always essential** to use a three-wire connection (cord) with ground (green) connected to the instrument case. Transformer insulation may fail leading to accidental connection of one side of power line to the case. With grounded case, the fuse blows and protects the user. The attachments of the ground wire to the case must be done by a "strain relief" wiring. All wiring going to the mains supply must be properly insulated possibly with heat shrinking tubing. A wiring convention must be observed (black for hot, white for neutral and green for ground).

Fuse

A fuse is a piece of wire, often in a casing that improves its electrical characteristics. If too much current flows, the wire becomes hot and melts. This effectively disconnects the power supply from its load, and the equipment stops working until the problem that caused the overload is identified and the fuse is replaced. There are various types of fuses used in power supplies.

- fast blow fuses cut the power as quick as they can
- slow blow fuses tolerate more short term overload

- wire link fuses are just an open piece of wire, and have poorer overload characteristics than glass and ceramic fuses.
- Some power supplies use a very thin wire link soldered in place as a fuse.

The fuse is an essential component with every piece of electronic equipment. A "slow-blow" type is preferred in the power-line circuit, due to large current transient at the turn-on. It is recommended to use a fuse at least 50% larger than the nominal load current. Fuses blow out more frequently due to fatigue if they are used near there rated currents.

Line Filter and Transient (Surge) Suppressor

Line filters prevent possible radiation of radio frequency interference from the instrument via the power line. At the same time, filter out incoming interference that may be present on the power line. Spikes as large as 1 kV to 5 kV are occasionally present at most power lines with smaller ones appearing more frequently. A line filter is reasonably effective in reducing such an interference.

A transient (surge) suppressor is a device that conducts when its terminal voltage is exceeded. It behaves as a bidirectional high-power zener and it can short out hundreds of amperes of harmful currents in form of spikes. It must be selected to have a turn-on voltage larger than the largest input voltage we nominally have. For example, the peak value of 127 Vrms line voltage is around 180 volts. As the line voltage fluctuates around the nominal value by 20%, this voltage will rise to 216 volts. Hence, a device with a higher voltage must be selected.

The Snubber

A snubber is a device used to suppress ("snub") voltage transients in electrical systems, pressure transients in fluid systems, or excess force or rapid movement in mechanical systems. Snubbers are frequently used in electrical systems with an inductive load where the sudden interruption of current flow often leads to a sharp rise in voltage across the device creating the interruption. This sharp rise in voltage is a transient and can damage and lead to failure of the controlling device. A spark is likely to be generated (arcing), which can cause electromagnetic interference in other circuits. The snubber prevents this undesired voltage by conducting transient current around the device.

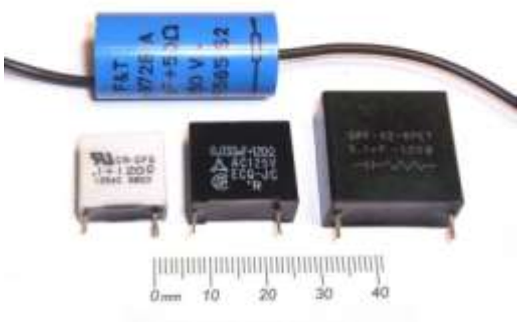


Figure 6.4 Examples of RC Snubbers

Figure 6.4 illustrates a few of commercially available snubbers. A simple snubber uses a small resistor (R) in series with a small capacitor (C). This combination can be used to suppress the rapid rise in voltage across a thyristor, preventing the erroneous turn-on of the thyristor; it does this by limiting the rate of rise in voltage (dV/dt) across the thyristor to a value which will not trigger it. Snubbers are also often used to prevent arcing across the contacts of relays and switches and the electrical interference and welding/sticking of the contacts that can occur. An appropriately-designed RC snubber can be used with either DC or AC loads. This sort of snubber is commonly used with inductive loads such as electric motors. The voltage across a capacitor cannot change instantaneously, so a decreasing transient current will flow through it for a small fraction of a second, allowing the voltage across the switch to increase more slowly when the switch is opened. While the values can be optimized for the application, a 100 ohm non-inductive resistor in series with a 100 nanofarad, or larger, capacitor of appropriate voltage rating is usually effective. Determination of voltage rating can be difficult owing to the nature of transient waveforms; the actual rating can be determined only by measuring temperature rise of the capacitor. This type of snubber is often manufactured as a single component.

A series combination of 100 ohms and 0.1 µF (1 kV) capacitor is useful in preventing the large inductive transient that the transformer would otherwise produce at turn-off as indicated in Figure 6.3. The snubber can be placed across the primary of the transformer or across the power on/off switch.

The Indicator Lamp

In some old power supply designs, a pilot light using a neon lamp and dropping resistor appears at the input section after the switch. Most new designs however, utilize a light emitting diode (led) that runs from the regulated voltage as shown in Figure 6.3.

The Transformer

The transformer has been discussed in Chapter 2. It has two functions in a power supply as:

- Stepping down the line voltage to levels required in electronics;
- Isolating the important parts of the electronic circuitry from the lines, hence providing electrical safety.

  The transformer must be selected to give us the voltage and current needed at worst case. For finding the voltage,

- Add the minimum required unregulated output voltage, ripple voltage, diodes forward voltage drop(s).

- Then, multiply the total with 0.707 (since the transformer's output voltage is expressed in rms) and $V_{transformer} = \frac{V_{unregulated} + V_{ripple} + V_{diodes}}{\sqrt{2}\,\eta}$

  This voltage must be supplied when the input has its lowest value. The efficiency of a transformer feeding a bridge rectifier and capacitive filter is around 0.81.

- Divide it by the expected efficiency of the transformer (η).

The current that is supplied by the secondary of the transformer depends upon the type of filter used. With capacitive filters, current flows for a very short duration of the period. Hence, the current can be taken as 0.7 of the load current (I_L) for an inductive filter, and 1.8 of I_L for a capacitive filter following a bridge rectifier.

Rectifiers

Diodes

Diodes allow electricity to flow in only one direction. The arrow of the circuit symbol shows the direction in which the current can flow as indicated in Figure 6.5. There is a small voltage across a conducting diode, it is called the forward voltage drop and is about 0.7 V for all normal diodes which are made from silicon. The forward voltage drop of a diode is almost constant whatever the current passing through the diode so they have a very steep characteristic (current-voltage graph) as shown in Figure 6.6. When a reverse voltage is applied a perfect diode does not conduct, but all real diodes leak a very tiny current of a few μA or less. This can be ignored in most circuits because it will be very much smaller than the current flowing in the forward direction. However, all diodes have a maximum reverse voltage (usually 50V or more) and if this is exceeded the diode will fail and pass a large current in the reverse direction, this is called breakdown.

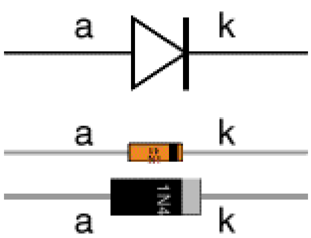


Figure 6.5 Circuit symbol and examples of diodes

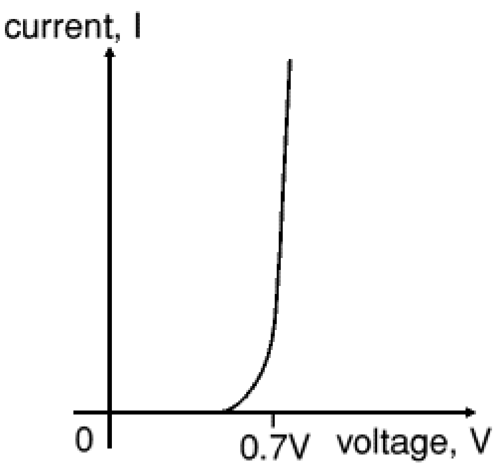


Figure 6.6 Forward current-voltage characteristic of a silicon diode

Ordinary diodes can be split into two types: Signal diodes which pass small currents of 100mA or less and rectifier diodes which can pass large currents. In addition there are light emitting diodes (LEDs) and Zener diodes.

Diodes must be connected the correct way round, the diagram may be labeled *a* or + for anode and *k* or - for cathode (yes, it really is k, not c, for cathode!). The cathode is marked by a line painted on the body. Diodes are labeled with their code in small print that may be difficult to read with bare eye (you may need a magnifying glass to read this on small signal diodes!) You can use a multimeter or a simple tester (battery, resistor and LED) to check that a diode conducts in one direction but not the other.

Rectifier Diodes

Ordinary signal diodes (like 1N4148) are designed for high speed, low leakage and low capacitance. They can handle currents up to about 100 mA with breakdown voltages rarely exceeding 100 volts. Rectifier diodes are used in power supplies to convert alternating current (AC) to direct current (DC), a process called rectification. They are also used elsewhere in circuits where a large current must pass through the diode. All rectifier diodes are made from silicon and therefore have a forward voltage drop of 0.7 V. For large current applications, the diode drop can be taken as 1.2 volt for a single diode (2.4 volts for a bridge rectifier). Rectifier diodes and bridges they can sustain currents up to 1 to 25 amps with surge currents even much greater. Their breakdown voltages ranges from 100 volts to 1000 volts. Their leakage is relatively high and junction capacitors are large making them unsuitable for signal operations. Table 6.1 shows maximum current and maximum reverse voltage for some popular rectifier diodes. The 1N4001 is suitable for most low voltage circuits with a current of less than 1A.

Table 6.1 Commonly used rectifier diodes

Diode	Maximum Current	Maximum Reverse Voltage
1N4001	1A	50V
1N4002	1A	100V
1N4007	1A	1000V
1N5401	3A	100V
1N5408	3A	1000V

There are four factors that must be considered in selection:

- Average rectified forward current (I_F) (averaged over a full cycle of operation). For famous 1N400x series it is 1 ampere.
- Surge current (I_{SFM}) is the maximum (peak) safe current for a given number of cycles. For 1N400x series it is about 30 A.
- Peak inverse voltage (PIV), V_{RM} is the maximum reverse voltage that can be applied across the diode before the onset of the avalanche breakdown. Values vary from 50 volts (1N4001) to a maximum of 1000 volts (1N4007).
- Forward voltage drop (V_F) is the DC voltage drop across the forward biased diode while the specified forward current I_F is flowing through. For 1N400x, $V_F \approx 1.1$ volt at $I_F = 1$ A.

There are several ways of connecting diodes to make a rectifier to convert AC to DC. The bridge rectifier is the most important and it produces **full-wave** varying DC. A full-wave rectifier can also be made from just two diodes if a centre-tap transformer is used, but this method is rarely used now that diodes are cheaper. A single diode can be used as a rectifier but it only uses the positive (+) parts of the AC wave to produce **half-wave** varying DC.

Three configurations are used with the rectifiers as the half-wave rectifier, full wave rectifiers with a center-tapped transformer and with a bridge rectifier.

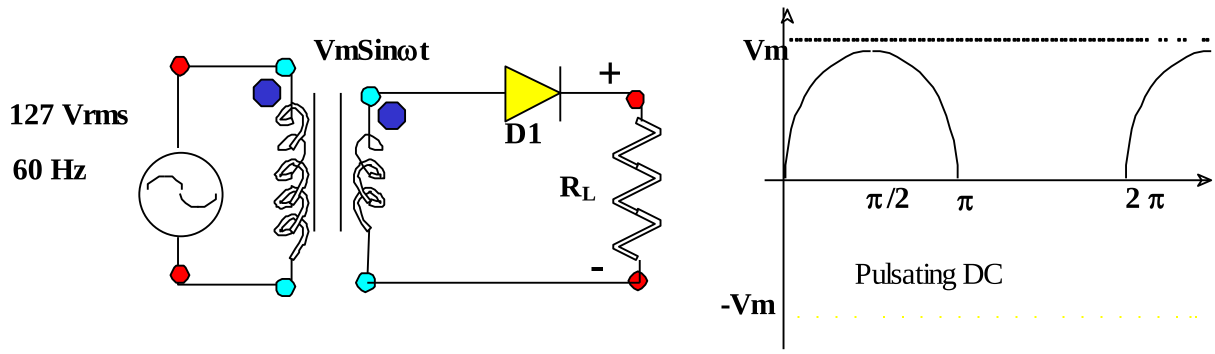


Figure 6.7 Half-wave rectified power supply

Half-Wave Rectifier

It is the simplest form as illustrated in Figure 6.7. A diode is used to clip the negative half of the input waveform. It is hard to smooth this sufficiently well to supply electronic circuits unless they require a very small current so the smoothing capacitor does not significantly discharge during the gaps.

$Vo = Vm - Vd$ with $Vd \cong 1$ volt. $Vdc = (Vm - Vd)/\pi$, $Vrms = (Vm - Vd)/2$

yielding a ripple factor (r) = 1.21

Full-Wave Rectifiers

They utilize both half of the input waveform. A center-tapped transformer provides the ground reference for the output as shown in Figure 6.8.

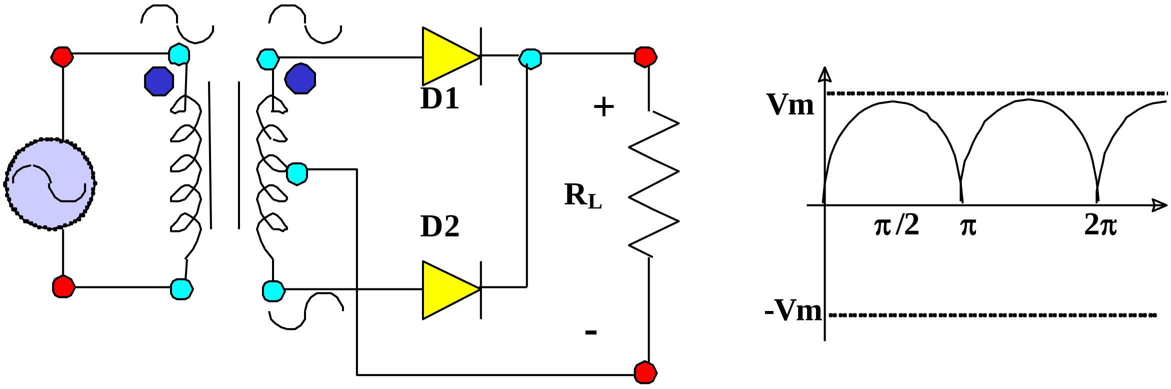


Figure 6.8 A full-wave rectified power supply with a center-tapped transformer

Vo = Vm - Vd with *Vd ≅ 1* volt. *Vdc = 2*(Vm - Vd)/π*, *Vrms = (Vm - Vd)/√2* yielding a much reduced ripple factor that is r = 0.483

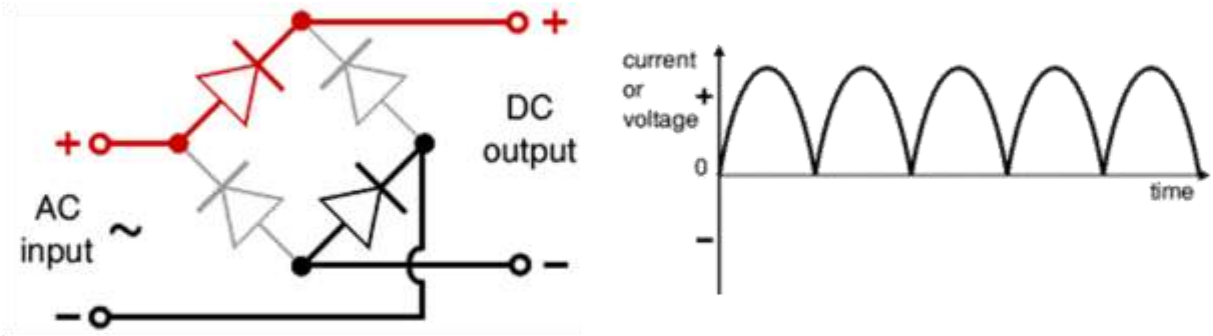


Figure 6.9 The bridge rectifier and its output waveform

An alternative and mostly used form is the bridge rectifier that uses four diodes but does not require a center tapped transformer. Figure 6.9 illustrates the bridge rectifier and its output waveform.

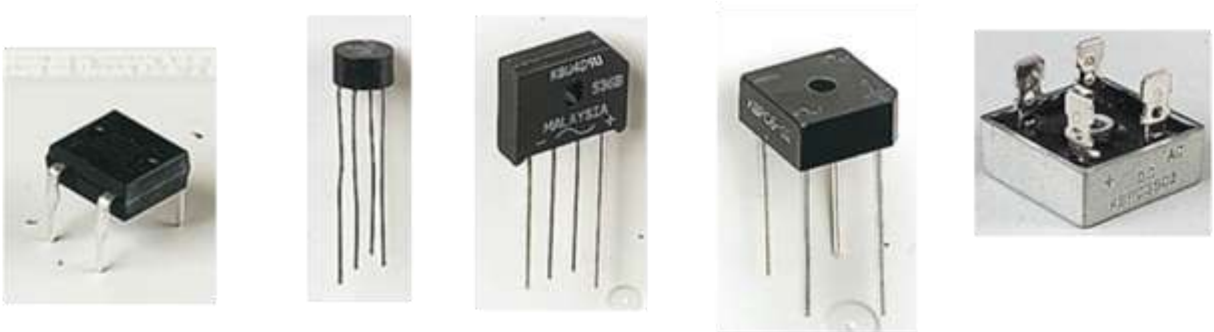


Figure 6.10 Various bridge rectifiers that are used in practice

It can be made using four individual diodes, but it is also available in special packages containing the four diodes required as shown in Figure 6.10. It is called a full-wave rectifier because it uses all the AC wave (both positive and negative sections). The output voltage is two diode drops below the input voltage. Rest of the parameters are the same as above. Bridge rectifiers are rated by the maximum current they can pass and the maximum reverse voltage they can withstand (this must be at least three times the supply RMS voltage so the rectifier can withstand the peak voltages). Figure 6.11 illustrates the full-wave rectified power supply with a bridge rectifier.

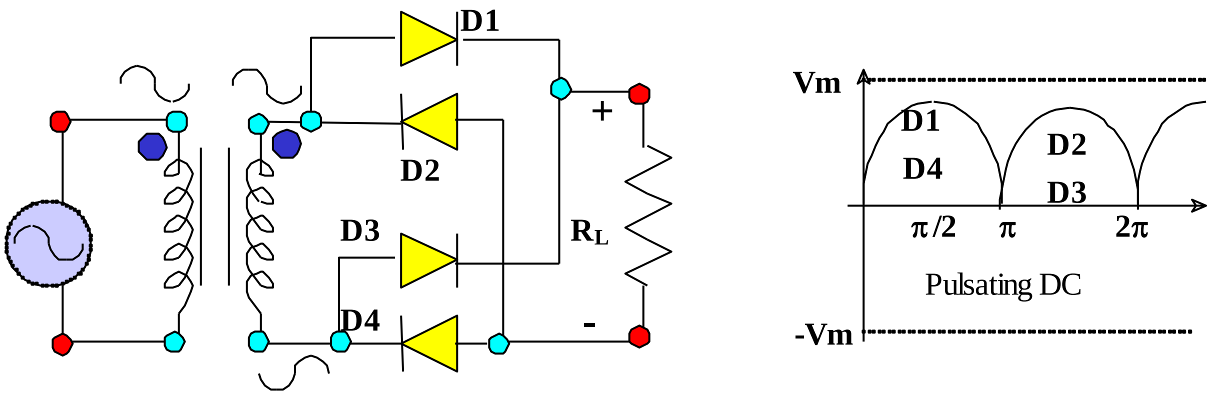


Figure 6.11 Full-wave rectified powers supply with a bridge rectifier

Smoothing Filters

Smoothing by Capacitive Filters

Smoothing is mostly performed by a large value electrolytic capacitor connected across the DC supply to act as a reservoir, supplying current to the output when the varying DC voltage from the rectifier is falling. It uses the principle that the voltage across a capacitor cannot change instantaneously. Hence, the capacitor behaves as an open circuit to DC and short circuit to AC components of the rectified signal. Figure 6.12 shows the unsmoothed varying DC (dotted line) and the smoothed DC (solid line). The capacitor charges quickly near the peak of the varying DC, and then discharges as it supplies current to the output.

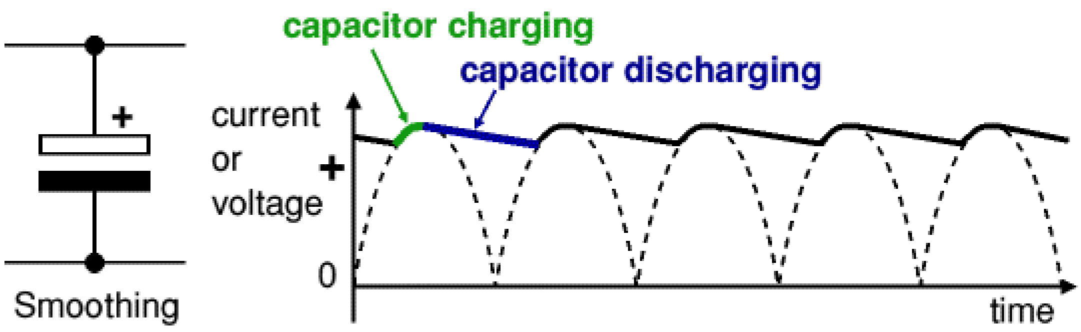


Figure 6.12 Output smoothing using an electrolytic filter capacitor

The smoothing significantly increases the average DC voltage to almost the peak value (1.4 × RMS value – diode voltage drops). For example 6V RMS AC is rectified to the peak value of about 8.4V RMS, with smoothing this increases to almost the peak value giving 6.4V smooth DC (2V is lost in the bridge rectifier).

Smoothing is not perfect due to the capacitor voltage falling a little as it discharges, giving a small **ripple voltage**. For many circuits a ripple which is 10% of the supply voltage is satisfactory and the equation below gives the required value for the smoothing capacitor.

$$Smoothing\ capacitor\ for\ 10\%\ ripple, C = \frac{5xIo}{V_s xf}$$

Where

- C = smoothing capacitance in farads (F)
- Io = output current from the supply in amps (A)
- Vs = supply voltage in volts (V), this is the peak value of the unsmoothed DC
- f = frequency of the AC supply in hertz (Hz), 50 Hz

Capacitor can be reduced by 20% if the frequency is 60 Hz instead of 50 Hz. A larger capacitor will give less ripple. The capacitor value must be doubled when smoothing half-wave DC.

Figure 6.13 shows the circuit diagram and output waveform for a capacitive filter. Here *Vi* is the rectified input, *Vo* is the filtered output, R_L is the effective load resistance *T* is the period of the Ac input, *T1* is the "off" and *T2* is the "on" time of the rectifier diodes. As the input voltage increases the capacitor charges to the maximum value of the input voltage as the rectifier diode turns on. As the input voltage starts decreasing, the voltage across the capacitor becomes greater than that of the output of the transformer. Hence, the rectifier diode turns off. The capacitor discharges slowly through the effective load resistance. In the second half cycle, as the input voltage becomes larger than the voltage across the capacitor, the diode turns on and charges the capacitor to the maximum voltage. The load causes the capacitor to discharge.

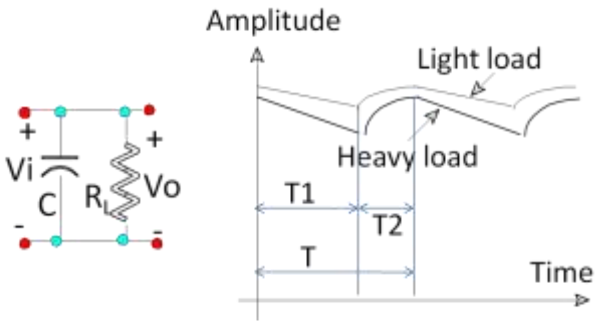


Figure 6.13 Output waveform of a capacitive filter

If we assume that the load current stays constant, the ripple voltage (peak to variation at the top of the waveform) can be approximately from the charge lost by the capacitor as $I_L=C*Vr/T_1$ yielding $Vr=I_L*T_1/C$.

The capacitor recovers the charge lost in T_2 as the diodes conduct. If T_2 is much smaller than T_1 (R_LC » T), than $T_1 \approx T = 1/f$. Hence we get $Vr=I_L/fC$ for half wave and $Vr=I_L/2fC$ for full wave rectification. The ripple factor and the DC output voltage can be estimated by $r=2400/R_LC$ and $Vdc=(Vi - 4200Idc/C)$ where C is in μF and frequency is 60 Hz.

Large electrolytic capacitors are used to obtain acceptably low ripple voltage. We can decrease the ripple voltage by increasing the value of the capacitor. However, this will cause a decrease in charging time T_2 and necessitates larger currents to flow through the rectifier diodes. Eventually, rectifier diodes and the transformer will be afflicted by increased I^2R heating.

The value of the capacitor is chosen according to the ripple voltage we can tolerate. In connecting electrolytic capacitors, attention must be paid to the polarity. The maximum DC voltage that the capacitor can withstand is mentioned as the working DC (WVDC). Capacitors have large tolerances (about 20%). Hence, the WVDC value must be taken safely above the maximum voltage that can appear across the capacitor (50% more than the maximum voltage is a good choice). Large electrolytic capacitors have appreciable series inductive components due to thick leads and wound plates to increase the capacitance to volume ratio. Thus, it may not behave as an effective capacitive element for high frequency spikes. This is usually corrected by adding a small parallel capacitor.

During charging interval, the current to the capacitor is limited by conduction resistance of the diode and wire resistance of the transformer. A small series resistance is added sometimes. This will cause a small drop at the output voltage, but improves the ripple factor considerably. It will limit the forward current; hence extend the life of diodes and transformer.

The charged capacitor retains some charge even after the supply switched off. This might damage some circuit components. A (bleeder) resistor (around 1 kΩ, 0.25 or 0.5 W) connected across discharges the capacitor in a few seconds. If a led indicator is connected, then there is no need for such a resistor.

Inductive Filters

Inductive filters have better control of the ripple for large load currents. The inductor behaves as a short circuit for the DC component. Hence, when $2fL \gg R_L$ the DC value of the output is approximately $2Vi/\pi$ and the ripple factor $r \approx 0.118R_l/fL$ where R_L is the effective load resistance, f is the frequency of the ripple and L is the inductance (in Henry). Figure 6.14 shows a symbolic diagram of an inductive filter.

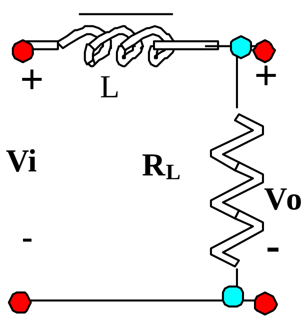


Figure 6.14 Inductive filter

With the inductive filter large current spikes do not hamper the transformer and rectifier as that occur in capacitive filters.

L-Section and π-Section Filters

Combination of capacitive and inductive elements is possible in form of L-section and π-section filters as shown in Figure 6.15. For the L-section filter the ripple factor is independent of the load and it is approximately *0.83/LC*. The DC value of the output is the same as that of the inductive filter.

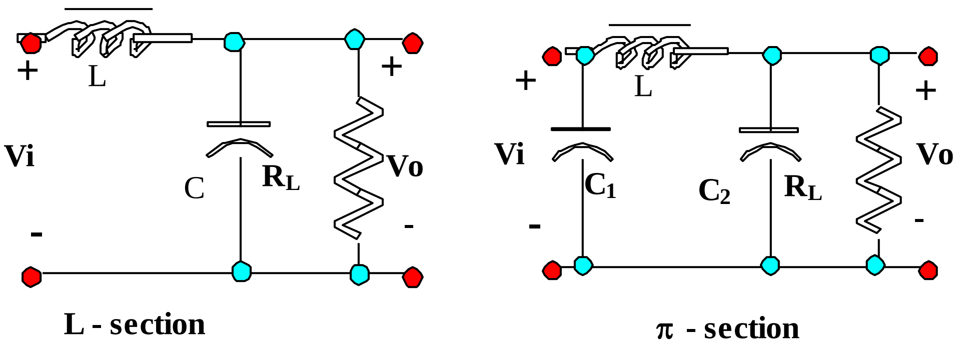


Figure 6.15 L and π section filters

For a π-section filter the DC value is the same as that of the capacitive filter. The ripple factor is inversely proportional to the product $C_1C_2LR_L$. Table 2 shows comparison of four filter types for 60 Hz ripple voltage.

Table 6.2 Comparison of four passive smoothing filters

Type of filter	Inductive	Capacitive	L-section	π-section
Ripple	$R_L/1600L$	$2400/(R_LC)$	0.83/(LC)	$3300/(C_1R_LC_2L)$
DC output volt	0.636Vm	Vm - 4200Idc/C	0.636Vm	Vm - 4200Idc/C

Linear (Dissipative)Regulators

Need for a Regulator

The output of unregulated supply contains an AC component that may cause interference to the

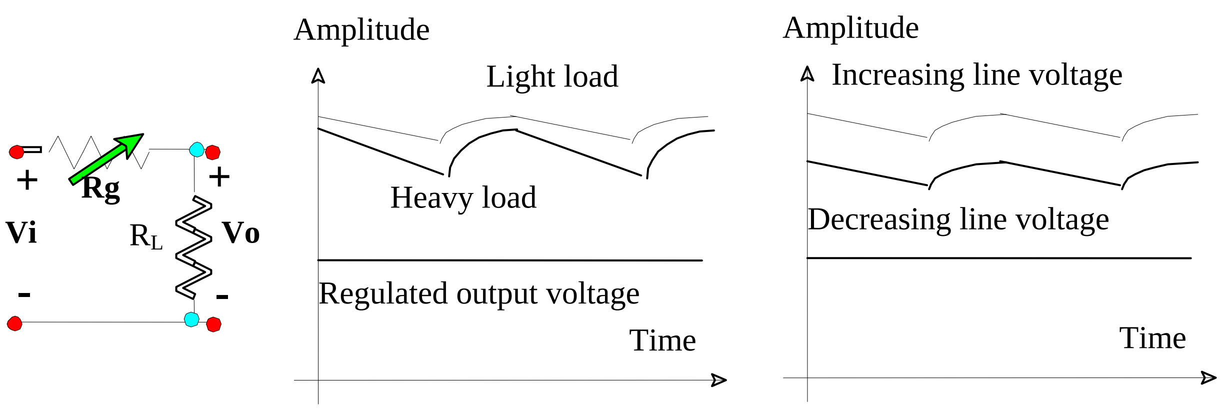


Figure 6.16 Function of the regulator as the unregulated input voltage fluctuates

electronic circuits. The average value of the voltage fluctuates as the load and/or the input line voltage changes as shown in Figure 6.16. The regulator behaves as a series variable resistor that changes in accordance with the load current, keeping the output voltage constant.

Many integrated circuits are developed to replace the discrete regulator circuits. Basic discrete circuits will be shown to illustrate the principle followed by an example of the IC regulator. A simplified diagram and output waveform of a regulated power supply is shown in Figure 6.17. The regulated DC output is very smooth with no ripple. It is suitable for all electronic circuits.

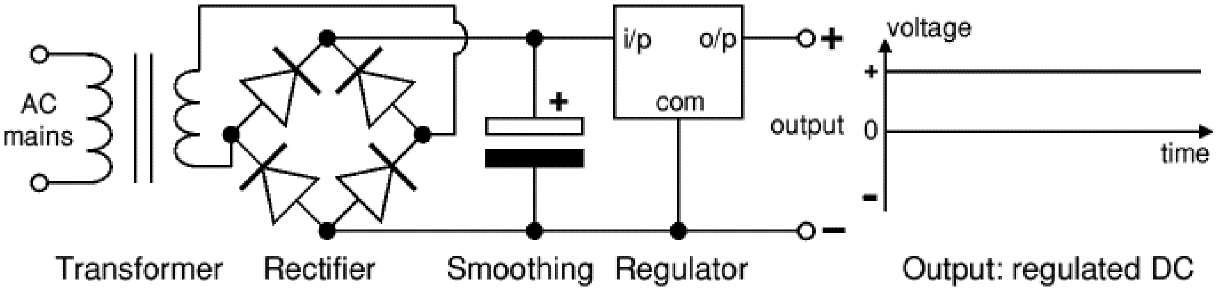


Figure 6.17 Simplified diagram of a regulated power supply

The Zener Diode Based Discrete Regulators

(This section is briefed mainly from http://www.electronics-tutorials.ws/diode/diode_7.html)

The DC output voltage from the half or full-wave rectifiers contains ripple superimposed onto the DC voltage. The load value changes causes the average output voltage to vary as well. The function of a regulator is to provide a constant output voltage to a load connected in parallel with it in spite of the ripples in the supply voltage or the variation in the load current. **Zener diodes** can be used to produce a stabilized voltage output with low ripple under these varying load current conditions. By passing a small current through the diode from a voltage source, via a suitable current limiting resistor (R_S), the zener diode will conduct sufficient current to maintain a voltage drop of *Vout*. Hence, by connecting a simple zener stabilizer circuit as shown in Figure 6.18 across the output of the rectifier, a more stable output voltage can be produced.

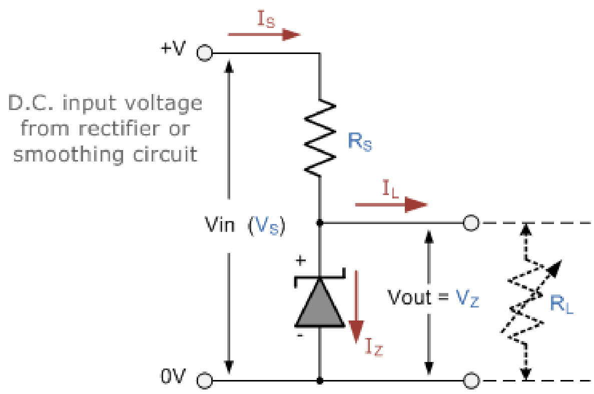


Figure 6.18 A zener diode based voltage regulator

The resistor, R_S is connected in series with the zener diode to limit the current flow through the diode with the voltage source, V_S being connected across the combination. The stabilized output voltage V_{out} is taken from across the zener diode. The zener diode is connected with its cathode terminal connected to the positive rail of the DC supply so it is reverse biased and will be operating in its breakdown condition. Resistor R_S is selected so to limit the maximum current flowing in the circuit. With no load connected to the circuit, the load current will be zero, (I_L = 0), and all the circuit current passes through the zener diode which in turn dissipates its maximum power. Also a small value of the series resistor R_S will result in a greater diode current when the load resistance R_L is connected and large as this will increase the power dissipation requirement of the diode so care must be taken when selecting the appropriate value of series resistance so that the zeners maximum power rating is not exceeded under this no-load or high-impedance condition. The load is connected in parallel with the zener diode, so the voltage across R_L is always the same as the zener voltage, ($V_R = V_Z$). There is a minimum zener current for which the stabilization of the voltage is effective and the zener current must stay above this value operating under load within its breakdown region at all times. The upper limit of current is of course dependent upon the power rating of the device. The supply voltage V_S must be greater than V_Z.

Example: Design a zener diode stabilized power supply that will provide an output voltage 5V at an output current of 60mA.

Steps in the design:

- Vz = 4.7V (nearest value available)
- Vs = 8V (it must be a few volts greater than Vz)
- Imax = 66mA (output current plus 10%)
- Pz > 4.7V × 66mA = 310mW, choose Pz = 400mW
- R = (8V - 4.7V) / 66mA = 0.05k = 50 , choose R = 47
- Resistor power rating P > (8V - 4.7V) × 66mA = 218mW, choose P = 0.5W

The simple voltage regulator based on the zener diode can be used if the load current is low and load is stable. General purpose voltage regulators can be designed inserting a common-base transistor in series with load and using the zener diode as a voltage reference. The transistor behaves as the variable resistor. There are several configurations available in the literature for such applications. However, zener diodes are very noisy especially operated around the avalanche region (for zener diodes with V_z>6 volts). The voltage drop across the zener varies with the input voltage causing slight variation of the output voltage. The zener diode, like all silicon devices, is effected by the temperature that causes a drift in the zener voltage. This can be compensated by complicated circuits.

Linear Regulator ICs

Problems of discrete regulators are solved by integrated circuit type linear regulators. There are many ICs with different rating available in the market. The most famous of them is the 7800 series three terminal positive voltage regulator shown in Figure 6.19. Input pin is connected to the unregulated supply voltage of the filters. The output pin delivers the regulated voltage. Current rating depends upon the package used. The plastic package can give up to 1 A and the metal package can safely supply up to 3 A. for larger currents, a current boosting transistor can be used. The central pin (the case in the metal one) is connected to ground for a fixed supply. This pin may be connected to ground through a zener diode to increase the output voltage. For a variable output voltage, the grounding terminal may be tied to the central pin of a potentiometer that is connected between the output and the ground. However, there are adjustable regulator ICs and they should be preferred instead for applications requiring variable output voltages.

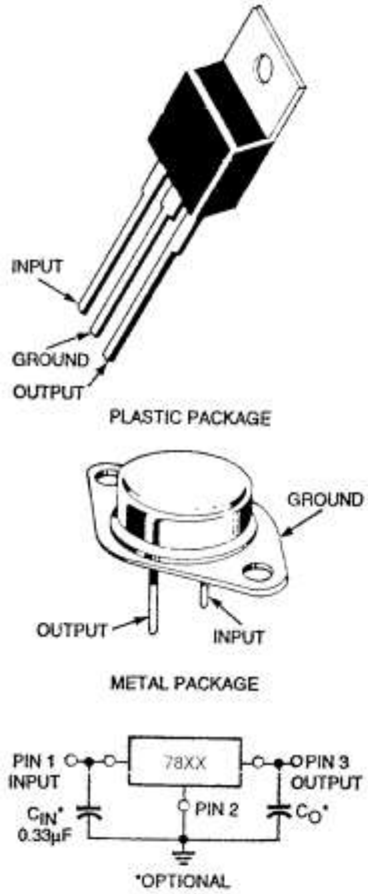


Figure 6.19 IC regulators

7900 series regulator ICs are the complementary of 7800 series to obtain negative regulated voltages.7800 and 7900 series are available with eight different output voltages; 5, 6, 8, 9, 12, 15, 18 and 24 volts. The output voltage appears as the suffix (i.e. 7806 for the 6-volt regulator). The input voltage is limited to 35 volts for 7805 to 7818 and 40 volts for 7812.

The minimum voltage drop across the regulator is about 2 volts. Hence, the input must be guaranteed to be at least 2 volts above the required output voltage. An input and output capacitor (value 0.22 to 1 µF) might be needed under certain conditions like the regulator is away from the filters and electronic circuits powered are away from the regulator.

Protection of Circuits in Case of Regulator Failure

Built-In Protection

7800 series regulators have built-in short circuit and over temperature protection. The chip shuts-down rather than blowing out to prevent the damage to the circuitry. However, if a boost transistor is driven by the chip to increase the current capability, then the transistor will see the full input voltage across without any limitation is the output current. Hence, an additional over-current (short circuit) protection becomes necessary.

The Over-Voltage Crowbar

The regulator circuit may not have the protection as above. Or a current boosting transistor may be used. Then, if the regulator fails and becomes short circuit, letting the full unregulated input voltage appearing across the load, damaging sensitive electronic components. A quick-blow fuse may be used at the output of the regulator to protect the circuits in case of excessive current. However, "the silicon fuse" may blow faster.

An over-voltage crowbars shown in Figure 6.20 may be added to provide the sufficient protection. A +5 V supply is shown as an example in the figure. TTL logic circuits require +5 V supply and they cannot tolerate more than +7 V without damage. The crowbar shown lets the thyristor (silicon controlled rectifier - SCR) to turn-on as the voltage goes over 6.5 V causing the fuse to blow due to excessive current drawn.

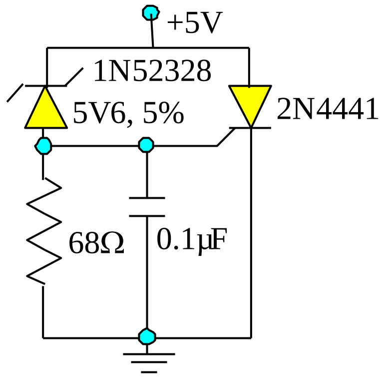


Figure 6.20 Over-voltage crowbar

SWITCH-REGULATED (SWITCHING) POWER SUPPLY

Linear Versus Switching

The linear regulator discussed above relies on receiving a power much higher than required from the source and dissipating some of it to keep the output voltage fixed immaterial of the current, provided it stays within the limits. It is cheap to install, but expansive in long run. It is mainly used for low power electronic devices either as a built-in unit or as a standalone unit. It best suits to applications where the output power varies considerably, like in laboratory power supplies.

Switching regulator chops the unregulated DC input voltage and provides the constant voltage required at the output by adjusting the chunks depending upon the demand from the load. It uses an inductor (choke) as an energy storage element. Regulation is not as good as the that of the linear type, but the efficiency is high. Expansive to install, but cheaper to run. It best suits to applications requiring high power and relatively constant power.

Principle of Operation

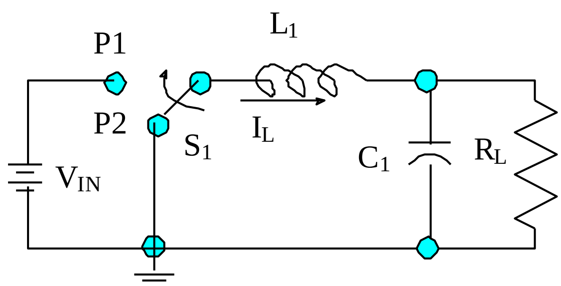


Figure 6.21 Elementary diagram for a switching power supply

The switching power supply relies on the switching regulator that is symbolically shown in Figure 6.21. Basically it is consists of a power source V_{IN}, duty cycle switch S_1 and an *LC* filter to provide constant output voltage across the load R_L.

Figure 6.22 shows typical current and switching waveforms. The switching transistor chops the DC input in such a way that it delivers constant volt-second energy pulses to the integrator. In this respect, the switching regulator is a power-controlled device and opposed to the linear regulator that is a current-controlled device. The integrator part of the regulator (the LC filter) smoothes out the pulsating DC. An inductor along with a capacitor stores sufficient electrical energy during the transistor on period to deliver to a regulated output voltage the load during the off period.

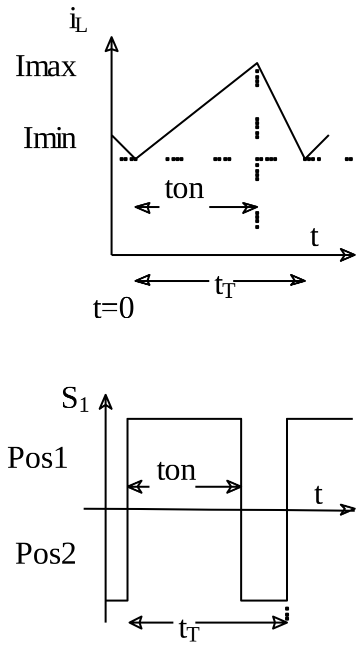


Figure 6.22 Waveforms

General Layout of the Switching Power Supply

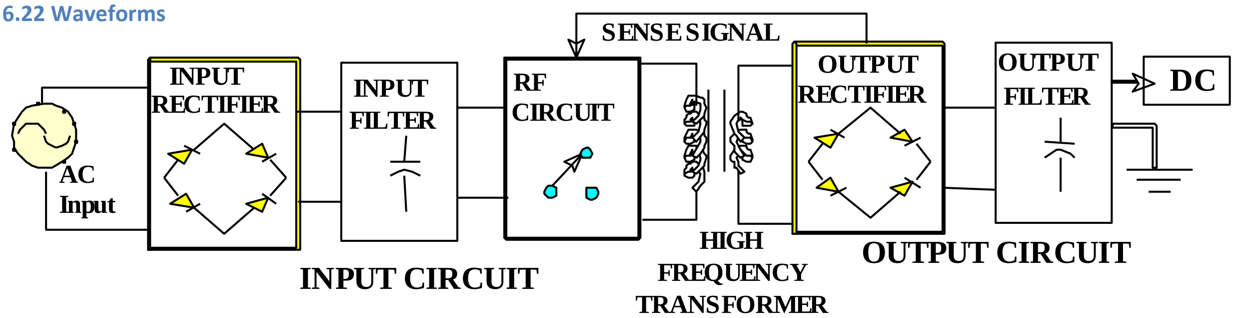


Figure 6.23 Functional block diagram of a switching power supply

Figure 6.23 shows block diagram of a complete switching power supply. It has:

- Input rectifier and filter that generates an unregulated DC from the power lines directly. At some low-power regulators, an input step-down transformer might be used. The input filter serves three purposes:
    - To smooth out spikes and high frequency transients with large peak values and small volt-second integrals.
    - To eliminate input ripple at the line frequency (50 Hz, 60 Hz or 400 Hz depending upon the application) for a half-wave rectified input and double the line frequency for a full-wave rectified input.
    - To attenuate AC components produced by transistor switching.
- A transistor switch that operates at high frequency (between 20 kHz and 1 MHz) chops the input DC.

- A high-frequency transformer steps down the chopped signal to the desired level.
- The output rectifier converts the signal from the transformer into unregulated DC and the output filter smoothes out the output.

The transformer and output rectifier are not necessary if the input voltage is at the same level as the required output voltage. The output is sensed and used to control the switching (on) time of the transistor.

Rectifiers and Filters of a Switching Power Supply

The Input Rectifier

It is similar to those used in the linear power supply. However, the input in this case is the line voltage directly. Thus, great care must be taken in handling the input components due to large voltage involved.

The bridge rectifier is used in almost all applications. It develops its own ground reference and isolates the rest from the AC line. In choosing the proper elements, the peak inverse voltage must be at least 50% larger than the maximum peak voltage at the input, and the forward current must be 2 to 5 times the average current required.

A small resistor or a thermistor connected between the bridge and the filter capacitor**Error! Bookmark not defined.** reduces surge currents that exist due to high frequency switching at peak line voltage.

Output Rectifiers

All three rectifier configurations discussed for the linear regulated supplies, half-wave (Figure 6.24), full wave with a center-tapped transformer (Figure 6.25) and full-wave with a bridge rectifier (Figure 6.26) are used.

The half-wave is the simplest form but not very effective. A full-wave rectifier either with a

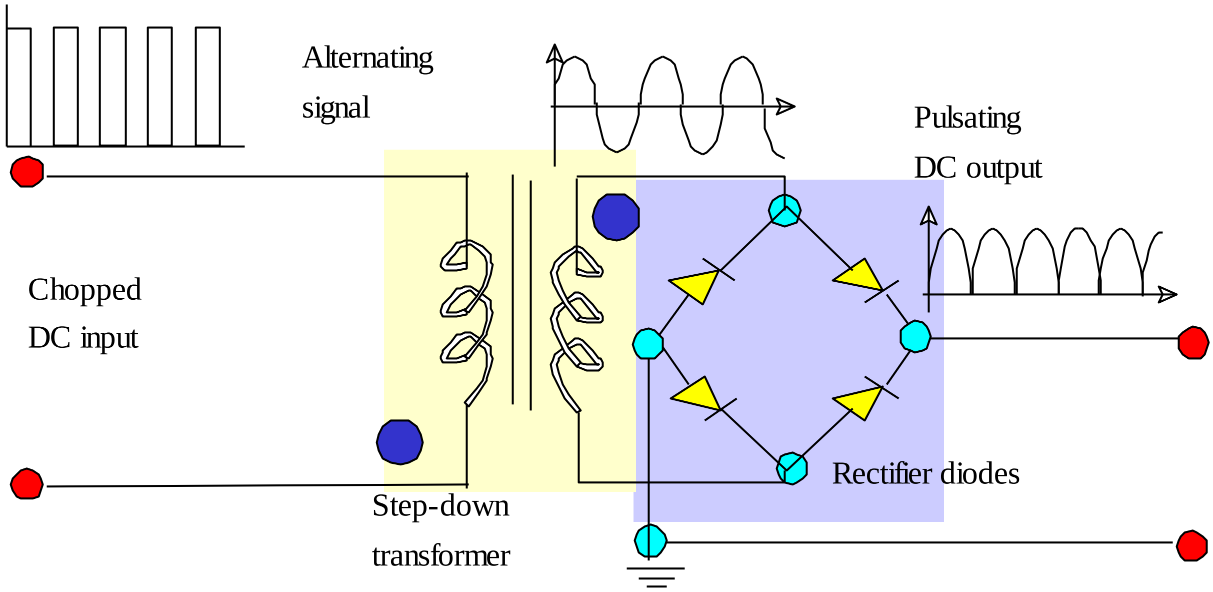


Figure 6.24 The full-wave with bridge rectifier type output rectifier

center tapped-transformer or with a bridge rectifier is mostly used. Figure 6.24 illustrates the full-wave with bridge rectifier version.

High-frequency rectifiers are needed. They represent largest single source of generated heat in a power supply. Schottky rectifiers and fast-recovery diodes are used. Schottky rectifiers are based on a metal-to-silicon junction called the Schottky barrier and they are the faster of the two types. They have small junction capacitances leading to smaller recovery times. Fast-recovery diodes are also divided into several categories and they approach to the Schottky diodes in terms of the recovery times.

Filters

They are similar to those used in linear regulators are utilized both for the input and output. Input filters involve capacitors between 1000 and 2200 µF (sometimes up to 5000 µF). Output filters may have capacitance up to 470 µF. Working DC voltage rating (WVDC) of the input filter capacitors must be about 150% of the peak voltage that may appear at the output of the input rectifier.

Capacitors have been designed to have higher capacitance to volume ratio, small equivalent series resistance (ESR) and series inductance for more effective operation at high frequencies. Aluminum electrolytic capacitors are used at the input filtering. It is preferable to place a tantalum or other low value capacitor with much smaller ESR in parallel. This second capacitor is generally placed close to the collector of the switching transistor. Multi-layer ceramic capacitors are used for output filtering at high frequencies. Electrolytic capacitors can also be used if the frequency of operation is low. High frequency operation requires smaller capacitor size.

Elements of the RF Regulator/Switching Network

The heart of every switching regulator is the RF regulator network shown in Figure 6.25. It chops the DC voltage from the input filter at 20 kHz or higher (up to 1 MHz is considered in recent designs). Pulse-Width-Modulation (PWM) shown in the figure is mostly used to drive the switching transistor for chopping. Pulse width varies according to the load (closed-loop control system). Basic components of the system involves the switching element, high frequency step-down transformer, output rectifier and filter discussed above, and sense amplifier and modulator.

The Switching Element

Power MOSFETs are mostly preferred over bipolar junction transistors. Power MOSFETs have the following major advantages:

- Can be driven directly by control ICs without a need for a drive circuitry.

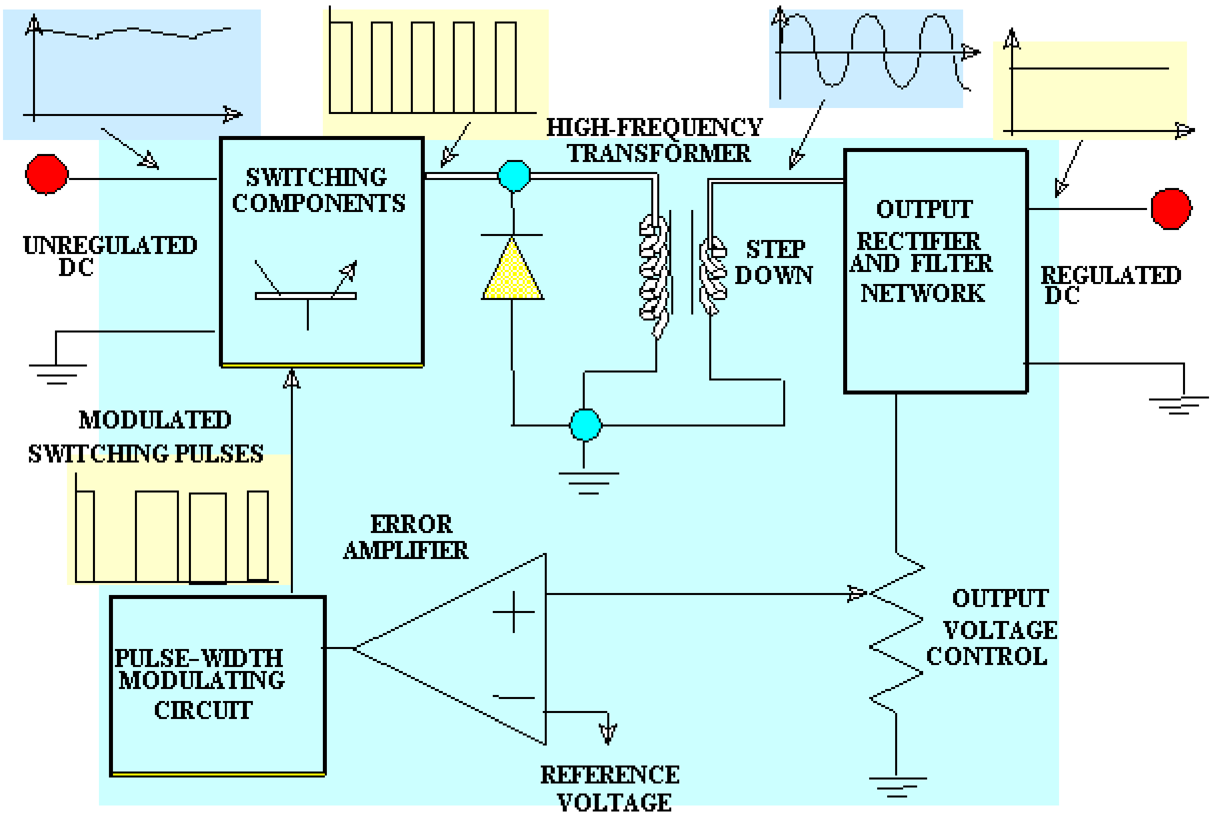


Figure 6.25 Block diagram of the switching network

- They don't store charge during saturation. Hence, they have very low transition time that allows them to work at high switching frequencies.

- They don't have destructive secondary break-down reducing or even eliminating the need for a speed limiting snubber network.

However, they have some disadvantages as:

- Large on resistance (4-5 Ω versus 0.1 Ω in bipolar).
- Sensitivity to reverse voltage spikes and,
- large die size.

In recent years, bipolar transistors have been developed that can switch amperes of currents in 2 μs or less and withstand voltage over 1000 volts.

The High-Frequency Transformer

A transformer is used to convert high-voltage, chopped DC into a lower voltage secondary AC signal. It must operate at the switching frequency of 20 kHz or higher. Although it uses the same principle of magnetic coupling as the transformer operating at line frequency (50 Hz, 60 Hz or 400 Hz depending upon the place of application), ordinary transformer will not work at high frequencies.

For switching supply applications toroidal transformers in which turns of wires wrapped around toroidal coils are used in medium to high power levels, where they are cost effective. At low power levels, ferrite E-cores are commonly used. Many ferrite materials work well at 100 kHz, but they fail at higher frequencies. Special core materials are developed for high-frequency operations. At high frequencies, proximity and skin effects in magnetic windings become dominant that limit the amount of copper that can be used. Litz wire (twisted bundle of fine wires), foil, and printed conductors are used to reduce losses.

The Regulator

There are three basic types of regulators as the Ferro resonant supply, pulse-frequency modulation, and pulse-width modulation.

The Ferro resonant supply is the simplest and most reliable one. It is composed of a Ferro resonant transformer, a resonating capacitor, and a rectifier and an output filter. No electronic regulation circuitry is involved and the regulation is achieved within the transformer core through a magnetic process. It is used in many industrial and commercial devices like microwave oven, but rarely appears in electronic applications.

The pulse frequency modulation reduces the duty-cycle by manipulating the interval between pulses, not the width of the pulses. It responds more closely changes in the load. Thus, the efficiency rises. It is very effective with high frequencies and light loads. Although the lower operating frequency is used, the longer pulse intervals causes filtering problems. The pulse width modulation (PWM) is the widely used approach. There are ICs manufactured for this purpose.

Switching Regulator Configurations

There are three basic configurations from which all others are driven as the buck (step-down) converter, the boost (step-up) converter, and the buck-boost (step-down / step-up or inverting) converter. Only the buck (step-down) converter will be summarized below. Interested readers are referred to the references for details of the buck converter and other switching regulators.

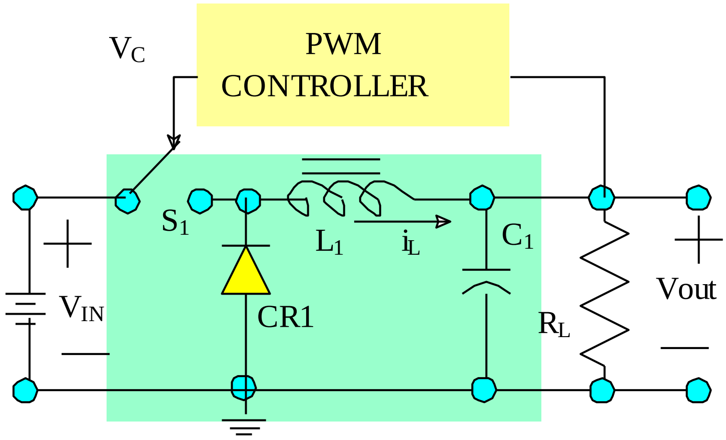


Figure 6.26 The buck (step-down) converter

Figure 6.26 shows the basic buck converter topology. The circuit interrupts the line and provides a variable pulse width rectangular wave to simple averaging filter L_1-C_1 such that the applied voltage is either V_{in} or 0.

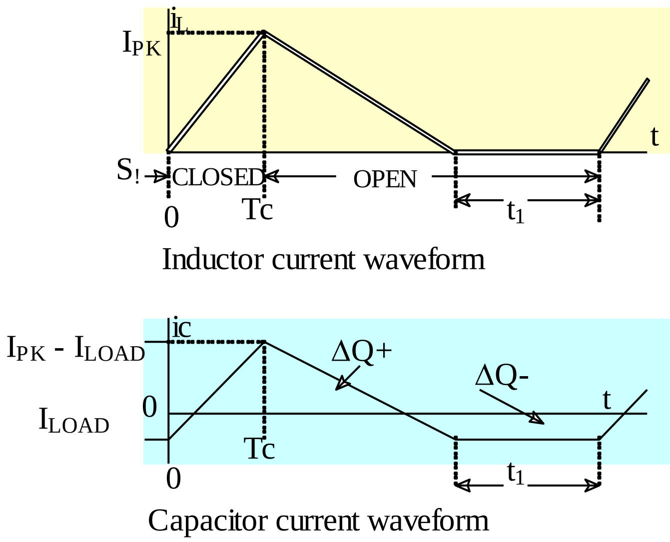


Figure 6.27 Inductor and capacitor waveform

When S_1 is closed, the diode CR1 is off (reversed biased) and when S_1 opens, the current through L_1 forces the diode to turn on. Figure 6.27 demonstrates typical inductor and capacitor current waveforms. The current i_L at any given time (t) is

$I = (V_{in} - V_{out})*t/L_1$ yielding $I_{pk} = (V_{in} - V_{out})*t_{on}/L_1$.

The duty cycle of the converter is $D= t_{on}/T = t_{on}/(t_{on}+t_{off})$

The output voltage V_{out} can be expressed in terms of the input voltage V_{in} and duty cycle D as $V_{out} = V_{in}D$.

L_1-C_1 combination behaves as a low-pass filter. For the output to remain constant, the net charge delivered to the filter capacitor must be zero. This means, the charge delivered to the capacitor from the inductor must be dissipated in the load. The charge developed in the inductor is fixed (constant on time) and the time required to dissipate it must vary according to the load conditions.

The figure shows the discontinuous operation, since the inductor current becomes *0* in certain period of the cycle. As the load continuously increased, a DC idle current will pass through the inductor and this is called the continuous mode of operation. In this mode, I_L never equals *0* and $t_1=0$.

The input current can be found as

$$I_{in}= (I_{out}*V_{out})/(\eta*V_{in})$$

where η is the efficiency of the regulator. The minimum achievable ripple voltage

$$V_{ripple}(min) = I_{PK}*(ESR)$$

where ESR is the series equivalent resistance of the filter capacitor. ($I_{pk} = (V_{in} - V_{out})*t_{on}/L_1$

The buck converter is the basis for many types of transformer coupled DC/DC converters.

Overall Look Into Advantages and Disadvantages of Switching Supplies

Some of the advantages and disadvantages of switching circuits are summarized in Table 3. They are far from ideal and present many problems. However, as the problems are identified correctly, it is possible to minimize their effects.

RF interference

Most of the advantages stated in the table are due to the presence of the switching transistor. However, in order to achieve that advantage, the input DC (unregulated) is chopped at a frequency above 20 kHz. Some current designs operate close to 500 kHz and in near future, up to 1 MHz will be available. Hence, the operating frequency falls within the RF (radio-frequency) spectrum. As a result each conductor in the high-frequency portion of the supply behaves as an antenna that transmits those frequencies to rather long distances. This causes interference to power supplies own circuitry, neighboring sensitive electronic instruments and circuits.

There are many techniques now available to eliminate the effects of the RF noise including:

- Careful grounding and shielding of switching components and outer case.
- Using well shielded interconnecting cables with the shield being the common-ground to the supply circuit.

- Using electronic filtering components, such as capacitors and inductors in the design to suppress the RF emission.
- Changing physical orientation and position of components in the supply, as well as location of the supply itself.

System Dynamics

Compared to its linear counterparts, the ability of a switching supply to adjust the output voltage continually under varying loading conditions is not as good. It is essential to have a minimum load to operate and it does not work under no load conditions. It is also slow in responding to transient changes at the output (load).

Table 6.3 Comparison of linear and switching mode power supplies

Parameter	Linear Supply	Switching supply
Efficiency	30 to 50%	60 to 80%
RF noise	Usually negligible	Can be problem unless shielded
Transformers	Requires bulky 60 Hz magnetics	Smaller, lighter. high-frequency magnetic
Ripple	1 to 5 mV peak to peak	10 to 40 mV peak to peak
Regulation	0.05 to 0.1% ($V_{Full\ Load}$)	0.3 to 1% ($V_{Full\ Load}$)
Power/Weight Ratio	14 Watts/kg (average)	7 Watts/kg (average)
Temperature Rise	50 to 100°C above ambient	20 to 40°C above ambient
Reliability	Runs much hotter and can degrade reliability	Cooler operation improves the reliability

Supply Service Precautions

Be Careful of High Voltage

Use extreme caution in taking measurements. Always unplug supply and allow sufficient time for large electrolytic capacitors to discharge. It is also good practice to discharge them manually.

Watch Out For Shielding

Replace and re-solder any shielding and resecure all grounds before operating the serviced supply.

Replacement Parts

Use only exact replacement parts. Otherwise, the switching frequency may shift causing an increased RF interference. Use the same type of components. For example if you should replace a tantalum capacitor, replace it with tantalum of the same value, not with an aluminum electrolytic capacitor.

Unless proper tools and instruments are available do not attempt to play with calibration adjustments. An improper adjustment may degrade the supply just as much as the use of an improper component.

Summary of Key Formulas that Help in Solving Power Supply Problem

$$Efficiency(\eta) = \frac{P_{OUT}}{P_{IN}} = \frac{V_O \, x \, I_L}{V_{IN} \, x \, I_{IN}};$$

$$Input\ regulation(\% / V_{IN}) = \frac{\Delta V_O}{\Delta V_{IN} \bullet V_O} x\,100;$$

$$Load\ regulation(\%) = \frac{V_{ML} - V_{FL}}{V_{\otimes}} 100 = \frac{R_0}{R_L} x\,100$$

$Smoothing\ capacitor\ for\ 10\%\ ripple, C = \frac{5xIo}{V_s x f}$, V_s is the peak value of the input

The ripple factor and the DC output voltage can be estimated by *r=2400/R_LC* and *Vdc=(Vi - 4200Idc/C)* where C is in µF and frequency is 60 Hz.

BATTERIES

Principles of Operation

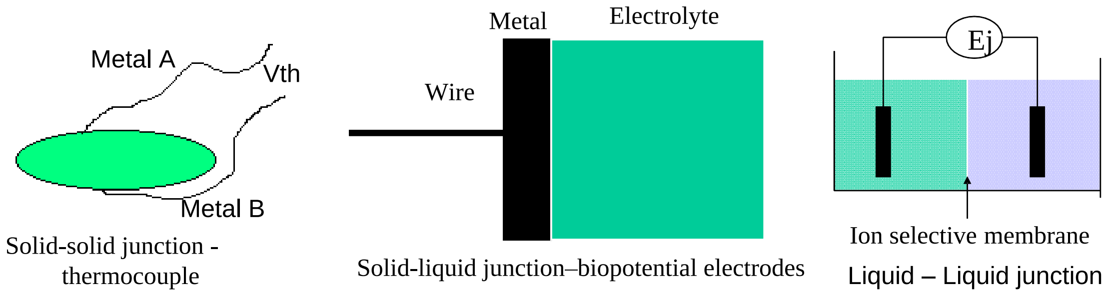


Figure 6.28 Junctions of dissimilar materials and junction potentials

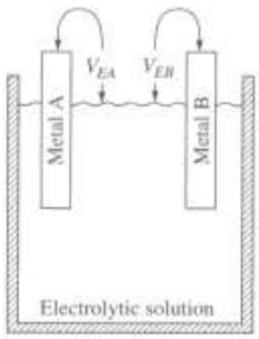


Figure 6.29 Electrochemical cells

Dissimilar materials can be brought together through a junction as shown in Figure 6.28 and a potential difference is established across this junction. The solid to solid junction is called the thermocouple that will be discussed in a special section. The solid to liquid junction appears in biopotential electrodes. Another similar junction to measure the potential as illustrated in Figure 6.29 . Hence, the solid to liquid junction potential is called the half-cell potential. Liquid to liquid junction is established by having two aqueous ionic solutions of different concentrations separated by an ion-selective semipermiable membrane.

Batteries are power sources for all portable electronic devices and electrical devices in remote areas. They are highly engineered electrochemical cells that convert chemical energy to electrical energy using three major materials: the anode (negative electrode), the cathode (positive electrode) , and the electrolyte. How these materials get picked for the job depends on how well they give up or attract electrons, something that must happen for an electric current to be generated. The anode is often a metal, the cathode is a metallic oxide and the electrolyte is the electricity conductor. The battery is one or more electrochemical cells that converts chemical energy directly to electrical energy.

The cell is the smallest unit based on chemical reactions. The cell voltage depends upon the electrode materials, electrolyte and its concentration and temperature. The current that can be supplied depends upon the internal resistance of the cell. Some cells use two half-cells with different electrolytes. A separator between half cells allows ions to flow, but prevents mixing of the electrolytes as shown in Figure 6.30. The voltage can be increased by adding cells in series and the current capacity can be increased by adding cells in parallel. Batteries are the multiple-cell entities.

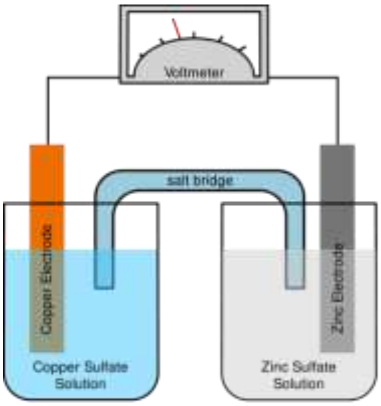


Figure 6.30 Two half-cells with two electrolytes

The electrical driving force or ΔV_{bat} across the terminals of a cell is known as the *terminal voltage (difference)* and is measured in volts. The terminal voltage of a cell that is neither charging nor discharging is called the open-circuit voltage and equals the emf of the cell. Because of internal resistance, the terminal voltage of a cell that is discharging is smaller in magnitude than the open-circuit voltage and the terminal voltage of a cell that is charging exceeds the open-circuit voltage. An ideal cell has negligible internal resistance, so it would maintain a constant terminal voltage until exhausted, then dropping to zero. If such a cell maintained 1.5 volts and stored a charge of one coulomb then on complete discharge it would perform 1.5 joule of work. In actual cells, the internal resistance increases under discharge, and the open circuit voltage also decreases under discharge. If the voltage and resistance are plotted against time, the resulting graphs typically are a curve; the shape of the curve varies according to the chemistry and internal arrangement employed.

Categories and Types

There are two types of batteries: primary batteries (disposable batteries), which are designed to be used once and discarded, and secondary batteries (rechargeable batteries), which are designed to be recharged and used multiple times. *Primary* batteries irreversibly (within limits of practicality) transform chemical energy to electrical energy. When the initial supply of reactants is exhausted, energy cannot be readily restored to the battery by electrical means. *Secondary* batteries can be recharged; that is, they can have their chemical reactions reversed by supplying electrical energy to the cell, restoring their original composition.

Primary Batteries

Primary batteries can produce current immediately on assembly. Disposable batteries are intended to be used once and discarded. These are most commonly used in portable devices that have low current drain, are only used intermittently, or are used well away from an alternative power source, such as in alarm and communication circuits where other electric power is only intermittently available. Disposable primary cells cannot be reliably recharged, since the chemical reactions are not easily reversible and active materials may not return to their original forms. Battery manufacturers recommend against attempting to recharge primary cells.

Common types of disposable batteries include zinc-carbon LeClanche, zinc chloride (heavy duty), zinc air, alkaline, mercury oxide, silver oxide and lithium batteries. Generally, these have higher energy densities than rechargeable batteries, but disposable batteries do not fare well under high-drain applications with loads under 75 Ω.

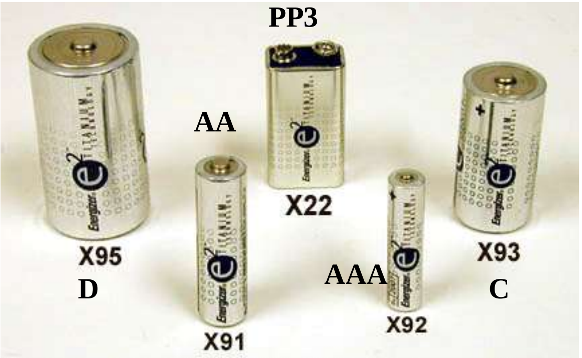


Figure 6.31 Commonly available sizes of batteries

Commonly available sizes are shown in Figure 6. 31 and descriptions of alkaline types are listed in Table 6.4. In addition, miniature cells are used to power devices such as hearing aids and wristwatches; larger batteries provide standby power for telephone exchanges or computer data centers. Mostly used primary batteries are the carbon zinc (or zinc chloride – heavy duty) and alkaline types. The alkaline batteries have several advantages over the zinc based ones as:

- Better discharge rate capability
- Lower and more stable internal resistance
- Better low temperature performance
- Better service maintenance
- Higher energy density
- More economical than Carbon Zinc in terms of cost per hour of use on high current drains
- Sloping discharge curve
- Relatively insensitive to changes in the discharge rate or duty cycle
- Available in voltages ranging from 1.5 to 12.0 and in a variety of shapes and sizes (commonly available one are shown in Figure 6.31).

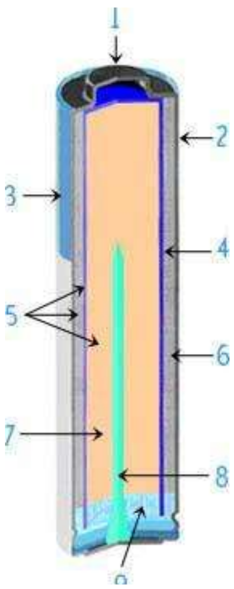


Figure 6.32 Anatomy of an alkaline battery

The anatomy of the alkaline battery is illustrated in Figure 6.32. It contains:

- Positive Pip: A formed protrusion in the bottom of the battery can which identifies it as the positive terminal.
- Steel Can: Nickel-plated steel which is formed into a container to hold chemicals; serves as the positive collector.

- Outer Jacket: A plastic sleeve which contains decorative printing identifying the cell type and size.
- Separator: Porous non-woven fibrous material which separates electrodes; holds electrolyte between electrodes.
- Electrolyte: A solution of potassium hydroxide in water which carries the ionic current inside the battery.
- Cathode: Manganese dioxide and graphite which take up electrons from the external circuits.
- Anode: Powdered zinc metal which serves as the source of electrons.
- Anode Collector: Tin-plated brass which serves as a path for the electrons from the anode to the external circuit.
- Seal/Vent: Molded plastic disc which holds internal components inside the cell and releases internal pressure when battery is abused.

Table 6.4 Information for commonly available alkaline batteries

Name	Size	Capacity * (mAh)	Voltage (nom.)	ANSI/ NEDA	IEC	Weight (g)	Diam. (max mm)	Height (max mm)	Length (max mm)	Width (max mm)
X22	9V	595	9	1604A	6LR61	45.6	N/A	48.5	26.5	17.5
X91	AA	2850	1.5	15A	LR6	23	14.5	50.5	N/A	N/A
X92	AAA	1150	1.5	24A	LR03	11.5	10.5	44.5	N/A	N/A
X93	C	8350	1.5	14A	LR14	66.2	26.2	50	N/A	N/A
X95	D	18000	1.5	13A	LR20	141.9	34.2	61.5	N/A	N/A

Secondary Batteries

Rechargeable batteries or *secondary cells* can be recharged by applying electric current, which reverses the chemical reactions that occur during its use. They must be charged before use; they are usually assembled with active materials in the discharged state. Devices to supply the appropriate current are called chargers or rechargers.

The oldest form of rechargeable battery is the lead-acid battery that contains a liquid in an unsealed container. However it is required that the battery be kept upright and the area be well ventilated to ensure safe dispersal of the hydrogen gas produced by these batteries during overcharging. The lead-acid battery is also very heavy for the amount of electrical energy it can supply. Despite this, its low manufacturing cost and its high surge current levels make its use common where a large capacity (over approximately 10A-H) is required or where the weight and ease of handling are not concerns.

A common form of the lead-acid battery is the modern car battery, which can generally deliver a peak current of 450 amperes. An improved type of liquid electrolyte battery is the sealed valve regulated lead acid (VRLA) battery, popular in the automotive industry as a replacement for the lead-acid wet cell. The VRLA battery uses an immobilized sulfuric acid electrolyte, reducing the chance of leakage and extending shelf life. VRLA batteries have the electrolyte immobilized, usually by way of a semi-solid electrolyte (called the gel cell) or absorbing the electrolyte in a special fiberglass matting (called the absorbed glass mat – AGM).

Other portable rechargeable batteries include several "dry cell" types, which are sealed units and are therefore useful in appliances such as mobile phones and laptop computers. Cells of this type (in order of increasing power density and cost) include nickel-cadmium (NiCd), nickel-zinc (NiZn), nickel metal hydride (NiMH) and lithium-ion (Li-ion) cells. By far, Li-ion has the highest share of the dry cell rechargeable market. Meanwhile, NiMH has replaced NiCd in most applications due to its higher capacity, but NiCd remains in use in power tools, two-way radios, and medical equipment. NiZn is a new technology that is not yet well established commercially.

Battery Capacity

The voltage developed across a cell's terminals depends on the energy release of the chemical reactions of its electrodes and electrolyte. Alkaline and carbon-zinc cells have different chemistries but approximately the same emf of 1.5 volts; likewise NiCd and NiMH cells have different chemistries, but approximately the same Nominal cell voltage (emf) of 1.2 volts at full charge 1.4 V for a fresh cell at immediate turn-on. On the other hand the high electrochemical potential changes in the reactions of lithium compounds give lithium cells emfs of 3 volts or more.

Because of the chemical reactions within the cells, the capacity of a battery depends on the discharge conditions such as the magnitude of the current (which may vary with time), the allowable terminal voltage of the battery, temperature and other factors. The available capacity of a battery depends upon the rate at which it is discharged. If a battery is discharged at a relatively high rate, the available capacity will be lower than expected. The battery capacity that battery manufacturers print on a battery is usually the product of 20 hours multiplied by the maximum constant current that a new battery can supply for 20 hours at 68 F° (20 C°), down to a predetermined terminal voltage per cell. A battery rated at 100 A-H

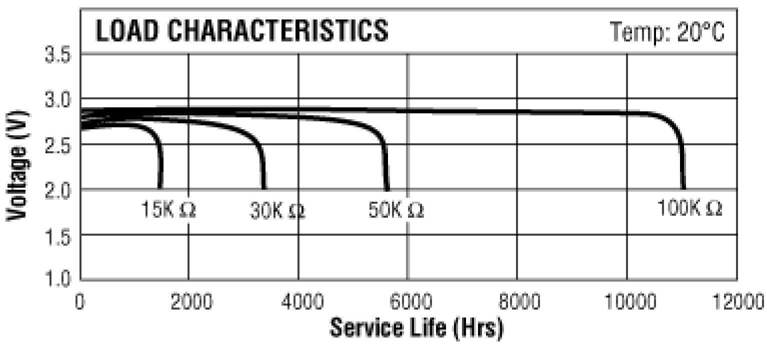


Figure 6.33 Load characteristics of a battery

will deliver 5 A over a 20 hour period at room temperature. However, if it is instead discharged at 50 A, it will have a lower apparent capacity. A typical load characteristic is shown in Figure 6.33.

Definitions of Different Drain Conditions

The drain conditions for a battery can be roughly defined as heavy, moderate and light drain.

- **Heavy drain** is defined as current that would discharge the battery within one day at room temperature.
- **Moderate drain** is defined as a current that would discharge the battery in approximately one week at room temperature.
- **Light drain** is defined as a current that would discharge the battery after one month or more at room temperature.

Life of Primary Batteries

Even if never taken out of the original package, disposable (or "primary") batteries can lose 8 to 20 percent of their original charge every year at a temperature of about 20°–30°C. This is known as the "self discharge" rate and is due to non-current-producing "side" chemical reactions, which occur within the cell even if no load is applied to it. The rate of the side reactions is reduced if the batteries are stored at low temperature, although some batteries can be damaged by freezing. High or low temperatures may reduce battery performance as illustrated in Figure 6.34. This will affect the initial voltage of the battery. For an AA alkaline battery this initial voltage is approximately normally distributed around 1.6 volts. An alkaline battery can be used down to 0.9 V.

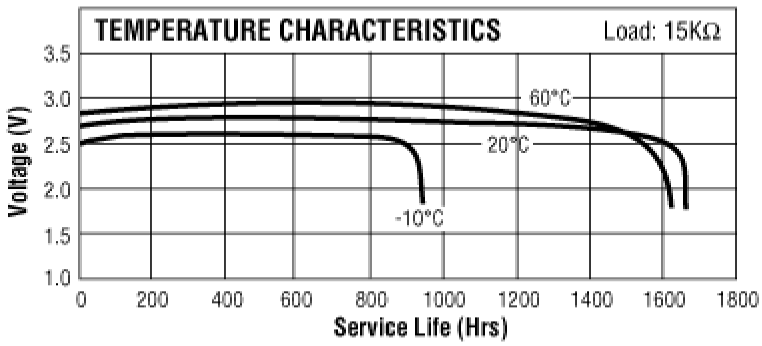


Figure 6.34 Effect of temperature on battery performance

The performance of a battery and eventually the battery voltage depends upon the load and temperature is shown in Figure 6.35. The figure illustrates the battery voltage at 50% discharged state against the load

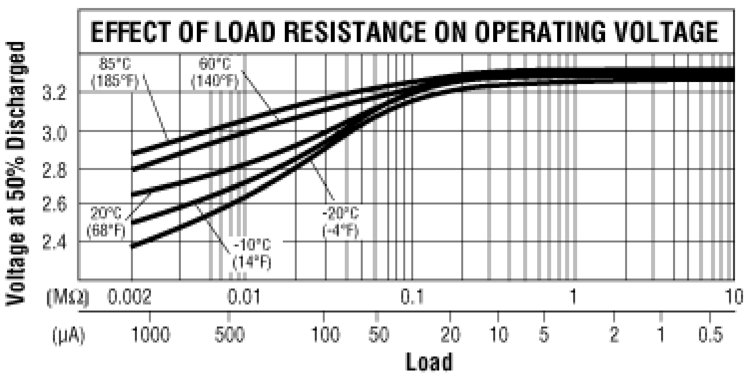


Figure 6.35 Effect of load resistance on operation voltage at 50% discharged

at various operating temperatures. At increased temperature, the voltage is higher under the same loading conditions. The effect of temperature is not apparent under low load conditions.

Life Span of Secondary Batteries

Old chemistry rechargeable batteries self-discharge more rapidly than disposable alkaline batteries, especially nickel-based batteries; a freshly charged NiCd loses 10% of its charge in the first 24 hours, and thereafter discharges at a rate of about 10% a month. However, NiMH newer chemistry and modern lithium designs have reduced the self-discharge rate to a relatively low level (but still poorer than for primary batteries). Most nickel-based batteries are partially discharged when purchased, and must be charged before first use. Newer NiMH batteries are ready to be used when purchased, and have only 15% discharge in a year.

Although rechargeable batteries have their energy content restored by charging, some deterioration occurs on each charge/discharge cycle. Low-capacity nickel metal hydride (NiMH) batteries (1700-2000 mA-H) can be charged for about 1000 cycles, whereas high capacity NiMH batteries (above 2500 mA-H) can be charged for about 500 cycles. Nickel cadmium (NiCd) batteries can sustain 1,000 charge – discharge cycles before their internal resistance permanently increases beyond usable values. They are unusable when the capacity drops below 80% of its nominal value. The amount time battery lasts is a function of discharge time. Normally a fast charge, rather than a slow overnight charge, will shorten battery lifespan. However, if the overnight charger is not "smart" and cannot detect when the battery is fully charged, then overcharging is likely, which also damages the battery. Degradation usually occurs because electrolyte migrates away from the electrodes or because active material falls off the electrodes.

NiCd batteries suffer the drawback that they should be fully discharged before recharge. Without full discharge, crystals may build up on the electrodes, thus decreasing the active surface area and increasing internal resistance. This decreases battery capacity and causes the "memory effect". These electrode crystals can also penetrate the electrolyte separator, thereby causing shorts. NiMH, although similar in chemistry, does not suffer from memory effect to quite this extent. When a battery reaches the end of its lifetime, it will not suddenly lose all of its capacity; rather, its capacity will gradually decrease.

The lead-acid cell is the most common form of storage battery. The positive electrode is lead peroxide; spongy lead is the negative electrode. Both are in a dilute solution of sulfuric acid as the electrolyte. The voltage output is approximately 2.1 V.

Lead-acid batteries are used for mobile (e.g. ambulance) and some high-power portable applications. The main benefit of the lead-acid battery is its low cost, easy availability and reliability. The main drawbacks are its large size and weight for a given capacity and voltage. Nominal voltage for automobile applications is 13.6 V DC. Many radio communication sets are also designed to operate from 13.6 V DC (also called 12-V). 6, 24, 28 and 32 Volt batteries are also available. Lead-acid batteries should never be discharged to below 20% of their full capacity, because internal resistance will cause heat and damage when they are recharged.

The relationship between current, discharge time, and capacity for a lead acid battery is approximated (over a certain range of current values) by Peukert's law:

$$t = \frac{Q_P}{I^k}$$

where

Q_P is the capacity when discharged at a rate of 1 amp.

I is the current drawn from battery (A).

t is the amount of time (in hours) that a battery can sustain.

k is a constant around 1.3.

For low values of I internal self-discharge must be included. Terminal voltage can be increased by connecting in series, while the current availability can be increased by connecting batteries in parallel.

Automotive lead-acid rechargeable batteries have a much harder life. Because of vibration, shock, heat, cold, and sulfation of their lead plates, few automotive batteries last beyond six years of regular use. Automotive starting batteries have many thin plates to provide as much current as possible in a reasonably small package. In general, the thicker the plates, the longer the life of the battery. Typically they are only drained a small amount before recharge. Care should be taken to avoid deep discharging a starting battery, since each charge and discharge cycle causes active material to be shed from the plates.

Battery Testing

The open circuit voltage (OCV) yields a rough estimate of the freshness of the battery and can be used to determine the amount of service life of a battery. However, the closed circuit voltage (CCV) is a better measure. This is accomplished by putting the battery under load for one to two seconds and measuring the CCV. If the battery voltage is greater than or equal to 1.1 volts, the battery has approximately 20% service left. The load is determined by the size and type of battery. In the case of a single cylindrical 1.5 volt Alkaline or Carbon Zinc battery, the load would be approximately 8 ohms. Otherwise, an OCV reading of 1.5 volts or greater for a single cylindrical 1.5 volt Alkaline or Carbon Zinc battery indicates essentially an undischarged battery or one that has been discharged less than 10%.

Care and Maintenance of Batteries

Battery Charging Protocols

Charging current that is less than 5% of the A-H rating of the battery will not be effective. Hence, the charging current I > A-H/20. It is safe to use I = A-H/10. A battery can charge up to 140% of the capacity; i.e. we can charge 14 hours at I = 0.1*A-H. Do not use charging current over A-H/10 unless specifically instructed by the battery manufacturer. A battery loses energy from merely sitting and it will be kept alive by a *trickle charge* at rate A-H/50 < I < A-H/30.

Periodic charging and discharging of batteries is essential. A battery or cell shall be charged fully and discharged fully with a resistor that draws a current of A-H/10 for 8 to 9 hours for multicell batteries and 10 hours for a single cell. Then, it must be recharged at the A-H/10 rate for 14 to 16 hours. Polarity reversal can occur in multicell batteries and the battery shall discharge only 10 to 20 % of capacity. Another problem with the batteries is the dendrite growth especially after leaving it discharged for a long time. These batteries can be revitalized by temporarily connecting them to a fully charged battery as illustrated in Figure 6.36. By pressing the pushbutton or a spring loaded switch, the high current in the circuit vaporizes the internal dendrites that shorts the plates together. We must be careful of explosion! And use safety goggles. We can't rely on revitalized ones and we must replace them as soon as possible. To charge a lead-acid battery, connect it to a dc voltage equal to approximately 2.5 V per cell. Connecting the positive terminal of the battery to the positive side of the charging source and the negative terminal to the negative side results in charging current through the battery.

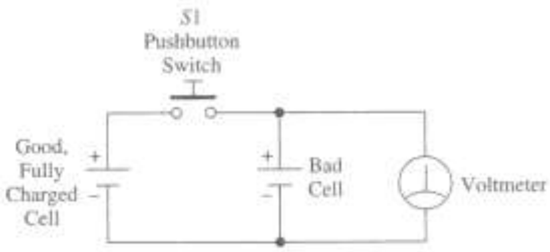


Figure 6.36 A flash revitilization circuit for batteries

A battery doesn't allow deep discharge after repeated shallow discharges; i.e. if it is discharged up to 80 % of full capacity repeatedly, it appears as if it is fully discharged when 80 % point is reached. In case of a premature failure, the battery can be reformed by repeatedly fully charging followed by immediately deep discharging it.

Explosion

A battery explosion may be caused by the misuse or malfunction of a battery, such as attempting to recharge a primary (non-rechargeable) battery, or short circuiting a battery. With car batteries, explosions are most likely to occur when a short circuit generates very large currents. In addition, car batteries liberate hydrogen when they are overcharged (because of electrolysis of the water in the electrolyte). Normally the amount of overcharging is very small, as is the amount of explosive gas developed, and the gas dissipates quickly. However, when "jumping" a car battery, the high current can cause the rapid release of large volumes of hydrogen, which can be ignited by a nearby spark (for example, when removing the jumper cables).

When a battery is recharged at an excessive rate, an explosive gas mixture of hydrogen and oxygen may be produced faster than it can escape from within the walls of the battery, leading to pressure build-up and the possibility of the battery case bursting. In extreme cases, the battery acid may spray violently from the casing of the battery and cause injury. Overcharging—that is, attempting to charge a battery beyond its electrical capacity—can also lead to a battery explosion, leakage, or irreversible damage to the battery. It may also cause damage to the charger or device in which the overcharged battery is later used. Additionally, disposing of a battery in fire may cause an explosion as steam builds up within the sealed case of the battery.

Leakage

Figure 6.36 A leaked alkaline battery

Figure 6.36 shows a leaking alkaline battery. Many battery chemicals are corrosive, poisonous, or both. If leakage occurs, either spontaneously or through accident, the chemicals released may be dangerous. For example, disposable batteries often use a zinc "can" as both a reactant and as the container to hold the other reagents. If this kind of battery is run all the way down, or if it is recharged after running down too far, the reagents can emerge through the cardboard and plastic that form the remainder of the container. The active chemical leakage can then damage the equipment that the batteries were inserted into. For this reason, many electronic device manufacturers recommend removing the batteries from devices that will not be used for extended periods of time.

Environmental Concern

The widespread use of batteries has created many environmental concerns, such as toxic metal pollution. Battery manufacturing consumes resources and often involves hazardous chemicals. Used batteries also contribute to electronic waste. Some areas now have battery recycling services available to recover some of the materials from used batteries. Batteries may be harmful or fatal if swallowed. Recycling or proper disposal prevents dangerous elements (such as lead, mercury, and cadmium) found in some types of batteries from entering the environment. In the United States, Americans purchase nearly three billion batteries annually, and about 179,000 tons of those end up in landfills across the country.

ELECTRICAL SAFETY

Scope and Purpose of Electrical Safety

Today, man is surrounded by electrical and electronic equipment. Some of them simple, some of them complicated, some considered essential, and some convenience, they are all intended to **serve** us. At times, however, we observe that they **harm** us. One of the ways that electrical equipment could cause physical harm is the **electrical shock** (Figure 6.37).

Figure 6.37 The electric shock

Electrical safety is containment or limitation of hazards:

- Electric shock to the patients, employees, and visitors in form of
    - Macroshock (both contacts are external to the body)
    - Microshock (one of the contact is inside of the body)

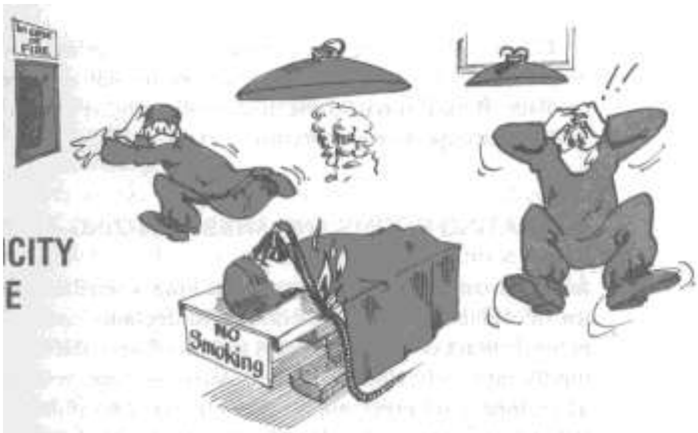


Figure 6.38 Fire caused by electricity

- Explosions that may result from electrical contact sparks that ignite variety of explosive gases, such as ether, or cyclopropane anesthetics.
- Fire (Figure 6.38)
- Damage to equipment and buildings

Hazards can be minimized but not eliminated. It is not static phenomena; rather it is a dynamic and continuous course of action involving hazard detection and correction. The scope of electrical safety includes any electrically operated equipment used in laboratories and public utilization areas of the Department. Safety is provided via power distribution and equipment design. Preventive maintenance procedures involving frequent equipment inspections and safety checks, uncovering early degradation of parts and replacements are needed for safe operation of equipment in the laboratories of the Department. Education and training of the lab engineers and students are essential ingredients of the safety measures.

What Is the Electrical Shock?

Electrical shock is defined as the undesirable biological damaging effect of an electrical current passing through the body. Electrical current could affect the body in three basic ways:

1. Resistive heating,
2. Electrical stimulation of nerves and muscles, and
3. Electrochemical burns (especially for DC current).

As a result it causes:

- Uncontrollable muscle contraction or unconsciousness,
- Ventricular fibrillation
- Injury to tissues
- Electrical burns
- Chemical burns (for dc currents)
- Muscular paralysis, injuries, pain and fatigue
- Breaking the bones and tendons
- Secondary (side) effects as falling of the ladder or spilling hot oil etc.

Figure 6.39 Direct contact with power lines

Electrical current flows through the body due to:

- Direct contact with power lines (Figure 6.39)
- Power line leakage in equipment to chassis (Figure 6.40)

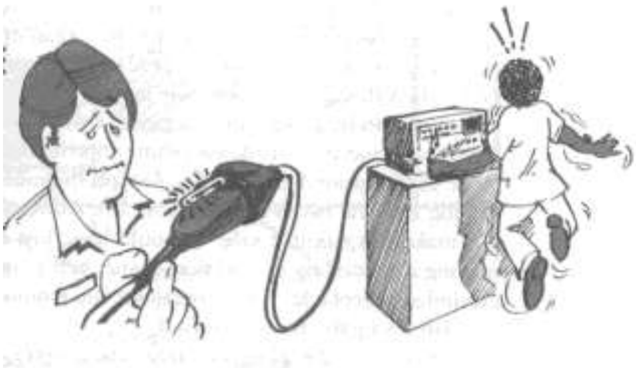


Figure 6.40 Power line leakage

- Leakage to the body from diagnostic and therapeutic equipment
- Uncontrolled electricity in the body during medical practices
- Defibrillator currents
- Electro surgical currents
- Diathermy currents

The severity of these effects depends on:

- Point of contact and the density,
- Frequency, and
- Duration of the current passing through the body.

Figure 6.41 illustrates the physiological effect of electricity. A current level below 0.5 milliampere at 60 Hz frequency will not be felt even if the person grips the conductor. However, as low as 0.2 milliampere may be sensed if the conductor makes a point contact. At low levels, it gives a tingling sensation and the victim can run away from further dangers of the electricity. As a rough guide, a current more than 10 milliamperes at 60 Hz frequency, for a duration of a few tenths of a second entering the body from one arm and leaving from the other arm or from the leg could be lethal. Yet, at current levels lower than 10 milliamperes, anywhere from just a tingling sensation to involuntary muscle contractions could result depending on the individual, raising the possibility of secondary physical injuries, such as falling from a ladder.

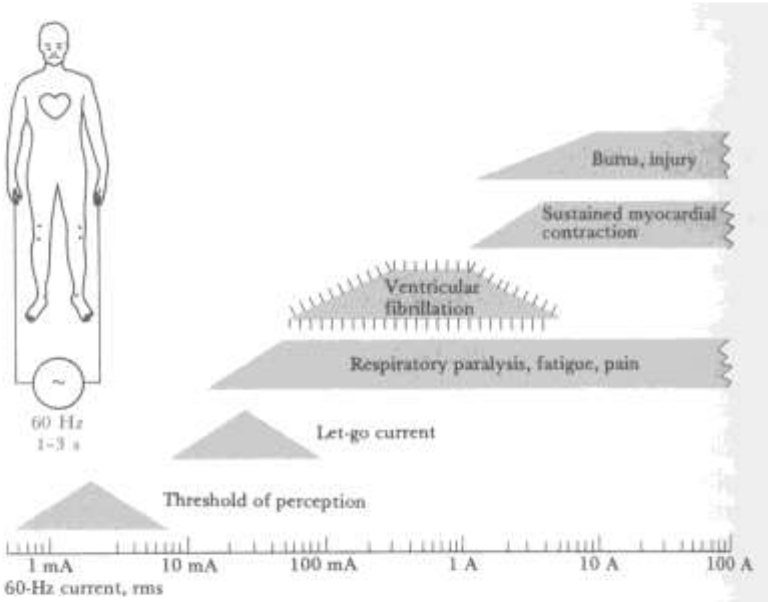


Figure 6.41 Ranges for the physiological effect of electricity

At current levels progressively higher than 10 milliamperes, respiratory paralysis, **ventricular fibrillation**, and burns result as illustrated in Figure 6.41. The figure represents estimated values given for each effect in a 70-kg male for 1-3 seconds exposure to 60 Hz current applied to copper grasped by hands. Among these, the ventricular fibrillation, a certain failure of the heart, is the major cause of death due to electric shock. The sensitivity of the individual varies. Women are more susceptible than the men. There is statistical variation in the level current to cause certain effects.

The amount of current required to cause a dangerous electric shock increases at frequencies below about 10 Hz, and above about 1000 Hz. This means that the 50 and 60 Hz frequency used for the mains supply is among the most dangerous, although technically and economically the most appropriate.

If the duration of the current passing through the body is less than about 0.1 second, even higher levels of current will not do any harm. The biological effects of electricity depend directly on the amount of **current** passing through the body, but not directly on the potential difference (voltage) applied to the body.

The voltage, being the force pushing the current though any circuit determines how much current would pass in relation to the total electrical resistance in the circuit. (Ohm's law: Current =Voltage/Resistance.) Since the total resistance is very difficult to predict in a typical electrical shock situation, safety standards for electrical shock are expressed directly in terms of current levels, rather than their voltage equivalents. However, it could be stated that voltages less than about 30 volts (rms) would not usually be able to cause dangerous amounts of current pass through the body under most **macro shock** conditions.

How the Electrical Shock Occurs?

An electric current could flow through the body unintentionally in one of the two situations explained below.

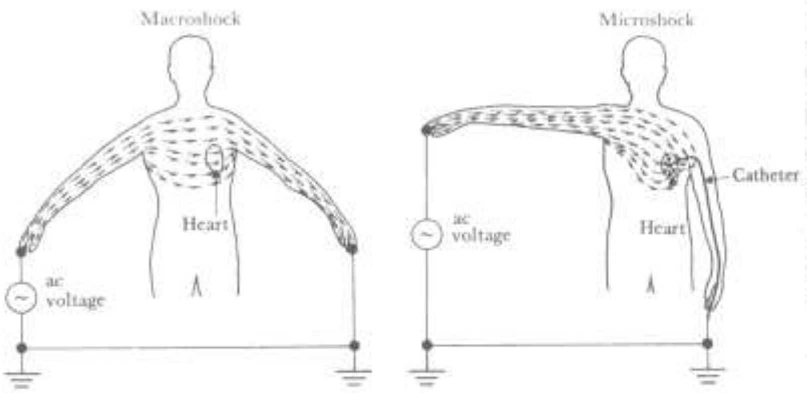


Figure 6.42 Illustration of macroshock and microshock (cardiac shock)

Macroshock Hazard

If an undesirable electric current enters and leaves the body through contacts on a limb such as the hand, arm, or foot, this is called a macro shock hazard, as shown in Figure 6.42. In this case the path of the current is quite wide as it passes through the chest where the heart is located. Only a small part of the total current affects the heart. Therefore the hazard is less. The dangerous current level of 10 milliamperes stated above is for a macro shock hazard.

Microshock (Cardiac Shock) Hazard

If in any way an electric current passes through the body with a direct electrical contact on the heart, this is called a micro-shock or cardiac shock hazard. Since all of the current would pass through the heart, the hazard is much more in the sense that even very small currents could damage the heart. The dangerous level of current directly applied to the heart could be as low as 10 microamperes. The micro shock hazard is normally limited to medical administration of electrically operated equipment on patients.

The prevention of the above-mentioned electric shock hazards share many common and some specific techniques, as summarized below.

How to Prevent Electrical Shocks?

At present, the potential causes of electric shock are well understood and comprehensive safety measures have been standardized. In many countries, these standards are obligatory and they are strictly enforced in the manufacturing and operation of all electrical equipment. However, even if rare, equipment not conforming to such safety standards might be available in the market. Also, properly manufactured equipment might lose its safety after some use or abuse. Therefore, the **educated buyer** or the user of electrical equipment should have an idea of the essential techniques of preventing the electric shock hazard both as built-in features of equipment and in the course of its utilization.

Electrical safety or protection from electric shocks can be achieved at three levels, namely

1. At the power distribution level,
2. At the equipment design level, and
3. At the utilization level.

Electrical Safety in Power Distribution

The present state of the electrical engineering science dealing with the distribution of electrical power dictates that one of the wires carrying the mains power be grounded (earthed) as illustrated in Figure 6.43. This grounding or earthing is done before it reaches the

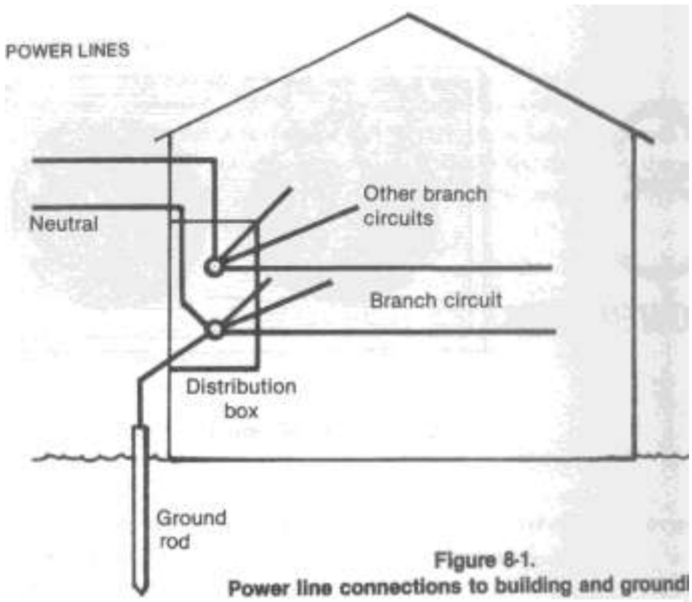


Figure 6.43 Distribution of electrical power

utilization point, usually at the transformer feeding a building. The grounded wire is called the "neutral". The other wires are called "phase", or "line", or "live", or "hot".

The requirement of grounding one of the power wires brings together the possibility that even if a person touches just a single wire, he could get an electric shock. If he touches the **neutral wire**, it is like touching ground (almost) and nothing will happen. But if he happens to touch one of the phase wires, rightfully called **live** or **hot**, the circuit will be completed through his feet touching the ground! Obviously, as illustrated in Figure 6.42, if both a phase and neutral wire, or two phase wires are contacted by two hands, an electrical current will pass through the body even if the feet are completely isolated from the ground.

The following safety measures are called in the distribution of electrical power in buildings.

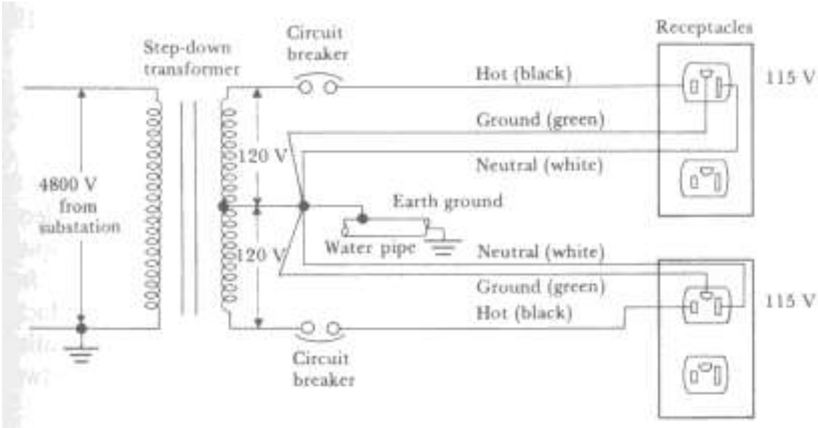


Figure 6.44 Simplified electrical power distribution for 115V/60 Hz

Figure 6.44 shows a simplified electrical power distribution in the US. Circuit breakers and switches to interrupt power, or to turn equipment on and off should be placed on the "hot" wire (phase), but not on the neutral wire. If a neutral wire going to equipment is interrupted, the equipment will not work, although the phase wire will still carry the dangerous mains voltage with respect to the earth.

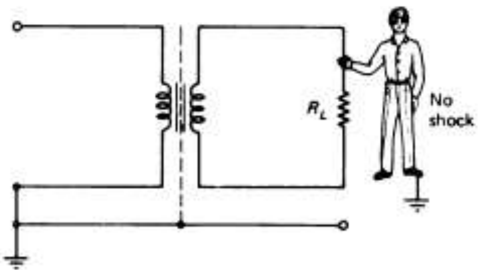


Figure 6.45 Utilization of isolation transformer

From the power distribution point of view, it is permissible to isolate the two mains wires from the ground in **limited** areas. This technique is called the "isolated power system", and utilized in wet areas and in operating rooms of hospitals. The transformer employed in this system (Figure 6.45) is called an **isolation transformer**. Its secondary winding is electrically insulated from the primary, and has some other special construction features. "Auto-transformers" commonly available in the market do not have an insulated secondary and they cannot be used for this purpose.

If an undesirable electrical connection occurs between the phase wire and the chassis of equipment, anybody touching the chassis will have an electrical current going through his body to the ground. In such a situation, instead of all of the current leaving the phase wire passing through the neutral, some is diverted to the ground. This is called a ground fault or earth leakage. This condition can be detected by monitoring the difference between the currents in the phase and neutral wires. They will be equal unless there is a ground fault. Simple and low cost devices are available in the market to continuously measure the difference and if a significant difference occurs, break the circuit immediately. These protection devices, called Ground Fault Circuit Interrupters (GFCI), or Earth Leakage Circuit Breakers (ELCB) are highly recommended for domestic use, and they are a **must** in the distribution of any wet area or outdoor installations (Figure 6.46). GFCI's are also available as an adapter to existing wall outlets.

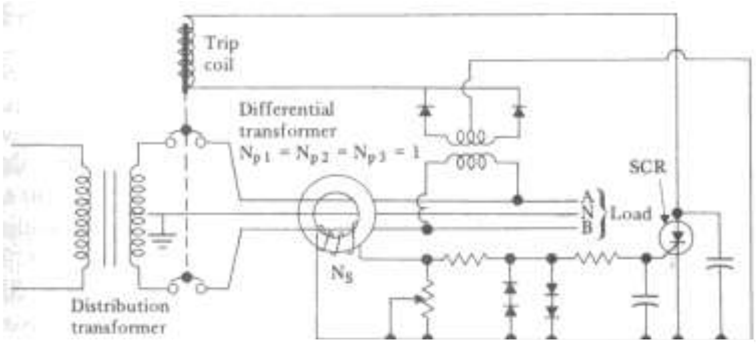


Figure 6.46 Ground Fault Circuit Interrupter (GFCI)

As detailed below, any exposed conducting surface of electrical equipment should be connected to the ground in order to discharge any current leaking to it. For this purpose, a local **grounding electrode system** is required to be established for each installation (i.e., building) as illustrated in Figure 6.47. This is the responsibility of the owner of the building, not the power company. In many countries the owner will be obliged to provide a grounding system in accordance with the applicable standards. The ground electrode connection should be brought to the central distribution board for the building, and from there on the ground wire will be carried along with the power lines in the distribution system inside. In this way, chassis grounding is conveniently done by the use of a three-way plug and socket pair. A direct connection to a metal water pipe buried under the ground could serve the purpose of grounding if certain conditions are satisfied.

The use of the neutral wire as the only way of grounding equipment is never permissible. Any failure of the neutral connection within the building could cause the phase voltage to appear on the chassis of equipment resulting in unexpected electrical shock accidents as illustrated in Figure 6.48.

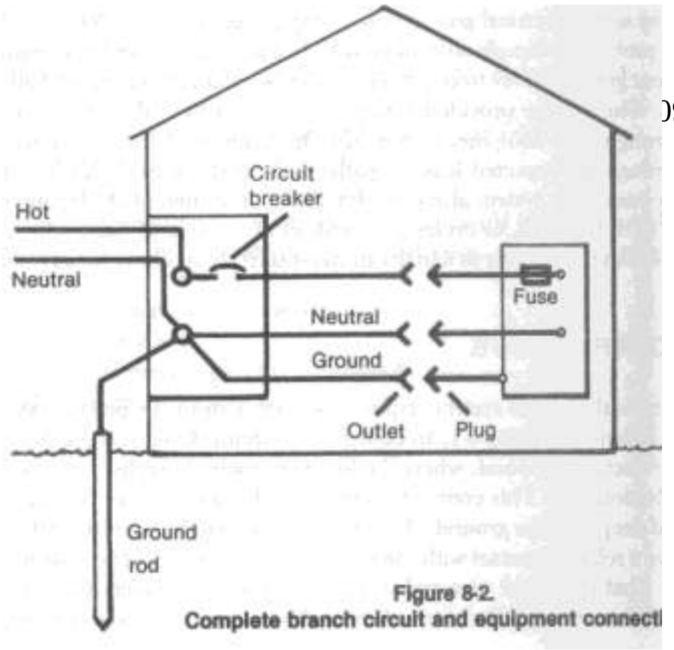


Figure 6.47 A complete branch circuit

We have to be careful in using the water pipe as a grounding point in Jeddah, since the pipe does not go to the ground; rather it goes to the tank in the roof. Such a case will electrify the whole building in case of a serious leakage.

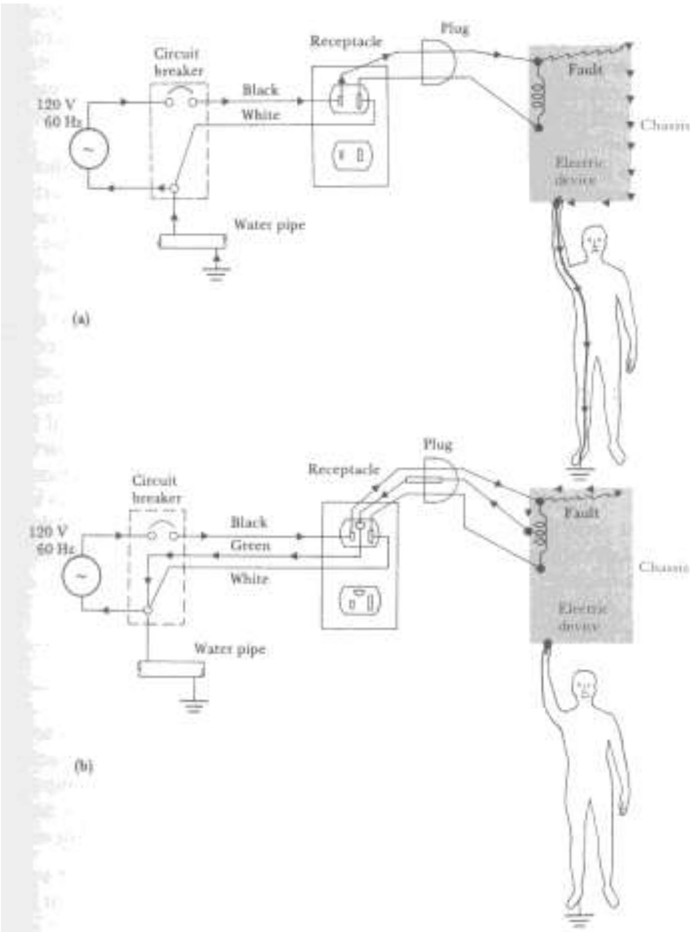


Figure 6.48 Ungrounded and grounded chassis

Electrical Safety in Equipment Design

Any metallic or otherwise conducting surface exposed on electrical equipment should be connected to the ground in order to discharge any current leaking to it.

Figure 6.48 (a) shows equipment with ungrounded chassis. The equipment works without any problem since the grounding of chassis is not essential for normal operation of it. However, a person touching the chassis drains all the leakage current to ground through his body.

Figure 6.48 (b) illustrates how safety is provided via the chassis grounding. High current flows through the circuit breaker in case of any serious fault developing in the equipment. This leads to tripping of the circuit breaker and interruption of the power to the equipment. Continuity of the safety ground wire and receptacle must be tested periodically.

This important safety requirement is relieved only if given equipment does not have any exposed metallic surfaces, or such surfaces are insulated from the current carrying conductors by a

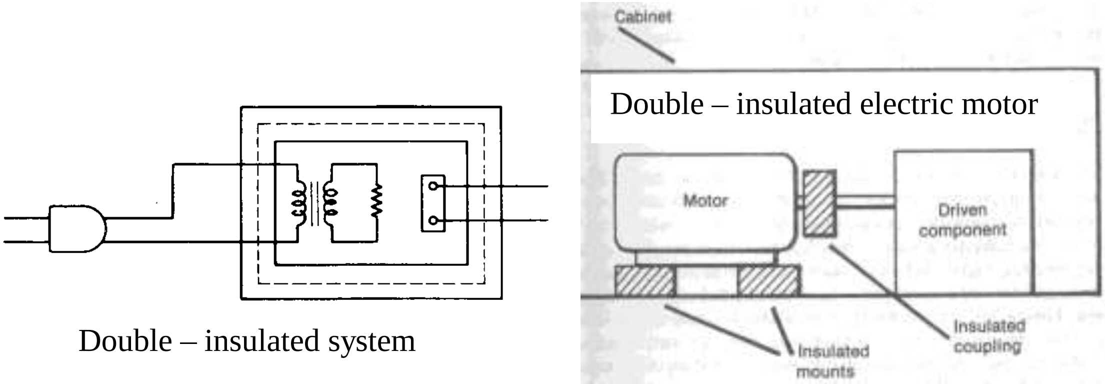


Figure 6.49 Double insulated system and an electric motor

double layer of insulation as illustrated in Figure 6.49. Such equipment is called "double insulated". However, since water entering this type of equipment could provide a leakage path to the outside, they cannot be employed in wet areas and outdoor applications safely.

Whenever the power requirements of equipment permit, it should be designed to operate from a low enough voltage to limit the current, which could pass in an accident. A voltage level below 30 volts (rms) could be considered safe in many applications. The low voltage should be obtained from batteries, or from an **isolation type** transformer feeding from the mains.

An isolation transformer has its secondary winding electrically insulated from the primary and some other special construction features. "Auto-transformers" commonly available in the market do not have an insulated secondary and they cannot be used for this purpose.

If equipment has signal connections to outside, such as existing in audio and video equipment, these should be electrically isolated from the mains voltage. This requirement can be satisfied in most applications by utilizing an isolating power transformer feeding all the circuits in equipment. In medical applications where direct body connections are required, special isolation techniques are utilized to limit the current, which could flow even at the worst cases.

Electrical Safety in Utilization

The first obligation of the buyer and user of electrical equipment is to make sure that it is conforming to the electrical safety guidelines stated above. If any significant deviations from these are suspected,

- Either the equipment should be rejected or
- A specialist in the field should be consulted.

It should be made sure that the electrical power distribution system at hand is satisfying the safety requirements. If equipment has a grounded, three-terminal plug, it should not be "adapted" to a mains outlet, which does not have a grounding terminal.

A fuse in the power distribution circuit or inside equipment not only protects against possible fire or extensive damage to the equipment, but also provides a line of defense against electrical shocks. In case a short circuit provides a current path from a phase wire to the grounded chassis in equipment, the excessive amount of current drawn will trip the fuse and immediately remove power from the equipment. If a fuse is over-rated or simply replaced by a thick wire this protection obviously fails.

Office Electrical Safety

Electricity is essential to the operations of a modern automated office as a source of power. Electrical equipment used in an office is potentially hazardous and can cause serious shock and burn injuries if improperly used or maintained.

Electricity travels through electrical conductors, which may be in the form of wires or parts of the human body. Most metals and moist skin offer very little resistance to the flow of electrical current and can easily conduct electricity. Other substances such as dry wood, porcelain, or pottery offer a high resistance and can be used to prevent the flow of electrical current. If a part of the body comes in contact with the electrical circuit, a shock will occur. The electrical current will enter the body at one point and leave at another. The passage of electricity through the body can cause great pain, burns, destruction of tissue, nerves, and muscles and even death. Factors influencing the effects of electrical shock include the type of current, voltage, resistance, amperage, pathway through body, and the duration of contact. The longer the current flows through the body, the more serious the injury. Injuries are less severe when the current does not pass through or near nerve centers and vital organs. Electrical accidents usually occur as a result of faulty or defective equipment, unsafe installation, or misuse of equipment on the part of office workers.

Types of electrical hazards found in an office environment include the following paragraphs.

Ungrounded Equipment

Grounding is a method of protecting employees from electric shock. By grounding an electrical system, a low-resistance path to earth through a ground connection is intentionally created. When properly done, this path offers sufficiently low resistance and has sufficient current-carrying capacity to prevent the build-up of hazardous voltages. Most fixed equipment such as large, stationary machines must be grounded. Cord and plug connected equipment must be grounded if it is located in hazardous or wet locations, if operated at more than 150 volts to ground, or if it is of a certain type of equipment (such as refrigerators and air conditioners). Smaller office equipment, such as typewriters and coffee pots, would generally not fall into these categories and therefore would not have to be grounded. However much of the newer office equipment is manufactured with grounded plugs as a precaution (three prong plugs). In such cases, the equipment should be used in accordance with the manufacturer's instructions. In any case, never remove the third (grounding) prong from any three-prong piece of equipment.

Overloaded Outlets

Insufficient or overloading of electrical outlets should be avoided. A sufficient number of outlets will eliminate the need for extension cords. Overloading electrical circuits and extension cords can result in a fire. Floor mounted outlets should be carefully placed to prevent tripping hazards.

Unsafe/Non-Approved Equipment

The use of poorly maintained or unsafe, poor quality, non-approved (by national testing laboratory) coffee makers, radios, lamps, etc. (often provided by or used by employees) should be discarded. Such appliances can develop electrical shorts creating fire and/or shock hazards. Equipment and cords should be inspected regularly, and a qualified individual should make repairs.

Defective, Frayed or Improperly Installed Cords for Electrically-Operated Office Equipment

Some common lethal electrical hazards are shown in Figure 6.50. When the outer jacket of a cord is damaged, the cord may no longer be water-resistant. The insulation can absorb moisture, which may then result in a short circuit or excessive current leakage to ground. If wires are exposed, they may cause a shock to a worker who contacts them. These cords should be replaced. Electric cords should be examined on a routine basis for fraying and exposed wiring.

Improper Placement of Cords

A cord should not be pulled or dragged over nails, hooks, or other sharp objects that may cause cuts in the insulation. In addition, cords should never be placed on radiators, steam pipes, walls, and windows. Particular attention should

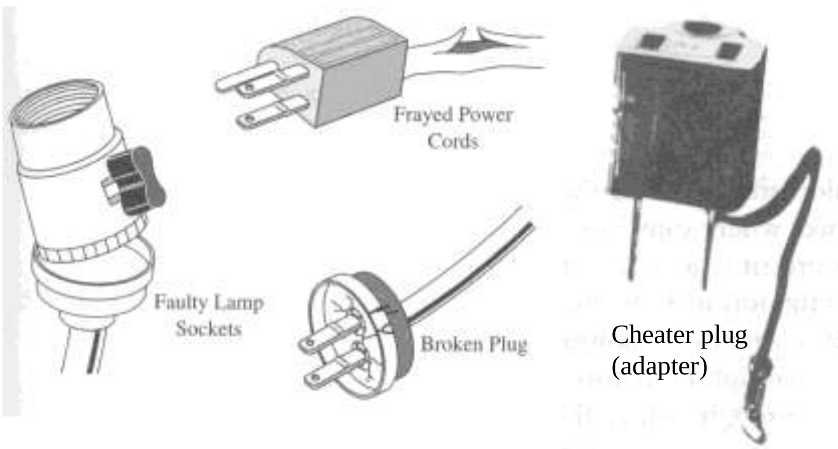


Figure 6.50 Common lethal electrical hazards

be placed on connections behind furniture, since files and bookcases may be pushed tightly against electric outlets, severely bending the cord at the plug.

Electrical Cords across Walkways and Work Areas

An adequate number of outlet sockets should be provided. Extension cords should only be used in situations where fixed wiring is not feasible. However, if it is necessary to use an extension cord, never run it across walkways or aisles due to the potential tripping hazard. If you must run a cord across a walkway, either tape it down or purchase a cord runner.

Live Parts Unguarded

Wall receptacles should be designed and installed so that no current-carrying parts will be exposed, and outlet plates should be kept tight to eliminate the possibility of shock.

Pulling of Plugs to Shut Off Power

Switches to turn on and off equipment should be provided, either in the equipment or in the cords, so that it is not necessary to pull the plugs to shut off the power. To remove a plug from an outlet, take a firm grip on and pull the plug itself. Never pull a plug out by the cord.

Working on "Live Equipment"

Disconnect electrical machines before cleaning, adjusting, or applying flammable solutions. If a guard is removed to clean or repair parts, replace it before testing the equipment and returning the machine to service.

Blocking Electrical Panel Doors

If an electrical malfunction should occur, the panel door, and anything else in front of the door will become very hot. Electrical panel doors should always be kept closed, to prevent "electrical flashover" in the event of an electrical malfunction.

PROBLEMS ON SOURCES OF ELECTRICAL ENERGY

Review Questions

1. What is a power supply?
2. Why do you need a DC power supply?
3. What are the critical factors effecting the choice of a power supply?
4. How a laboratory power supply differs from an instrument power supply?
5. What is the ripple factor?
6. What are the load and input regulations?

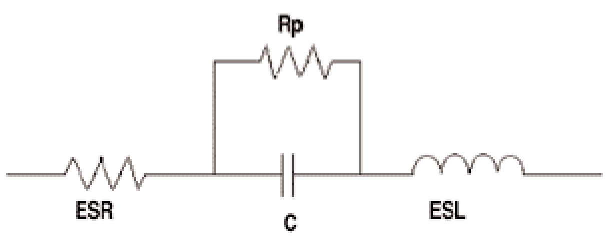
7. What is the efficiency of a power supply?
8. What are the indispensible components of a power supply?
9. What are the AC line components of a power supply?
10. What is a fuse?
11. What type of a fuse is preferred in power supplies?
12. What is the meaning of the voltage rating of a fuse?
13. What is a feasible link?
14. What is the transient suppressor and why it is used at the input section of a power supply?
15. What is the function of the line filter in power supplies?
16. What is the snubber, what is its function and in what position you expect to see it in a power supply?
17. What are the components of a snubber and what are their important properties?
18. What is special about the transformer used in power supplies?
19. What is the function of the rectifier diode in a power supply?
20. What are the differences between rectifier diodes and other types of diodes that you know?
21. How can you test a diode using a multimeter?
22. Why half-wave rectifiers are not commonly used although they are very simple?
23. What is the peak inverse voltage of a rectifier diode and how it is used in selecting rectifier diodes?
24. Why do you need smoothing in power supplies?
25. What are the circuit modalities used for smoothing in power supplies?
26. How can you choose a smoothing capacitor for a given power supply application?
27. What is the "bleeding" resistor, where and why it is used?
28. Why a small non-electrolytic capacitor is connected in parallel with the electrolytic smoothing capacitor in power supplies?
29. Why do you need for a voltage regulator in power supplies that are used in electronics?
30. What is a zener diode and how it differs from an ordinary rectifier diode?
31. What are the advantages of integrated circuit regulators over the discrete ones?
32. What is a crowbar and how it is used in protecting power supplies?
33. What is a switched regulator and how it differs from the linear regulator?
34. What are the major advantages of switching regulators over the linear ones?
35. What are the major disadvantages/limitations of switching power supplies?
36. What is the function of the high frequency switch in switching regulators?
37. What are the similarities and differences between the input and output rectifiers used in switching power supplies?

38. What are the similarities and differences between the input filter capacitors and output filter capacitors?
39. What is the function of the pulse width modulator (PWM) in regulating the output voltage?
40. Why we have problem of RF interference in switching supplies and how it can be eliminated?
41. What element contributes most to the weight of the power supply and why the switching supply is much lighter than its linear counterparts?
42. What is a battery and what is its function in electronics?
43. What are the anode and cathode as referred to a battery?
44. What is the principle of operation of batteries?
45. What is a primary battery and what are the commonly available ones?
46. What are the advantages of alkaline batteries?
47. What is a secondary battery and how it differs from the primary battery?
48. What are the meanings of "a dry cell" and "a wet cell"?
49. How is the battery capacity expressed?
50. What is the meaning of "shelf life" for a battery?
51. What are the factors that affect the life of a battery?
52. What are the commonly used battery charging protocols for secondary batteries?
53. What is the trickle charge?
54. Why does the battery leak?
55. Why may the battery explode?
56. What is electricity and electric shock?
57. What is electrical safety?
58. What is the scope of electrical safety?
59. Why the birds can sit on electrical conductors and yet do not get electrical shock?
60. What are the electrical hazards that might be faced in a regular office environment?
61. What are the electrical hazards that might be faced in a medical environment?
62. Why the patients with electrodes are more susceptible to electrical shock?
63. What are the important levels of 60 Hz electrical current for an average individual?
64. What are the macroshock and microshock hazards?
65. What is the safety ground and how it can prevent the electric shock?
66. Why the water pipe cannot be used for grounding in domiciliary environment in Jeddah?
67. What is an isolated power system?
68. What is a ground fault circuit interrupter and how it can be used for a three-phase power system?
69. What are the ways of protection against electrical shock by means of equipment design?

70. Why can a double-insulated operate safely without a ground connection?

Exercises on Power Supplies

1. Define the following terms related to the power supplies:
   a. Ripple factor
   b. Load regulation
   c. Input regulation
   d. Efficiency
2. Draw the block diagram of a linear regulated power supply and describe the major function each block briefly.
3. Explain the function of the fuse in power supplies. What type of a fuse is preferred in power supplies?
4. Explain shortly the function of a transformer in a power supply with a simple circuit symbol and input/output waveforms.
5. What are the critical factors in selecting the transformer for a power supply?
6. Define the efficiency of the transformer in a power supply.
7. Discuss how to select a transformer for a given power supply application with an example.
8. Discuss the function of the rectifier diode and the difference between rectifier diodes and other types of diodes that you know.
9. Discuss the reasons for half-wave rectifiers not being commonly used although they are very simple.
10. Describe how to test a diode using a multimeter.
11. Define the forward current (I_F), surge current (I_{SFM}), forward diode voltage (V_D) and peak inverse voltage (PIV) for a rectifier diode with a simple sketch.
12. Mathematically determine the average and effective values and the ripple factor for half wave and full wave rectified voltages.
13. Discuss the determination of the peak inverse voltages in selecting rectifier diodes.
14. Discuss the necessity for smoothing and circuit modalities used for this purpose.
15. Calculate the smoothing capacitor required for a supply with output voltage 12 V, current 0.5 A, frequency of the main's supply 60 Hz and ripple factor 10%.
16. Calculate the approximate charging and discharging times at steady state for the capacitor in the previous question. Determine the approximate value of the average charging current at steady state.



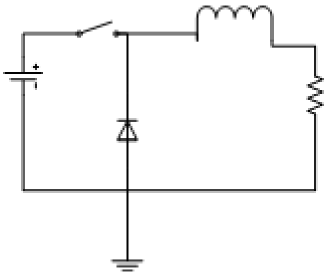
17. Figure shows the equivalent circuit of a smoothing capacitor. Define each

component in the circuit and discuss how they affect the performance of the capacitor in a power supply.

18. Explain the reason for heaving a small non-electrolytic capacitor across the smoothing capacitor.
19. Discus the reason for adding a small resistance between the output of the rectifier and smoothing capacitor.
20. Discuss how to choose a smoothing capacitor for a given power supply application.
21. What is the "bleeding" resistor, where and why it is used?
22. An unregulated power supply has 2200 μF aluminum electrolytic smoothing capacitor in parallel with 0.1 μF polystyrene capacitor. The nominal value of the output voltage is 10 V for the output current of 0.5 A and main's voltage 220 V at 60 Hz.
    a. Calculate the DC component of the output voltage and the ripple voltage for the load current of 0.1A.
    b. Repeat (a) for the load current of 1 A.
    c. Calculate the output voltage and ripple for the output current 500 mA as the main's voltage dropping to 200 V.
    d. Repeat (c) for the main's voltage rising to 240 V.
23. Generate a comparison table and discuss the effect of load current and input voltage variations on the performance of the power supply.
24. Discuss the need for a voltage regulator in power supplies that are used in electronics.
25. Design a zener diode regulated power supply assuming that:
    - The required output voltage is 5 V
    - The output current is between 0 and 100 mA
    - Transformer used is 220 V / 6 V.

    a. Using commercial components, select the rectifier, smoothing capacitor and limiting resistor.
26. Search for 5 IC voltage regulators from component catalogs and/or web and make a table of comparison for their characteristics.
27. Design a linear regulated dual power supply that would provide 1 A load current at ± 6 V from a mains supply of 220 V / 60 Hz. Use practical values for the components and justify your selections.
28. Explain the function of the high frequency switch in switching regulators.
29. Draw the functional block diagram of a switching power supply and explain the similarities and differences between a regular transformer used in ordinary power supplies and high-frequency transformer used in switching power supplies.

30. Explain the similarities and differences between the input and output rectifiers used in switching power supplies.
31. Explain the similarities and differences between the input filter capacitors and output filter capacitors.
32. Explain the function of the pulse width modulator (PWM) in regulating the output voltage.



33. The circuit shown is driven off by a 12 V DC supply. The inductor is 10 mH and the resistor is 100 Ω. The switch works at 1 kHz with 40% duty cycle (i.e. it is "on" for 0.4 ms and "off" for 0.6 ms in a 1 ms cycle). Determine and draw the waveform of the voltage across the resistor. What happens if the frequency of the switch goes to 10 kHz? What happens if the switch works at 100 kHz? (Assume that the diode is ideal, i.e. it works as an electronic switch).
34. The current in a 10 Ω resistor is 5*sin(314t) A
    a. Draw the waveform of the current
    b. Define and calculate the following values for the current:
        i. Peak
        ii. Peak to peak
        iii. Average
        iv. Root Mean Square (RMS)
    c. Calculate the value of the power dissipated by the resistor
    d. How much would be the current if it would be DC to generate the same power on the resistor?
35. For a transformer in a power supply, the required average output voltage is 10 V, the ripple voltage is 1 V and the voltage drop across the rectifier is 2 V and the required output current (average) is 1 A. The efficiency (η) of the transformer is 0.8. Calculate:
    a. the required output voltage of the transformer
    b. the input current of the transformer if the input voltage is 220 V
    c. the output power delivered by the power supply
    d. the power loss by the transformer.
36. A series R-L circuit has R = 0.1 kΩ and L = 10 mH. The circuit is excited by Vi = 5 + 10 sin(1000t) V
    a. Draw the circuit diagram
    b. Calculate the voltages across R and L.

Exercises on Batteries

Multiple-Choice Questions

1. Which one of the following cell is not a primary cell?
   a. Carbon-zinc
   b. Alkaline
   c. Zinc-chloride
   d. Lead-acid
2. The dc output of a C-size alkaline cell is
   a. 1.2 V
   b. 1.5 V
   c. 2.1 V
   d. About 3 V
3. Which of the following cell is a secondary cell?
   a. Silver oxide
   b. Lead-acid
   c. Nickel-cadmium
   d. Both b and c
4. What happens to the internal resistance, r_i, of a voltaic cell as the cell deteriorates?
   a. It increases
   b. It decreases
   c. It stays the same
   d. It usually disappears
5. The output voltage of a lead-acid cell is
   a. 1.35 V
   b. 1.5 V
   c. 2.1 V
   d. About 12 V
6. Cells are connected in series to
   a. Increase the current capacity
   b. Increase the output voltage
   c. Decrease the voltage output
   d. Decrease the internal resistance
7. Cells are connected in parallel to
   a. Increase the current capacity

b. Increase the output voltage

c. Decrease the output voltage

d. Decrease the currents capacity

8. Five D-size alkaline cells in series have a combined voltage of

a. 1.5 V

b. 5.0 V

c. 7.5 V

d. 11.0 V

9. A battery has no load voltage of 9 V. It's terminal voltage drops to 8.25 V when a load current of 200 mA is drawn from the battery. The internal resistance ri equals

a. 0.375 Ω

b. 3.75 Ω

c. 41.25 Ω

d. 4.5 Ω

10. The main difference between the primary and secondary cell is that

a. A primary cell can be recharged and a secondary cell cannot

b. A secondary cell can be recharged and a primary cell cannot

c. A primary cell has an unlimited shelf life and a secondary cell does not

d. A primary cell produce a dc voltage and secondary cell produce ac voltage

11. Which one of the following batteries has a cell voltage of 1.2 V?

a. Lead-acid

b. Zinc-chloride

c. Nickel-cadmium

d. Lithium

12. Five nickel-cadmium cells in series have a combined voltage of

a. 5.0 V

b. 6.0 V

c. 7.5 V

d. 11.0 V

13. What type of battery or cell would likely be used to power this portable drill?

a. A mercury oxide button battery

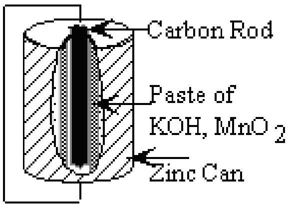
b. A lead storage battery

c. A nickel-cadmium battery

Black and Decker
Portable Power Drill

d. A hydrogen-oxygen fuel cell

14. This type of alkaline cell is commonly used to power flashlights and other similar objects. Which is the anode of the cell?
    a. Carbon rod
    b. Paste of KOH, MnO_2
    c. Z inc can
    d. Water



General Questions on Batteries

1. Many high-end cell phones are equipped with lithium ion batteries. Use the resources of the Web to find out more about this type of battery by searching for "lithium battery chemistry."
2. Are lithium-based batteries better than nickel-metal hydride ones? Use the Web to find details about these two types of batteries. Then, write a brief summary of your findings and give your conclusion as to which battery would be more suitable for use in an electric vehicle.
3. Why is this battery suited for use in portable devices?
4. What materials form the anode and the cathode of a lithium ion battery?
5. What is the voltage of a lithium ion battery?
6. What other types of batteries are used in cell phones? What are their advantages and disadvantages compared to lithium ion batteries?
7. Draw the circuit diagram of a battery charger that has 15 V output and used two charge two 12 V lead-acid batteries simultaneously.
8. How much is the energy in Joule stored in D-size alkaline battery?
9. Make a web search and find out the type of cell that is typically used in watches, hearing aids, cameras, etc. Explain the reason for its preference over others.
10. How long it will take to charge a flat (no initial charge) 1.8 A-H battery from a constant current source that supplies 200 mA into the battery during charging.

Exercises on Electrical Safety

Multiple-Choice Questions – A

Multiple Choice: In the following group of questions select the statements, which are correct (there may be more than one correct statement in each problem).

1. Physiological effect of electricity depend:
    a. Solely on the voltage applied to the body since it is the high enough voltage which breaks down the skin insulation and causes an electric shock;

   b. On the current which passes through the body;
   c. On both voltage and the total impedance of the circuit since these determine the current.
2. The dangerous levels of electric shock depend:
   a. Only on the total amount of current passing through the body;
   b. On the current density across critical organs.
3. In the following statements, the electrical current mentioned passes two hands of an adult male for about 1 second:
   a. The minimum current perceivable by the most sensitive person is about 0.5mA;
   b. The most fortunate person can take his hands off the hot conductors at current levels up to 100mA;
   c. Respiratory paralysis can occur at current levels < 20mA;
   d. The most dangerous form of electric shock hazard, ventricular fibrillation occurs between about 50mA and 5 Amperes;
   e. Currents > 6A does not usually cause fibrillation or any known damage to the heart, but it may cause respiratory paralysis.
4. The most dangerous frequency for electric shock is:
   a. Low frequencies (approximately 10Hz to 100Hz);
   b. Zero frequency (direct current);
   c. High frequencies.
5. In Jeddah, the power distribution to non-industrial districts is by:
   a. Only a single line conductor at 220V plus a neutral;
   b. Two line conductors plus a neutral, that is two phases 180 degrees apart;
   c. Three phase system, line to neutral voltage being 127 Volts and line-to-line voltage 220 Volts.
6. The precautions that can be taken against the macro-shock electric hazards are:
   a. Driven right leg circuit for ECG equipment;
   b. Double insulation of the equipment;
   c. Optical isolation of the amplifier circuits;
   d. Proper grounding of the equipment cases;
   e. Isolation transformers for the power distribution.
7. The precautions against the micro-shock hazard could be:
   a. Running individual grounding conductors form each equipment to a central ground terminal in every patient room;
   b. Battery operated, double insulated equipment;

   c. Isolation transformers for the power distribution;
   d. Isolation transformers for supplying power to OPAMP circuits and for output connections.

8. In arm-to-arm passage of 60 Hz current, levels above 6 Amps. generally does not cause ventricular fibrillation because:
   a. Current is well distributed throughout the chest leaving negligible amount through the heart;
   b. It stimulates the whole heart;
   c. Patient dies as soon as it is applied;
   d. 60Hz does not stimulate the active cells.
9. According to the U.S. NFPA standards, the leakage current limits for electrical appliances are:
   a. For appliances not intended to contact patients, chassis leakage = 100mA;
   b. For appliances likely to contact patients, chassis leakage = 100mA and patient lead (electrode) leakage = 10mA;
   c. For appliances with "isolated" patient leads, chassis leakage = not applicable, patient lead = 10mA.
10. Ground fault circuit interrupter devices are usually used in the:
    a. Operating room;
    b. EEG laboratories;
    c. Hemodialysis ward.
11. An equipotential ground system:
    a. Consists of a separate additional ground wire connections from each equipment chassis and metal surface to a central ground terminal;
    b. Consists of a separate ground wire connecting the metal surfaces of each equipment to each other in cascade order (one after another);
    c. Reduces differential potentials between surfaces to zero;
    d. Used in operating rooms, ICU and CCU.

Multiple-Choice Questions – B

Fill in the Blank Spaces in the following group of questions.

1. In power systems, the black wire is ________________, the white wire is ______________, and the green wire is _______________.
2. The maximum differential voltage between metal surfaces in critical care areas is ______________mV.

3. Specialized hospital electrical safety test equipment measures ___________ resistance, _______________ polarity, ____________ spring tension and _______________ current.
4. Leakage current can be reduced by adding a _________ wire from equipment metal chassis to a common _____________ terminal.
5. Leakage current standards are _________ microamper or less for critical care areas, ______________ microamper or less for patient care areas, and _____________ microamper or less for public areas of the hospital.

General Questions

Solve the Following Problems in Detail.

1. The patient's right hand touches the bed-rail, which is coupled to 220V rms above ground through 1600pF leakage capacitor of the driving electric motor. The left hand of the patient touches the metal base of a lamp, which is grounded. A saline-filled catheter (R=20KΩ) for measuring blood pressure is connected to the patient's heart. Some of the pressure transducer strain - gage wiring is grounded, and the transducer is somewhat isolated electrically. However, there is 100 pF capacitance between the ground and the saline. Assume the skin resistance of the patient is 100KΩ.
    a. Draw a complete equivalent circuit indicating the paths of leakage currents through the patient's body;
    b. Compute the rms current through the patient's heart for the above situation;
    c. Is there a microshock or macroshock hazard, why?
2. Draw the circuit diagram for a ground fault circuit interrupter for a three-phase power system.
3. State the ways of protection against electrical shock by means of equipment design.
4. Define safety.
5. List types of hazards that might be faced in a medical environment.
6. Define each hazard you have and list at least three types for each category. Discuss ways of protection for each type.
7. Define electricity.
8. Define electrical shock.
9. Define the scope of electrical safety.
10. Draw a symbolic electrical diagram that indicates the patient and conditions of electrical shock.
11. Explain why the patients with electrodes are more susceptible to electrical shock.
12. Explain the response of the human body to electrical current at 60 Hz. What are the important levels for an average individual?
13. Explain the macroshock and microshock hazards.

BIBLIOGRAPHY

Further Reading

Power supplies

E.R. Hnatek, *Design of solid state power supplies*, Van Nostrand Reinhold, NewYork, 2nd ed, 1989.

S.J. Bigelow, All about switching power supplies, *Electronics Now*, pp. 40-47, August 1997.

L.R. Luchi, Power supply regulation, *Electronics Now*, pp. 69-76, December 1994.

Batteries

R.M. Dell and D.A.J. Rand, *Understanding Batteries*, RSC Paperbacks, 2001, ISBN 0-85404-605-4

D. Linden and T.B. Reddy, *Handbook of Batteries*, 3rd ed., McGraw-Hill, 1995, ISBN 0-07-135978-8

G. Pistoia, *Batteries for Portable Devices*, Elsevier, 2005

V. Pop et al, *Battery Management Systems*, Springer, 2008

T.R. Crompton, *Battery Reference Book*, 3rd ed, Newnes, 2000

Shultz, Grob's *Introduction to Electronics*, McGraw-Hill, 2007

J.J. Carr, and J.M. Brown, *Introduction to Biomedical Equipment Technology*, 3rd ed. Prentice-Hall, 1997.

Useful Websites

Power supplies

http://www.electronics-tutorials.ws/diode/diode_7.html (last visited in March 2011)

http://en.wikipedia.org/wiki/Switched-mode_power_supply (last visited in March 2011)

fuse_bible_complete.pdf, www.swecheck.com.au

Fuse_ApplicationsGuide.pdf, www.bussmann.com

Wikipedia sources on batteries

-

TEMPERATURE MEASUREMENT

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

1. Define temperature.
2. Describe temperature scales.
3. Interpret reference temperatures.
4. List temperature measuring devices.
5. Explain principles of thermocouples.
6. Describe resistance temperature devices.
7. Describe the principles and applications of radiation detectors (infrared sensors).
8. Explain the principles and applications of integrated circuit (I.C.) sensors.
9. Describe the principles and applications of bimetallic devices in temperature sensing.
10. Explain the principles and applications of fluid-expansion devices and chemical (change-of-state) sensors.
11. Compare practical temperature measurement devices.
12. Illustrate the principle of temperature measurement using thermocouples.
13. State the empirical laws of thermocouples.
14. Describe how to measure the thermocouple voltage using a digital voltmeter (DVM).
15. Discuss the importance of the reference junction.
16. Describe the reference circuit that replaces the function of the reference junction.
17. Describe the software compensation technique that replaces the function of the reference junction.
18. Discuss the reasons for commonly using thermocouples in temperature measurement.
19. Explain the principle of operation of thermistors.
20. Describe the thermistor linearization techniques.
21. Explain the thermistor thermometry.

BASIC PRINCIPLES

Definition of Temperature

Temperature is an expression for the kinetic energy of vibrating atoms and molecules of a matter. This energy can be measured by various secondary phenomena, e.g., change of volume or pressure, electrical resistance, electromagnetic force, electron surface charge, or emission of electromagnetic radiation. Many engineering applications require direct measurement of the temperature. Synthetic fuel research, solar energy conversion and new engine development are a few of these disciplines. All industries place new emphasis on energy efficiency. Hence, the fundamental measurement of temperature assumes new importance. Temperature also effects measurement of most physical variables and it must be measured for compensation purposes as well.

Temperature Scale

The most frequently used temperature scales are Celsius and Fahrenheit, which divide the difference between the freezing and boiling points of water into 100° and 180°, respectively.

$°C = (5/9)(°F - 32)$, and $°F = (9/5)\,°C + 32$

The thermodynamic scale begins at absolute zero, or 0 Kelvin, the point at which all atoms cease vibrating and no kinetic energy is dissipated.

$0\ K = -273.15°\ C = -459.67°\ F$

The official Kelvin scale does not carry a degree sign. The units are expressed in “kelvins,” not degrees Kelvin.

Reference Temperatures

We cannot build a temperature divider as we can a voltage divider, nor can we add temperatures as we would add lengths to measure distance. We must rely upon temperatures established by physical phenomena, which are easily observed and consistent in nature. The International Temperature Scale (ITS) is based on such phenomena. Revised in 1990, it establishes seventeen fixed points and corresponding temperatures. Reference temperatures include the triple-points (the temperature and pressure at which solid, liquid, and gas phases of a given substance are all present simultaneously in varying amounts) of several important engineering substances. Examples:

- Triple-point of water = 0.01°C,
- Triple-point of hydrogen = -259.3467°C, and
- Freezing point of silver = 961.78°C.

Since we have only these fixed temperatures to use as a reference, we must use instruments to interpolate between them. But accurately interpolating between these temperatures can require some fairly exotic transducers, many of which are too complicated or expensive to use in a practical situation.

TEMPERATURE MEASURING DEVICES

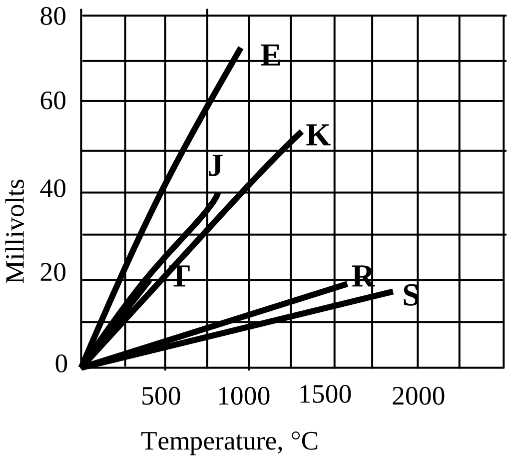
Temperature can be measured via a diverse array of sensors. All of them infer temperature by sensing some change in a physical characteristic of the device. The types with which an engineer is likely to come into contact are:

- Thermocouples,
- Resistance temperature devices (RTD's and thermistors),
- Infrared radiators,
- I.C. sensors,
- Bimetallic devices,
- Liquid expansion devices, and
- Change-of-state devices

In the chemical process industries, the most commonly used temperature sensors are thermocouples, resistive devices and infrared devices.

Thermocouples

Thermocouples consist essentially of two strips or wires made of different metals and joined at one end. An electromotive force (e.m.f) is induced between the other ends whose value is related to the temperature of the junction. As temperature goes up, this output e.m.f of the thermocouple rises, though not necessarily linearly. Output voltages for some popular thermocouples are plotted as a function of temperature in Figure 7.1. It is the most versatile temperature transducer.



Type of Metals

	+	-
E	Chromel vs	Constantan
J	Iron vs	Constantan
K	Chromel vs	Alumel
R	Platinum vs	Platinum 13% Rhodium
S	Platinum vs	Platinum 10% Rhodium
T	Copper vs	Constantan

Figure 7.1 Typical thermocouple characteristics

Resistance Temperature Devices

Resistance temperature devices capitalize on the fact that the electrical resistance of a material changes as its temperature changes;

$$R = R_0[1 + \alpha(T - T_0)]$$

Where R_0 is the resistance at $T=T_0$ and α is the temperature coefficient of the device. Two key types are the metallic devices (commonly referred to as RTD's), and thermistors.

RTD's

As their name indicates, RTD's rely on resistance change in a metal, with the resistance rising more or less linearly with temperature. The most common RTD's are made of either platinum, nickel, or nickel alloys. The economical nickel derivative wires are used over a limited temperature range. They are quite non-linear and tend to drift with time. For measurement integrity, platinum is the obvious choice. A typical RTD consists of a fine platinum wire wrapped around a mandrel and covered with a protective coating (also abbreviated PRTD). It is the most stable temperature transducer.

In the newest construction technique, a platinum or metal-glass slurry film is deposited or screened onto a small flat ceramic substrate, etched with a laser-trimming system, and sealed to form the film RTD. It offers substantial reduction in assembly time and has the further advantage of increased resistance for a given size. Due to the manufacturing technology, the device size itself is small, which means it can respond quickly to step changes in temperature. Film RTD's are less stable than their wire-wound counterparts, but they are more popular because of their advantages in size, production cost and ruggedness.

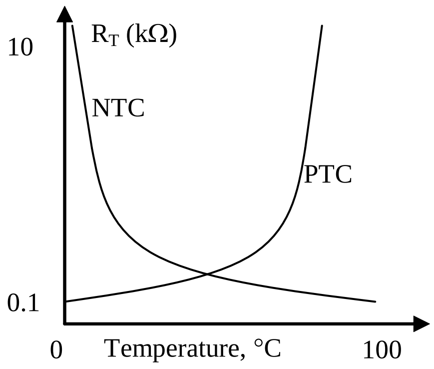


Figure 7.2 Illustration of NTC and PTC type thermistors

Thermistors

Like the RTD, the thermistor is also a temperature sensitive resistor. It is based on the resistance change in a ceramic semiconductor; the resistance drops nonlinearly with temperature rise. There are two types as the positive temperature coefficient (PTC) and negative temperature coefficient (NTC) as illustrated in Figure 7.2. Although positive temperature coefficient units are available, most thermistors have a negative temperature coefficient (TC); that is, their resistance decreases with increasing temperature. The negative TC can be as large as several percent per degree C, allowing the thermistor circuit to detect minute changes in temperature, which could not be observed with an RTD, or thermocouple circuit. The PTC type is used mainly in thermostat type applications in which the electrical power applied to an electrical element, like a motor, is interrupted as the temperature (of its winding) goes above a preset value.

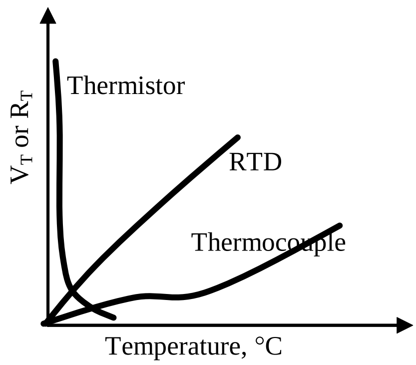


Figure 7.3 Three temperature measuring devices together

The thermistor is the most *sensitive* temperature transducer. Of the three major categories of sensors shown in Figure 7.3, the thermistor exhibits by far the largest parameter change with temperature. The price we pay for this increased sensitivity is loss of linearity. The thermistor is an extremely non-linear device, which is highly dependent upon process parameters. Consequently, manufacturers have not standardized thermistor curves to the extent that RTD and thermocouple curves have been standardized.

The resistance-temperature relationship of a NTC type thermistor is negative and highly nonlinear. This poses a serious problem for engineers who must design their own circuitry. However, using thermistors in matched pairs, in such a way that the nonlinearities offset each other, can ease the difficulty. Furthermore, vendors offer panel meters and controllers that compensate internally for thermistors' lack of linearity.

The Self-Heating Problem

Other important problem that effects the thermistor and all other resistance temperature devices is the self-heating. The current passing through the device causes conversion of the electrical energy to heat at a rate

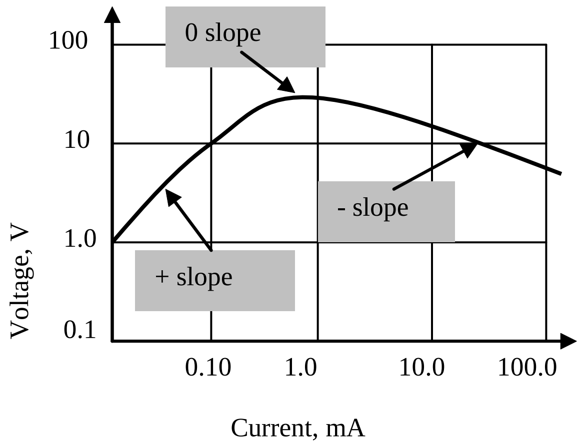


Figure 7.4 Self-heating in thermistors

$P = I^2R_t$

Heat generated is dissipated to the environment. Rate of dissipation is proportional to the difference between the case temperature of the thermistor and the ambient temperature. The equilibrium is reached as the rate of dissipation balances the rate of generation. An increase in the case temperature causes a decrease in the resistance of the thermistor. The voltage across the thermistor is $V = IR_t$

The V-I characteristic of a typical thermistor is shown in Figure 7.4. The device is used as a temperature transducer in the "+" slope region where the self-heating is negligible.

Radiation Detectors (Infrared Sensors)

Infrared (IR) sensors are non-contacting devices that infer temperature by measuring the thermal radiation emitted by the surface of a material as illustrated in Figure 7.5. Electro-magnetic energy radiates from all matters regardless of their temperatures. In many process situations, the energy is in the infrared region. As the temperature goes up, the amount of infrared radiation and its average frequency go up. Different materials radiate at different levels of efficiency. This efficiency is quantified as emissivity, a decimal number or percentage ranging between 0 and 1 or 0% and 100%.

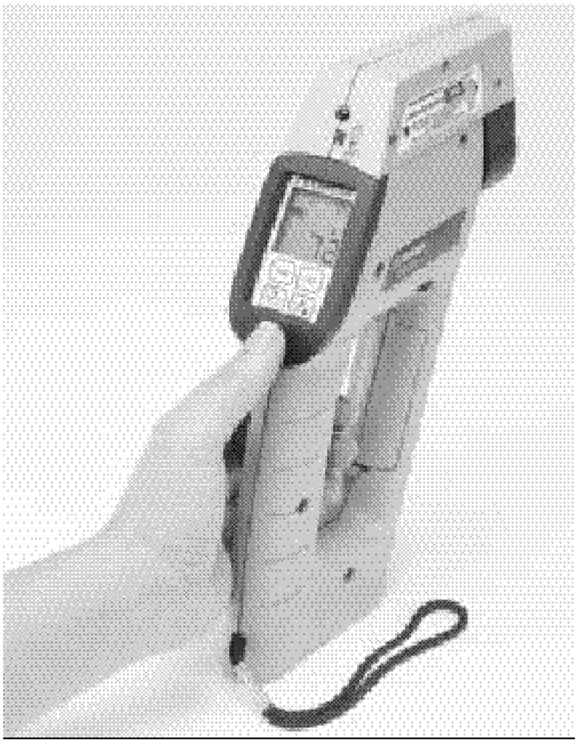


Figure 7.5 An IR type temperature measuring device

Most organic materials, including skin, are very efficient, frequently exhibiting emissivity of 0.95. Most polished metals, on the other hand, tend to be inefficient radiators at room temperature, with emissivity or efficiency often 20% or less. To function properly, an infrared measurement device must take into account the emissivity of the surface being measured. This can often be looked up in a reference table. However, we have to bear in mind that tables cannot account for localized conditions such as oxidation and surface roughness. A sometimes practical way to measure temperature with an infrared technique when the emissivity level is not known is to “force” the emissivity to a known level, by covering the surface with masking tape (emissivity of 95%) or a highly emissive paint.

Some of the sensor inputs may well consist of energy that is not emitted by the equipment or material whose surface is being targeted. Instead, there may be some rays being reflected by that surface from other equipment or materials reaching the sensor.

Emissivity pertains to energy radiating from a surface, whereas “reflection” pertains to energy reflected from another source. Emissivity of an opaque material is an inverse indicator of its reflectivity – substances that are good emitters do not reflect much incident energy, and thus do not pose much of a problem to the sensor in determining surface temperatures. Conversely, when one measures a target surface with only, say, 20% emissivity, much of the energy reaching the sensor might be due to reflection from, e.g., a nearby furnace at some other temperature. In short, we have to be wary of hot, spurious reflected targets. An infrared device is like a camera, and thus covers a certain field of view. It might, for instance, be able to “see” a 1- degree visual cone or a 100- degree cone.

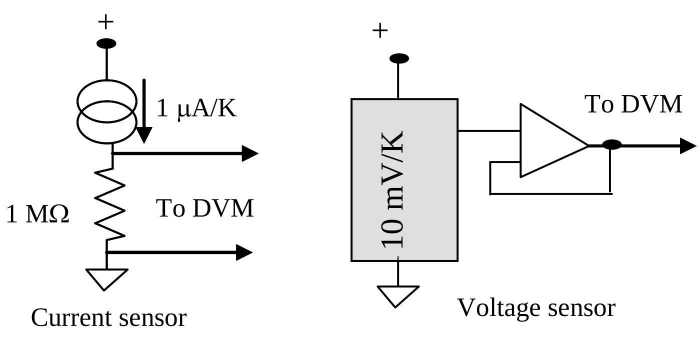


Figure 7.6 IC temperature sensors

Integrated Circuit (I.C.) Sensors

An innovation in thermometry is the integrated circuit temperature transducers shown in Figure 7.6. These are available in both voltage and current-output configurations. Both supply an output that is linearly proportional to absolute temperature. Typical values are 1 μA/K and 10 mV/K.

Some integrated sensors even represent temperature in a digital output format that can be read directly by a microprocessor. Except that they offer a very linear output with temperature, these IC sensors share all the disadvantages of thermistors. They are semiconductor devices and thus have a limited temperature range. The same problems of self-heating and fragility are evident and they require an external power source.

These devices provide a convenient way to produce an easy-to-read output that is proportional to temperature. Such a need arises in thermocouple reference junction hardware, and in fact these devices are increasingly used for thermocouple compensation.

Bimetallic Devices

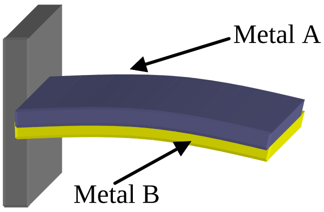


Figure 7.7 A bimetallic temperature sensor

Bimetallic devices take advantage of the difference in rate of thermal expansion between different metals. Strips of two metals are bonded together as illustrated in Figure 7.7. When heated, one side will expand more than the other, and the resulting bending is translated into a temperature reading by mechanical linkage to a pointer. These devices are portable and they do not require a power supply, but they are usually not as accurate as thermocouples or RTD's and they do not readily lend themselves to temperature recording.

Fluid-Expansion Devices

Typified by the household thermometer illustrated in Figure 7.8, fluid-expansion devices generally come in two main classifications:

- The mercury type, and
- The organic-liquid type.

Versions employing gas instead of liquid are also available.

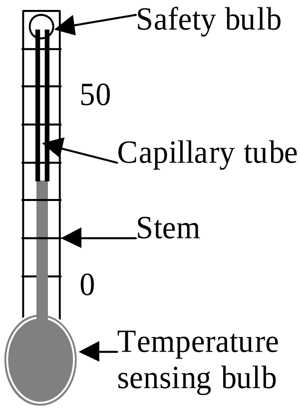


Figure 7.8 A mercury thermometer

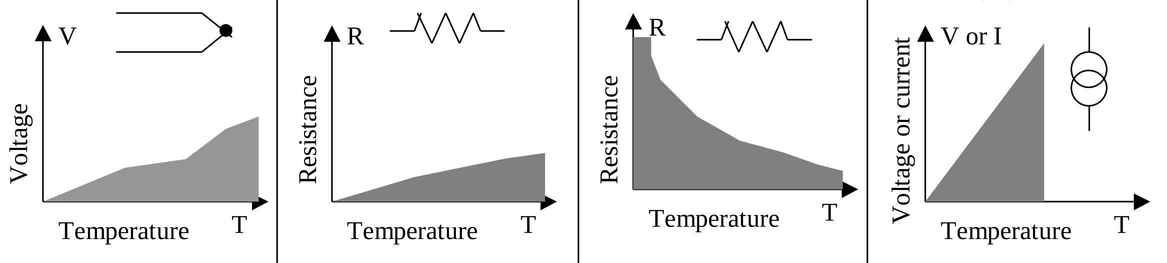
Mercury is considered an environmental hazard, so there are regulations governing the shipment of devices that contain it. Fluid-expansion sensors do not require electric power, do not pose explosion hazards, and are stable even after repeated cycling. On the other hand, they do not generate data that are easily recorded or transmitted, and they cannot make spot or point measurements.

Chemical (Change-of-State) Sensors

Change-of-state temperature sensors consist of labels, pellets, crayons, lacquers or liquid crystals whose appearance changes when a certain temperature is reached. They are used, for instance, with steam traps – when a trap exceeds a certain temperature, a white dot on a sensor label attached to the trap will turn black. Response time typically takes minutes, so these devices often do not respond to transient temperature changes, and accuracy is lower than other types of sensors. Furthermore, the change in state is irreversible, except in the case of liquid-crystal displays. Even so, change-of-state sensors can be handy when one needs confirmation that the temperature of a piece of equipment or a material has not exceeded a certain level, for instance for technical or legal reasons, during product shipment

Comparison of Practical Temperature Measurement Devices

The four most common temperature transducers are thermocouples, resistance-temperature detector's (RTD's), thermistors, and integrated circuit sensors. Their characteristics are shown and advantages and disadvantages are tabulated in Figure 7.9.



	Thermocouple	RTD	Thermistor	I.C. Sensor
Advantages	• Self powered • Simple • Rugged • Inexpensive • Wide variety of physical forms • Wide temperature range	• Most stable • Most accurate • More linear than thermocouple	• High output • Fast • Two-wire ohmic measurement	• Most linear • Highest output • Inexpensive
Disadvantages	• Non-linear • Low voltage • Reference required • Least stable • Least sensitive	• Expensive • Slow • Current source required • Small resistance change • Four-wire measurement	• Non-linear • Limited temperature range • Fragile • Current source required • Self-heating	• T < 250°C • Power supply required • Self-heating • Limited configurations

Figure 7.9 Comparison of four temperature measurement devices

TEMPERATURE MEASUREMENT USING THERMOCOUPLES

Principle of Operation

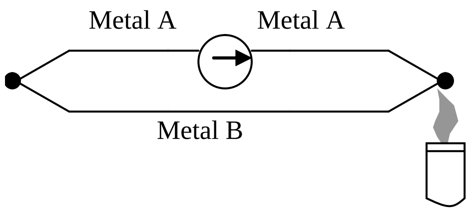


Figure 7.10 The thermoelectric circuit

When two wires composed of dissimilar metals are joined at both ends and one of the ends is heated, there is a continuous current which flows in the *thermoelectric* circuit as shown in Figure 7.10. This is called the Seebeck effect.

If this circuit is broken at the center as shown in Figure 7.11, the net open circuit voltage (the Seebeck voltage) is a function of the junction temperature and the composition of the two metals. All dissimilar metals exhibit this effect. For small changes in temperature the Seebeck voltage is linearly proportional to temperature: $V_{AB} = \alpha T$, where α, the Seebeck coefficient, is the constant of proportionality. (For real world thermocouples, α is not constant but varies with temperature.) If a voltage is applied, then there will be temperature change at the junction. This is called the Peltier effect and can be used for heating and cooling (refrigeration).

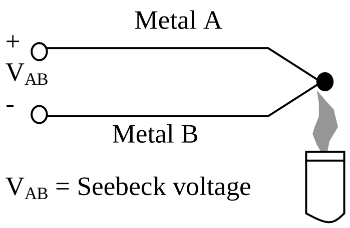


Figure 7.11 The Seebeck voltage

There is second effect that generates voltage and it is the temperature gradient along a single conductor as illustrated in Figure 7.12. The net e.m.f. due to this effect is proportional to the difference between the squares of the absolute junction temperatures. In this case, the thermocouple voltage is actually generated by the section of wire that contains a temperature gradient, and not necessarily by the junction. For example, if we have a thermal probe located in a molten metal bath, there will be two regions that are virtually isothermal and one that has a large gradient.

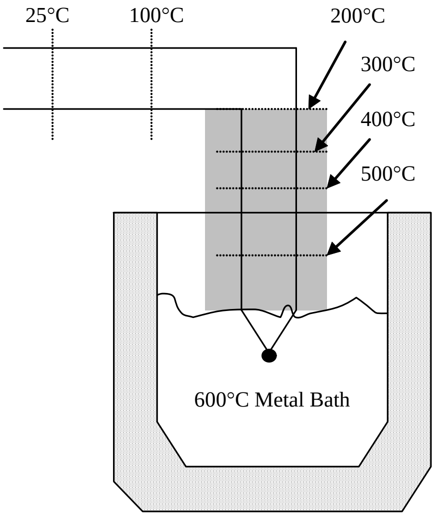


Figure 7.12 Temperature gradient along the wires

In Figure 7.12, the thermocouple junction will not produce any part of the output voltage. The shaded section will be the one producing virtually the entire thermocouple output voltage. If, due to aging or annealing, the output of this thermocouple had been found to be drifting, replacing only the thermocouple junction would not solve the problem. We would have to replace the entire shaded section, since it is the source of the thermocouple voltage.

The output voltage "V" of a simple thermocouple (with a reference temperature T_0 = 0°C = 32°F) is:

$$V = AT + \frac{1}{2}BT^2 + \frac{1}{3}CT^3 \quad \text{volts,}$$

where T is the temperature of the measuring junction in °C, A, B, and C are constants that depend upon the thermocouple material. The sensitivity

$$S = \frac{\partial V}{\partial T} = A + BT + CT^2 \quad \text{volt/°C}$$

Empirical Laws of Thermocouples

The "laws" governing the operation of the thermocouple are obtained experimentally. They are exemplified below and are useful in understanding and diagnosing thermocouple circuits. Examples below assume the measurement wires are homogeneous; that is, free of defects and impurities. The isothermal block is an electrical insulator, but a good heat conductor.

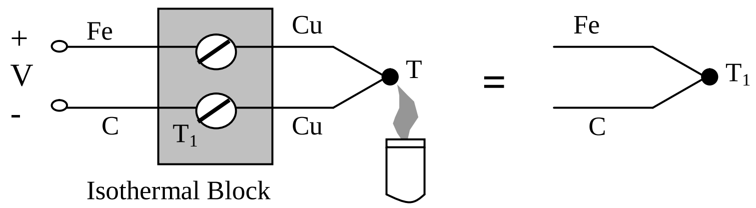


Figure 7.13 Law of intermediate metals

Law of Intermediate Metals

Inserting the copper lead between the iron and constantan (a metal alloy with %60 copper and %40 nickel) leads will not change the output voltage V, regardless of the temperature of the copper lead. The voltage V is that of a Fe-C thermocouple at temperature T_1 as illustrated in Figure 7.13.

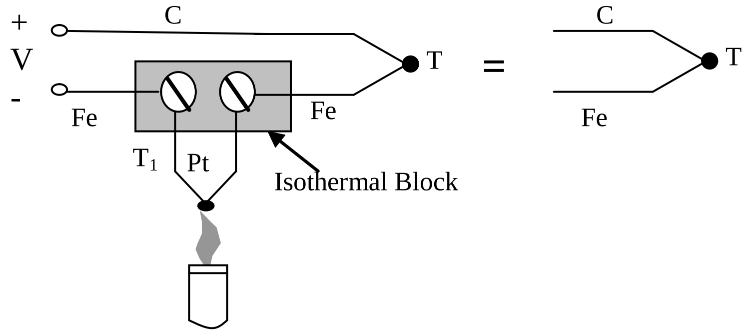


Figure 7.14 Law of inserted metals

Law of Interior Temperatures

The output voltage V will be that of a Fe-C couple at temperature T, regardless of the external heat source applied to either measurement lead. This is illustrated in Figure 7.14.

Law of Inserted Metals

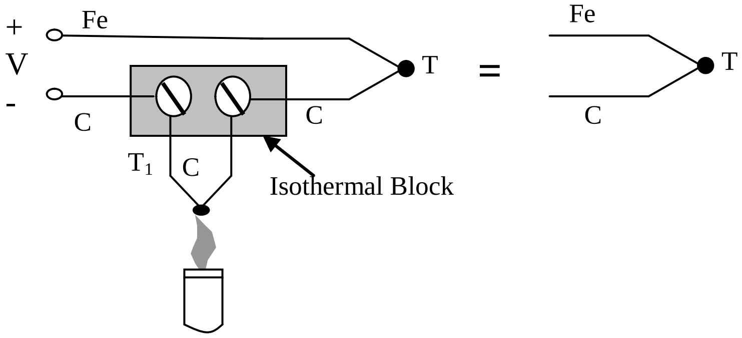


Figure 7.15 Law of inserted metals

The voltage V will be that of a Fe-C thermocouple at temperature T, provided both ends of the platinum wire are at the same temperature. The two thermocouples created by the platinum wire (Fe-Pt and Pt -Fe) act in opposition as shown in Figure 7.15.

Measuring Thermocouple Voltage with a Digital Voltmeter (DVM)

We can't measure the Seebeck voltage directly because we must first connect a voltmeter to the thermocouple, and the voltmeter leads, themselves, create a new thermoelectric circuit. Let's connect a voltmeter across a copper-constantan (Type T) thermocouple and look at the voltage output.

We would like the voltmeter to read only V_1, but by connecting the voltmeter in an attempt to measure the output of Junction J_1 we have created two more metallic junctions: J_2 and J_3. Since J_3 is a copper-to-copper junction, it generates no thermal e.m.f. ($V_3 = 0$) but J_2 is a copper-to-constantan junction which will add an e.m.f. (V_2) in opposition to V_1 . The resultant voltmeter reading V will be proportional to the temperature *difference* between J_1 and J_2 as illustrated in Figure 7.16. This implies that we can't find the temperature at J_1 unless we first find the temperature of J_2.

The Reference Junction

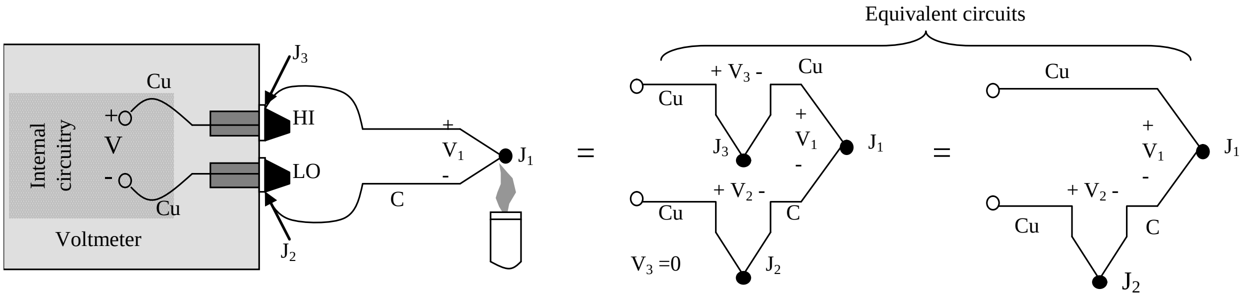


Figure 7.16 Model for measuring the thermocouple voltage with a DVM

External Reference Junction

One way to determine the temperature J_2 is to physically put the junction into an ice bath, forcing its temperature to be 0°C and establishing J_2 as the *Reference Junction* as illustrated in Figure 7.17. Since both voltmeter terminal junctions are now copper-copper, they generate no thermal e.m.f. and the reading V on the voltmeter is proportional to the temperature *difference* between J_1 and J_2 . Now the voltmeter reading is: $V = (V_1 - V_2) = \alpha(Tj_1 - Tj_2)$

If we specify Tj_1 in degrees Celsius: $Tj_1(°C) + 273.15 = Tj_1\,(K)$

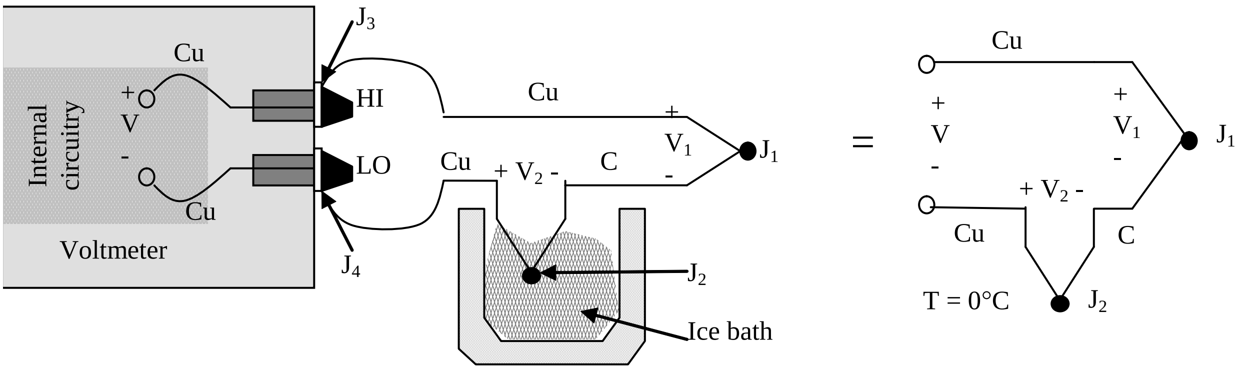


Figure 7.17 Temperature measurement using an ice bath to keep the reference junction

Then the equation can be rewritten and *V* becomes:

$V = V_1 - V_2 = \alpha\,[(Tj_1(°C) + 273.15) - (Tj_2(°C) + 273.15)] = \alpha(Tj_1(°C) - Tj_2(°C)) = \alpha(Tj_1(°C) - 0(°C))$ yielding $V = \alpha Tj_1(°C)$

We use this derivation to emphasize that the ice bath junction output V_2 is not zero volts. It is a function of absolute temperature.

By adding the voltage of the ice point reference junction, we have now referenced the reading V to 0°C. This method is very accurate because the ice point temperature can be precisely controlled. The ice point is used by the National Institute of Standards and Technology (NIST) as the fundamental reference point for their thermocouple tables, so we can now look at the NIST tables and directly convert from voltage V to Temperature Tj_1(°C).

The Iron-Constantan Couple

The copper-constantan thermocouple shown is a unique example because the copper wire is the same metal as the voltmeter terminals.

Let's use an iron-constantan (Type J) thermocouple instead of the copper-constantan (Type T) as shown in Figure 7.18. The iron wire increases the number of dissimilar metal junctions in the circuit,

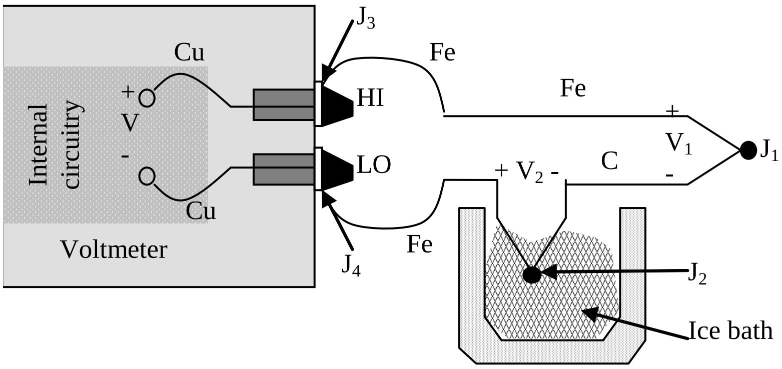


Figure 7.18 Temperature measurement using an iron-constantan couple

as both voltmeter terminals become Cu-Fe thermocouple junctions.

Junction Voltage Cancellation

$V_1 = V$ if $V_3 = V_4$, i.e. $T_{j3} = T_{j4}$

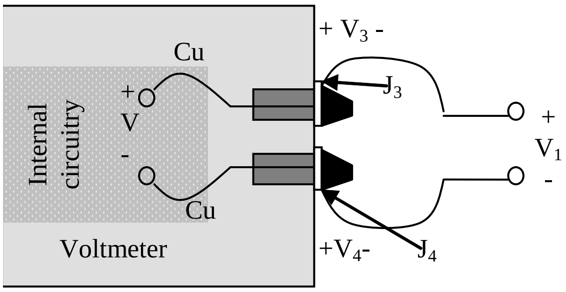


Figure 7.19 Junction voltage cancellation

This circuit provides moderately accurate measurements as long as the voltmeter *high* and *low* terminals (J_3 & J_4) shown in Figure 7.19 act in opposition. If both front panel terminals are not at the same temperature, there will be an error. For a more precise measurement, the copper voltmeter leads should be extended so the copper-to-iron junctions are made on an isothermal (same temperature) block.

Removing Junctions from the DVM Terminals

The isothermal block is an electrical insulator but a good heat conductor and it serves to hold J_3 and J_4 at the same temperature. The absolute block temperature is unimportant because the two Cu-Fe junctions act in opposition. In this way, the junctions are removed from the DVM terminals as illustrated in Figure 7.20.

Reference Circuit: External Reference Junction – No Ice Bath

The circuit described in the previous section will give us accurate readings, but it would be nice to eliminate the ice bath if possible.

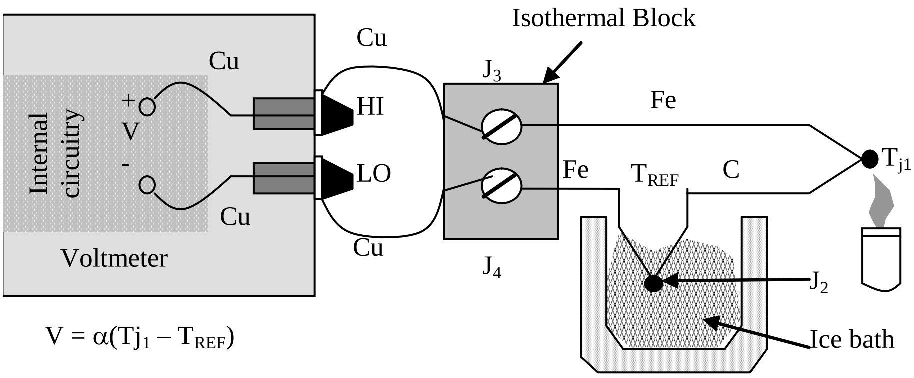


Figure 7.20 Removing junctions from the DVM terminals

Eliminating the Ice Bath Using Isothermal Blocks

Let's replace the ice bath with another isothermal block as shown in Figure 7.21. The new block is at Reference Temperature T_{REF} , and because J_3 and J_4 are still at the same temperature we can again show that: $V = \alpha(T_1 - T_{REF})$

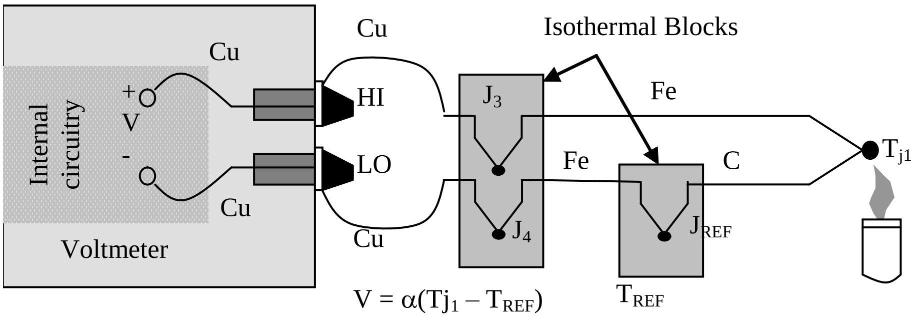


Figure 7.21 Eliminating ice bath using isothermal blocks

This is still a rather inconvenient circuit because we have to connect two thermocouples. Let's eliminate the extra Fe wire in the negative (LO) lead by combining the Cu-Fe junction (J_4) and the Fe-C junction (J_{REF}).

Joining the Isothermal Blocks

We can do this by first joining the two isothermal blocks as shown in Figure 7.22. We haven't

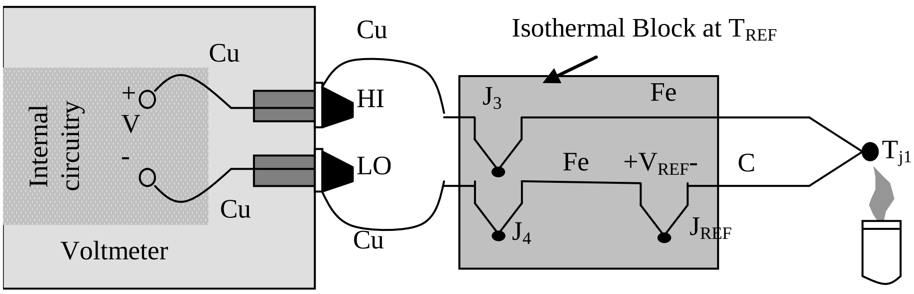


Figure 7.22 Joining isothermal blocks

changed the output voltage V. It is still: $V = \alpha(T_1 - T_{REF})$

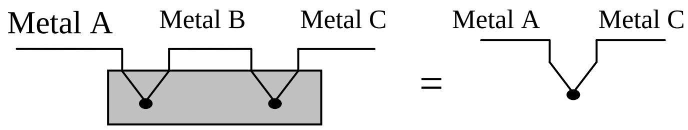


Figure 7.23 A way of using law of inserted metals

Now we call upon the law of intermediate metals to eliminate the extra junction as illustrated in Figure 7.23. This empirical law states that a third metal (in this case, iron) inserted between the two dissimilar metals of a thermocouple junction will have no effect upon the output voltage as long as the two junctions formed by the additional metal are at the same temperature.

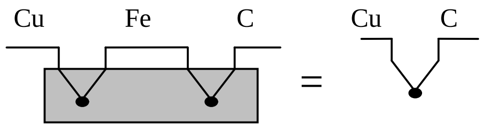


Figure 7.24 Eliminating Fe junction

This is a useful conclusion, as it completely eliminates the need for the iron (Fe) wire in the LO lead as shown in Figure 7.24. Again $V = \alpha(T_1 - T_{REF})$

where "α" is the Seebeck coefficient for a Fe-C thermocouple. Junctions J_3 and J_4 take the place of the ice bath. These two junctions are combined to become the *reference junction*.

External Reference Junction – No Ice Bath

Software Compensation

Now we can proceed to the next logical step: Directly measure the temperature of the isothermal block (the reference junction) and use that information to compute the unknown temperature, Tj_1 as illustrated in Figure 7.25.

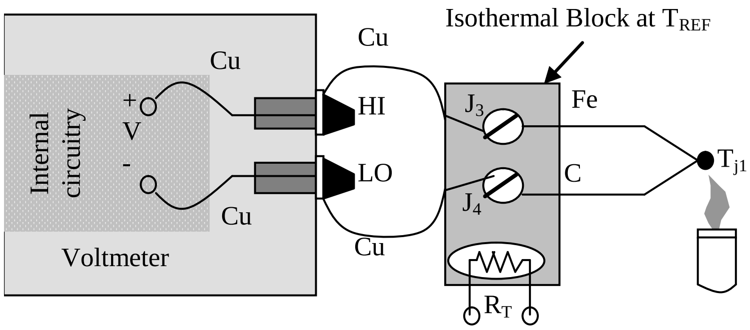


Figure 7.25 External reference junction without ice bath

A thermistor, whose resistance R_T is a function of temperature, provides us with a way to measure the absolute temperature of the reference junction. Junctions J_3 and J_4 and the thermistor are all assumed to be at the same temperature, due to the design of the isothermal block. Using a digital multimeter (DMM), we simply:

- Measure R_T to find T_{REF} and convert T_{REF} to its equivalent reference junction voltage, V_{REF}
- Measure V and add V_{REF} to find V_1 and convert V_1 to temperature Tj_1 .

This procedure is known as *software compensation* because it relies upon software in the instrument or a computer to compensate for the effect of the reference junction. The isothermal terminal block temperature sensor can be any device, which has a characteristic proportional to absolute temperature; an RTD, a thermistor, or an integrated circuit sensor.

Hardware Compensation

Rather than measuring the temperature of the reference junction and computing its equivalent voltage as we did with software compensation, we could insert a battery to cancel the offset voltage of the reference junction as illustrated in Figure 7.26. The combination of this hardware compensation voltage and the reference junction voltage is equal to that of a 0°C junction.

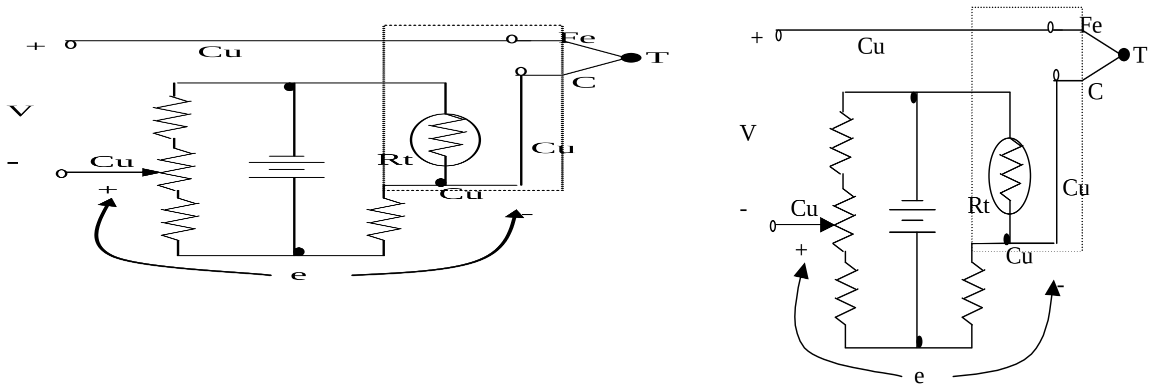


Figure 7.26 Hardware compensation of the thermocouple junction

The compensation voltage, e, is a function of the temperature sensing resistor, R_T. The voltage V is now referenced to 0°C, and may be read directly and converted to temperature by using the NIST tables.

Why Thermocouple is Used?

Ease and Reliability in Application

It seems logical to ask: If we already have a device that will measure absolute temperature (like an RTD or thermistor), why do we even bother with a thermocouple that requires reference junction compensation? The single most important answer to this question is that the thermistor, the RTD, and the integrated circuit transducer are only useful over a certain temperature range. Thermocouples, on the other hand, can be used over a range of temperatures, and optimized for various atmospheres. They are much more rugged than thermistors, as evidenced by the fact that thermocouples are often welded to a metal part or clamped under a screw. They can be manufactured on the spot, either by soldering or welding. In short, thermocouples are the most versatile temperature transducers available and since the measurement system performs the entire task of reference compensation and software voltage-to-temperature conversion, using a thermocouple becomes as easy as connecting a pair of wires.

Monitoring Large Number of Data Points

Thermocouple measurement becomes especially convenient when we are required to monitor a large number of data points. This is accomplished by using the isothermal reference junction for more than one thermocouple element as shown in Figure 7.26. A relay scanner connects the voltmeter to the various thermocouples in sequence. All of the voltmeter and scanner wires are copper, independent of the type of thermocouple chosen. In fact, as long as we know what each thermocouple is, we can mix thermocouple types on the same isothermal junction block (often called a *zone box*) and make the appropriate modifications in software. The junction block temperature sensor, R_T is located at the center of the block to minimize errors due to thermal gradients. Software compensation is the most versatile technique we have for measuring thermocouples. Many thermocouples are connected on the same block, copper leads are used throughout the scanner, and the technique is independent of the types of thermocouples chosen. In addition, when using a data acquisition system with a built-in zone box, we simply connect the thermocouple as we would a pair of test leads. All of the conversions are performed by the instrument’s software. The one disadvantage is that it requires a small amount of additional time to calculate the reference junction temperature. For maximum speed we can use hardware compensation.

Series and Parallel Connection of Thermocouples

An arrangement of multiple-junction thermocouples is referred to as a *thermopile*. Increased sensitivity may be achieved by connecting a number of thermocouples in series, all of them measure the same temperature and using the same reference junction. Parallel combinations may be used to measure average temperature.

Examples for Thermocouple and Temperature Measurement

Table 7.1. Data for commonly used thermocouples

Temperature		emf (mV) with reference at 0 °C				
°C	°F	T	E	J	K	S
-184.4	-300	-5.284	-8.30	-7.52	-5.51	
	-250	-7.747		-6.71	-4.96	
-128.9	-200	-4.111	-6.40	-5.76	-4.29	
	-150	-3.380		-4.68	-3.52	
-73.3	-100	-2.559	-3.94	-3.49	-2.65	
	-50	-1.654		-2.22	-1.70	
-17.78	0	-0.670	-1.02	-0.89	-0.68	
	50	0.389		0.50	0.40	
37.78	100	1.517	2.77	1.94	1.52	0.221
	150	2.711		3.41	2.66	0.401
93.33	200	3.967	5.87	4.91	3.82	0.595
	250	5.280		6.42	4.97	0.800
148.9	300	6.647	9.71	7.94	6.09	1.017
	350	8.064		9.48	7.20	1.242
204.4	400	9.525	13.75	11.03	8.31	1.474
	450	11.030		12.57	9.43	1.712
260	500	12.575	17.95	14.12	10.57	1.956
	600	15.773	22.25	17.18	12.86	2.458
371.1	700	19.100	26.65	20.26	15.18	2.977
	800		31.09	23.32	17.53	3.506
537.8	1000		40.06	29.52	22.26	4.596
	1200		49.04	36.01	26.98	5.726
815.6	1500		62.30		33.93	7.498
	1700		70.90		38.43	8.732
1093	2000				44.91	10.662
	250				54.92	13.991
1649	3000					17.292

Example 7.1

For the configuration shown:

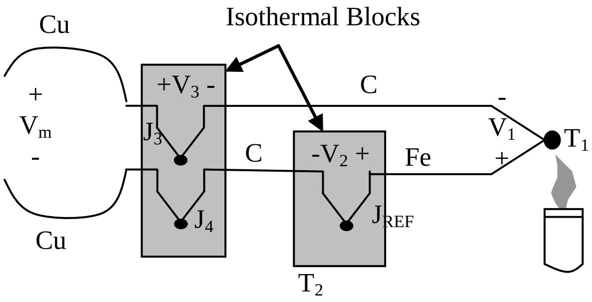


Figure for example 7.1.

Find the voltage across the measuring junction and sensitivity at T_1 = 260 °C

- Linear interpolation using the data for J type thermocouple: V_1 = 14.12 mV
- Sensitivity before 260°C = 55.56μV/°C
- Sensitivity after 260°C = 55.27μV/°C
- Average of the two = 55.42μV/°C
- Find the voltage across J_3 at T_3 = 25 °C
- From the data for T type V_3 = 1.517 mV
- Find the output voltage for T_2 = 0 °C
- $V_m = - V_1 = -14.12$ mV

Assume now the isothermal blocks are combined and kept at T= 25 °C. Find the output voltage in this condition and sensitivity of the output voltage to $T_2 = T_{Ref}$.

V_2 = 1.94x25/37.78 = 1.284mV; $V_m = -V_1 + V_2$ = -12.836 mV

Example 7.2

Assume that you can add a battery in series with the loop in the following circuit.

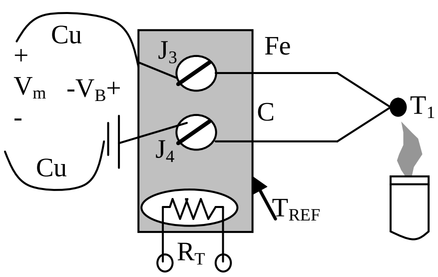


Figure for example 7.2.

How much is the required voltage to have the output voltage is V_1 only at the reference junction is kept at 25 °C?

$V_m = V_1 - V_2 + V_B = V_1$ Hence, $V_B = V_2$ =1.2.84 mV

R_T is a resistance type temperature sensor. $R_T = R_0[1+\alpha(T-T_0)]$ where R_0 = 120 Ω at T= 25 °C, α = 4x10^{-4} /°C. Design a temperature measurement set-up around R_T that produces an output voltage equivalent to V_B and has the same sensitivity to temperature variations at the reference junction as the output voltage in the previous problem. So, the circuit can replace the battery.

We can form a Wheatstone bridge and place $R_T = R_4$. The sensitivity of V_m to T_2 is:

$$\frac{\partial V_m}{\partial T_2} = -\alpha_2 = -\frac{1.94}{37.78} = -51.35\mu V/°C .$$

The bridge output V_b must have this sensitivity. At the same time $V_b\big|_{T=0°C} = 0V$.

$$V_b = E_b\left(\frac{R_4}{(R_1+R_4)} - \frac{R_3}{(R_2+R_3)}\right); \quad \frac{\partial V_b}{\partial R_4} = E_b\frac{R_1}{(R_1+R_4)^2} \quad \text{and} \quad \frac{\partial R_4}{\partial T}\bigg|_{T=25°C} = \alpha R_0 = 0.048\Omega/°C ;$$

$$\frac{\partial V_b}{\partial T} = \frac{\partial V_b}{\partial R_4}\frac{\partial R_4}{\partial T} = 51.35 \quad \text{yielding} \quad \frac{\partial V_b}{\partial R_4} = \frac{51.35}{0.048} = 1.07mV/\Omega = \frac{E_b R_1}{(R_1+R_4)^2} .$$

Let the value of $R_{40} = R_4$ at T=0°C. From the equation R_{40} = 118.8 Ω. To balance the bridge $R_2R_{40} = R_1R_3$. With $R_1 = R_2$ and $R_3 = R_{40}$ and taking E_b = 5 volts, the above equations can be solved simultaneously and yield

$R_1 = R_2$ = 4430 Ω.

Example 7.3

A thermopile is formed as shown in the figure.

The thermocouples are all of the same copper (Cu) - constantan (C) (Type T). The isothermal block is kept at the reference temperature T_0 = 0°C. The e.m.f. $E_{cu\text{-}C}$ (T,T_0) (mV) versus temperature (T) of copper – constantan thermocouple is given in the table. The output voltage E_T= 2.05 mV Calculate the e.m.f. for junctions (B) and (C); temperature of junction (A) and (C).

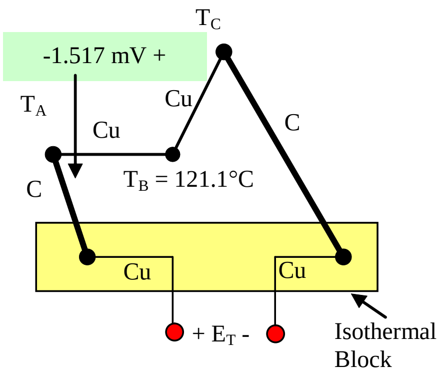


Figure for example 7.3

Table 7.2 Data for example 7.3

T(°C)	-128.9	-73.3	-17.78	37.78	93.33	148.9	204.4	260
E(mV)	-4.111	-2.559	-0.670	1.517	3.967	6.647	9.525	12.575

V_B = 0 V (Cu-Cu junction); V_A = -1.517 mV; $V_C = E_T – V_A$ = 2.05+1.517 = 3.567 mV

T_A = 37.78 °C (from the table). T_C can be found through interpolation as follows:

Two known points around the unknown temperature are marked on a graph a a linear relation is expected in between. Then, from the similarities of triangles;

$$\frac{T_C - 37.78}{93.33 - 37.78} = \frac{3.567 - 1.517}{3.967 - 1.517}$$

yields T_C = 84.26°C

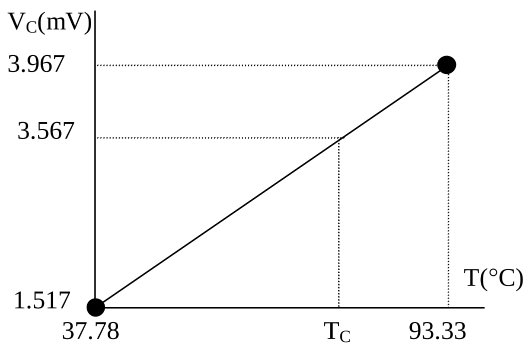


Figure for interpolation in example 7.3.

Example 7.4

It is required to measure temperature in the range 50-200 °C by means of a thermocouple having a sensitivity of 50 μV/°C ± 1.5%. The reference temperature T_0 = 0 ± 0.1 °C. The available millivoltmeter has uncertainty of ± 40 μV. Find the temperature and its uncertainty for an output of 2.5 mV and 10 mV.

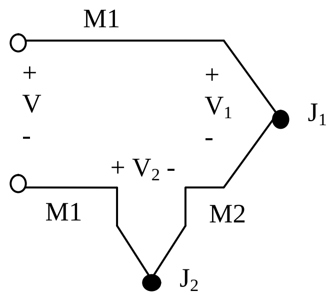


Figure for example 7.4.

$V = V_1 - V_2 = \alpha(T_1 - T_2)$; $T_2 = T_0 = 0 \pm 0.1$ °C and $\alpha = 50$ μV/°C ± 1.5%.

For V = 2.5 mV; $\bar{T} = \dfrac{2.5x10^{-3}}{50x10^{-6}} =$ 50°C; and

For V = 10 mV; $\bar{T} = \dfrac{10x10^{-3}}{50x10^{-6}} =$ 200°C

After rearranging the voltage equation:

$T = \dfrac{V}{\alpha} + T_0$ yielding $\dfrac{\partial T}{\partial V} = \dfrac{1}{\alpha}$; $\dfrac{\partial T}{\partial \alpha} = -\dfrac{V}{\alpha^2}$ and $\dfrac{\partial T}{\partial T_0} = 1$

$\Delta V = \pm 40$ μV; $\Delta T_0 = \pm 0.1$ °C and $\Delta\alpha = \pm 1.5x50/100$ (μV/°C) = ± 0.75 μV/°C.

The uncertainty equation: $(\Delta T)^2 = \left(\dfrac{\partial T}{\partial V}\right)^2 (\Delta V)^2 + \left(\dfrac{\partial T}{\partial \alpha}\right)^2 (\Delta\alpha)^2 + \left(\dfrac{\partial T}{\partial T_0}\right)^2 (\Delta T_0)^2$ yields

ΔT = 1.1 °C = 2.2% for V = 2.5 mV; and ΔT = 3.11 °C = 1.55% for V = 10 mV

TEMPERATURE MEASUREMENT USING THERMISTORS

Principle of Operation

An individual NTC type thermistor curve shown in Figure 7.2 can be very closely approximated through use of the Steinhart-Hart equation:

$$\frac{1}{T} = A + B(\ln R) + C(\ln R)^3$$

where; T = Kelvins, R = Resistance of the thermistor, and A,B,C = curve-fitting constants

A, B, and C are found by selecting three data points on the published data curve and solving the three simultaneous equations. When the data points are chosen to span no more than 100°C within the nominal center of the thermistor's temperature range, this equation approaches a rather remarkable ±0.02°C curve fit.

Somewhat faster computer execution time is achieved through a simpler equation:

$$T = \frac{1}{(\ln R) - A} - C$$

where A, B, and C are again found by selecting three (R,T) data points and solving the three resultant simultaneous equations. This equation must be applied over a narrower temperature range in order to approach the accuracy of the Steinhart-Hart equation.

Thermistors are usually designated in accordance with their resistance at 25° C. The most common of these ratings is 2252 ohms; among the others are 5,000 and 10,000 ohms. If not specified to the contrary, most instruments will accept the 2252 type of thermistor. The resistance of the thermistor (R_T) at a temperature T (K) can also be expressed in terms of its resistance R_0 at a reference temperature T_0 (K) as:

$$R_T = R_0 e^{\beta\left(\frac{T_0 - T}{TT_0}\right)}$$

where β is the material constant for thermistor, in kelvins (K).

The temperature coefficient can be found by differentiating the above equation as,

$$\alpha = \frac{1}{R_T}\left(\frac{dR_T}{dT}\right) = -\frac{\beta}{T^2}$$

and it indicates that α is temperature dependant and decreases with increasing temperature.

Thermistor Linearization

It is difficult to design a linear-reading thermometer due to the inherent non-linearity of the

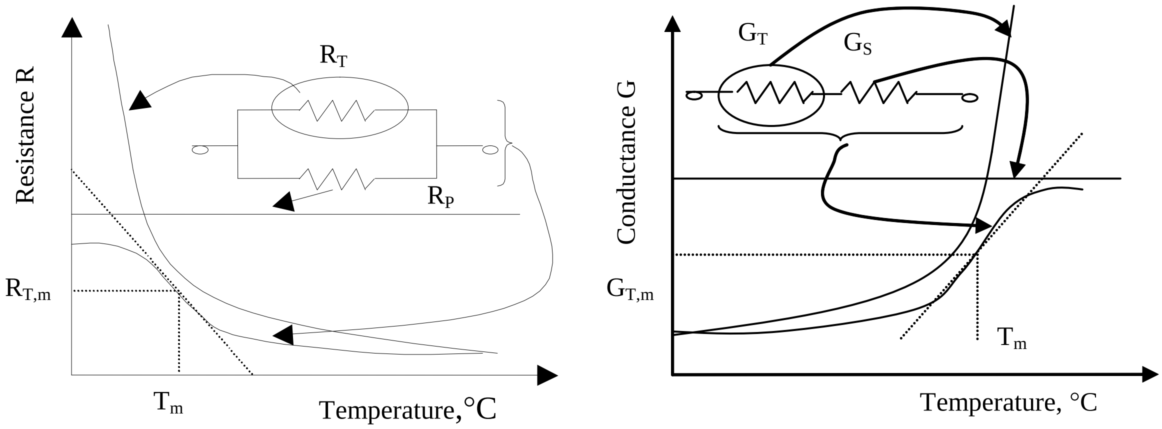


Figure 7.27 Thermistor linearization by shunt and series connected resistors

resistance-versus-temperature characteristics of thermistors. Approximate linearization can be achieved over a limited temperature range by adding series or parallel resistors to the thermistor as illustrated in Figure 7.27. Both characteristics can be approximated to straight lines around their turning (inflection) points at $T=T_m$. The shunt (parallel) compensation is used if the network is fed from a constant current source and the voltage across is measured. The series combination is the choice when a voltage is applied to the network and the current passing through is used to indicate the temperature.

Linearity of the temperature indication is achieved if the inflection point is set to the mid-range of the measurement. For medical applications, the range used is from 32°C to 42°C in general. The resolution however, is 0.1°C. Then, 37°C is taken as the midrange. The inflection point of any curve can be found by taking its second derivative and equating it to zero. Hence, differentiating the equations for the equivalent resistances twice and equating them to zero, we can calculate proper values of shunt and/or series resistors. This yields

$$R_P = R_{T,M}\left(\frac{\beta - 2T_M}{\beta + 2T_M}\right)$$

where $R_{T,M}$ is the resistance of the thermistor at the mid-scale temperature T_M (in Kelvin). In a similar manner

$$\frac{1}{R_P} = G_S = G_{T,M}\left(\frac{\beta - 2T_M}{\beta + 2T_M}\right)$$

where $G_{T,M}$ is the thermistor conductance at T_M.

The improved linearity comes with a decrease in the effective temperature coefficient of the combination that can be given by

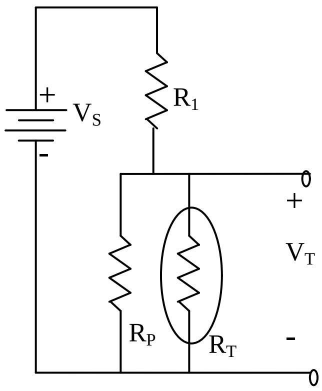
$$\alpha_{eff} = \frac{-\left(\beta / T_M^2\right)}{\left[\left(R_{T,M} / R_P\right) + 1\right]} \qquad \text{(Parallel)}$$

$$\alpha_{eff} = \frac{\left(\beta / T_M^2\right)}{\left[\left(G_{T,M} / G_S\right) + 1\right]} \qquad \text{(Series)}$$

It is reported that, with careful design, the maximum deviation from the linearity can be as low as 0.03°C for a ± 10°C span and 0.1°C for a span of ± 15°C. More complex circuit arrangements must be used for a better linearization over wider temperature ranges.

Thermistor Thermometry

In a thermistor thermometry, either the voltage across or the current through the network is used to indicate the temperature. Figure 7.28 shows conversion of temperature to voltage using a shunt



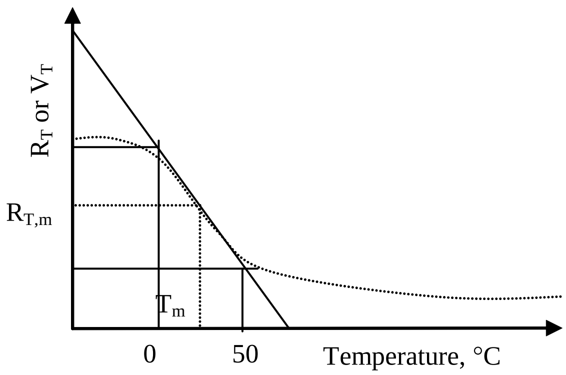


Figure 7.28 Converting temperature to voltage with a parallel-compensated thermistor

compensated thermistor. The characteristic of the equivalent resistance (R_{eff} = R_p//R_T) is shown as dashed line and it is linearized around T_M as indicated by the solid line. The output voltage of the circuit becomes

$$V_T = V_S \frac{R_{eff}}{R_1 + R_{eff}}$$

And with $R_1 >> R_{eff} = R_T//R_P$, the voltage can be computed using

$$V_T \cong \frac{V_S}{R_1} R_{eff}$$

that indicates a linear relationship between the voltage and the resistance.

The above equation doesn't yield a linear relationship between the temperature and the voltage. It

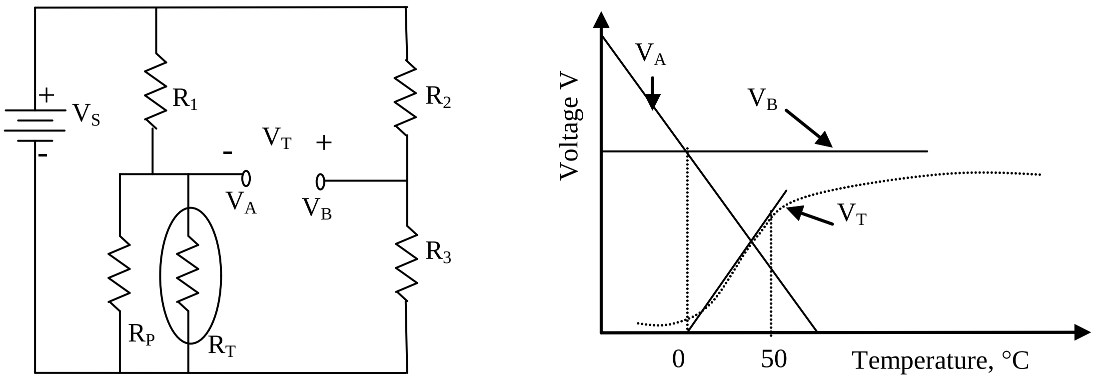


Figure 7.29 Converting temperature to voltage with a bridge network

may become linear around the mid-range if the voltage V_T is subtracted from $V_{T(0°C)}$. This can be easily managed using the bridge network shown in Figure 7.29. [1]The voltage V_A is the same as V_T in Figure 7.28. The balancing voltage

$$V_B = \frac{V_S R_3}{R_2 + R_3}$$

As this voltage is set to the value of V_A at 0°C, the bridge voltage $V_T = ST$, where S is the sensitivity of the system and T is the temperature in degree Celsius. The actual response is illustrated by the dashed-line in the figure. The error due to linearization increases as we go away from the mid-point. The sensitivity S around the mid-point is

$$S = \frac{V_S}{R_1} R_{eff} \alpha_{eff}$$

where R_{eff} and α_{eff} are the effective resistance and temperature coefficient for the shunt compensated thermistor.

[1] The bridge circuit will be discussed in detail in the next chapter. Readers who do not have prior familiarity with such circuits are recommended to read the related section of the next chapter first.

The series compensated thermistor can also be used to obtain an output voltage proportional to the temperature. An example is shown in Figure 7.30. The inverting terminal of the operational amplifier (op-amp) behaves as a virtual ground. The current through the thermistor is

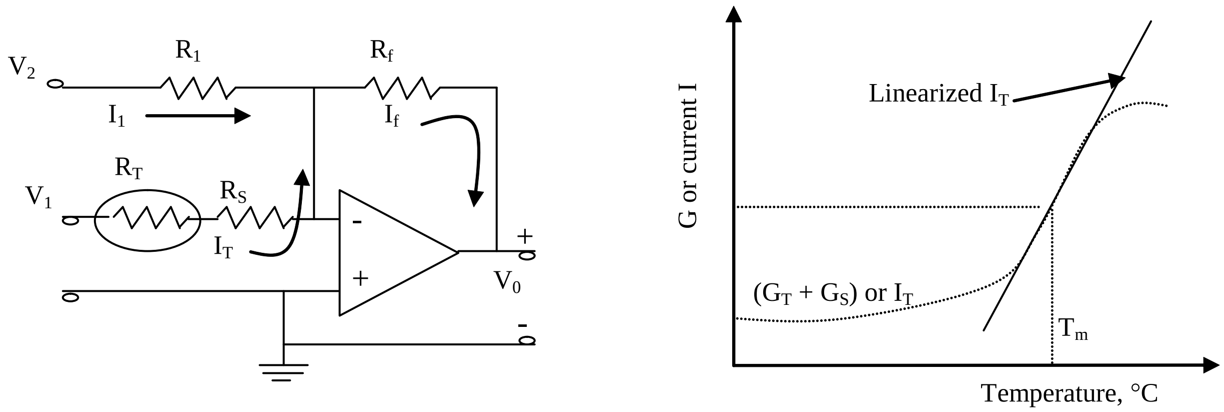


Figure 7.30 A thermometer based on a series compensated thermistor

$$I_T = \frac{V_1}{R_T + R_S} = V_1 G_{eff}$$

and it flows through the feedback resistor R_f together with the current I_1 yielding the output voltage,

$$V_0 = -R_f\left(I_1 + I_T\right) = -R_f\left(\frac{V_2}{R_1} + V_1 G_{eff}\right)$$

The sensitivity of the output voltage to temperature is (around the mid-range)

$$\frac{dV_0}{dT} = S = -R_f V_1 \frac{dG_{eff}}{dT} = -R_f V_1 G_{eff}\alpha_{eff}$$

where G_{eff} and α_{eff} are the effective conductance and temperature coefficient for the series compensated thermistor. S can be set to any value by adjusting the R_f and V_1. The output voltage can indicate the temperature in °C if V_2 and R_1 are selected to have

$$\frac{V_2}{R_1} = \left.\frac{-V_1}{R_S + R_T}\right|_{T=0°C}$$

PROBLEMS ON TEMPERATURE MEASUREMENTS

Review Questions

1. What is the temperature and how it can be used as an indicator of the heat energy?
2. What are the commonly used temperature scales and how they are related to each other?
3. What is the thermodynamic scale and how it is expressed?
4. What is the significance of a reference temperature?
5. What are the reference temperatures used in practice?
6. What are the commonly used temperature measuring devices?
7. What is a thermocouple and how it works?
8. What are the resistance temperature devices?
9. What is a thermistor and how the ntc and ptc types differ from each other?
10. What is the self-heating problem in thermistor thermometry?
11. What is the radiation detector (infrared sensor) and how it can be used for temperature measurement?
12. What are the integrated circuit (I.C.) sensors used for temperature measurement?
13. How a bimetallic device is used in temperature sensing?
14. What is the function of a bimetallic device in temperature sensing?
15. What are the fluid-expansion devices and how it can be used in temperature measurement?
16. What are the chemical (change-of-state) sensors and they are used in temperature measurement?
17. How can you compare and contrast practical temperature measurement devices?
18. How do you measure temperature using thermocouples?
19. What are the empirical laws of thermocouples?
20. How can you measure the thermocouple voltage using a digital voltmeter (DVM)?
21. Why is the reference junction is important in temperature measurement using thermocouples?.
22. How does a reference circuit replace the function of the reference junction?
23. How does the software compensation technique replace the function of the reference junction?
24. Why are thermocouples commonly used in temperature measurements?
25. Why are the thermistors used for temperature measurement although their characteristics are nonlinear?
26. How can you linearize thermistors?
27. How does the thermistor thermometry work?

Questions with Solutions

1. Resistance versus temperature characteristic of a typical thermistor is shown in the figure. The thermistor curve can be very closely approximated through use of the Steinhart-Hart equation:

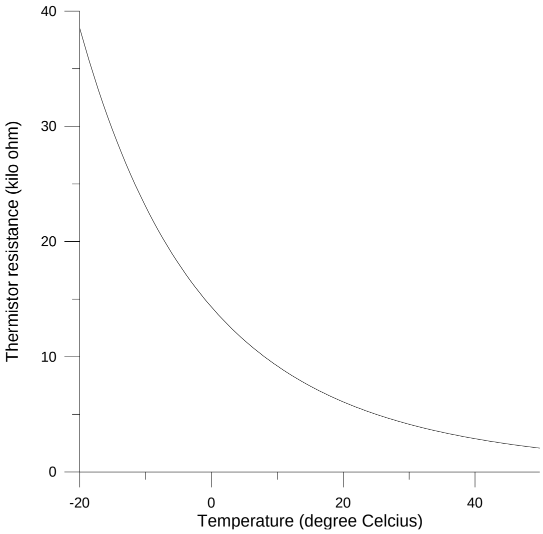


Figure for question 1

$$\frac{1}{T} = A + B(\ln R_T) + C(\ln R_T)^3$$

where; T = Kelvins, R_T = Resistance of the thermistor, and A,B,C = curve-fitting constants.

a. Show that the equation can be converted to

$$R_T = R_0 e^{\beta(\frac{T_0 - T}{T T_0})}$$

where R_T is the resistance of the thermistor at a temperature T (K) and R_0 is its resistance at a reference temperature T_0 (K) (assuming that the coefficient C in the previous equation is negligible)

Ans.

$$\frac{1}{T} = A + B(\ln R_T)$$

$$\frac{1}{T_0} = A + B(\ln R_0)$$

$$- \overline{\quad\quad\quad\quad\quad\quad\quad\quad}$$

$$\frac{1}{T} - \frac{1}{T_0} = B \ln \frac{R_T}{R_0}$$

Rearranging the equation and taking the exponential of both sides and letting β = 1/B yields the required result.

2. For a given thermistor β= 3420K and the resistance at 25°C is 5.00 kΩ ± 1%. The thermistor is used for a temperature measurement and the resistance measured is 2315 ± 4 Ω. Calculate the temperature and its uncertainty.

Ans. $\frac{1}{T} = \frac{1}{\beta} ln \frac{R_T}{R_0} + \frac{1}{T_0}$; 1/T = 0.003131, T = 319.43 K = 46 °C; we can use the "goal seek" function of the EXCEL as well.

Uncertainty in measuring the resistance is 400/2315 = 0.17%, uncertainty in R_T/R_0 is 1.17% that will be the uncertainty in T as well.

3. A thermopile is formed as shown in the figure. It has five junctions including the ones inside the isothermal block. Thermocouple data are given for copper-constantan (Cu-Con; type T) and iron-constantan (Fe-Con; type J) pairs in millivolt (mV) in the table. The isothermal block is at 25 °C. A thermal resistor is also placed into the isothermal block.

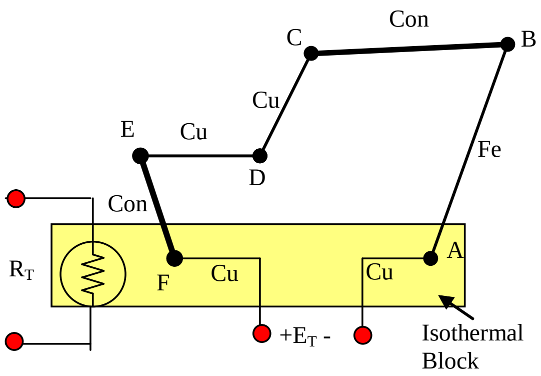


Figure for question 3

Temperatures at junctions C and D are 180 °C and 275 °C respectively. Voltages across junctions B and E are $V_B = V_{Fe\text{-}Con}$ =5.27 mV and $V_E = V_{Cu\text{-}Con}$ = 4.00 mV. Find the voltages across C [$V_C = V_{Cu\text{-}Con}$] and

Table for thermocouple data in question 3

T(°C)	-50	0	25	50	100	150	200	300	400
Cu-Con	-1.766	0	1.004	2.056	4.289	6.704	9.297	14.947	
Fe-Con	-2.40	0	1.28	2.59	5.27	8.00	10.79	16.33	27.428

D[$V_D = V_{Cu\text{-}Cu}$] and temperatures at B and E.

a. Find the voltages developed across junctions A($V_A = V_{Cu\text{-}Fe}$) and F ($V_F = V_{Cu\text{-}Con}$). (Hint: use the low of inserted metals for junction A.) Calculate the output voltage E_T.

b. A resistance temperature device is placed on the isothermal block. $R_T = R_0[1+\alpha(T-T_0)]$ where R_0 = 100Ω at T_0 = 0°C and α = 4x10^{-4}/°C. Calculate R_T and its sensitivity to T at the T=25°C.

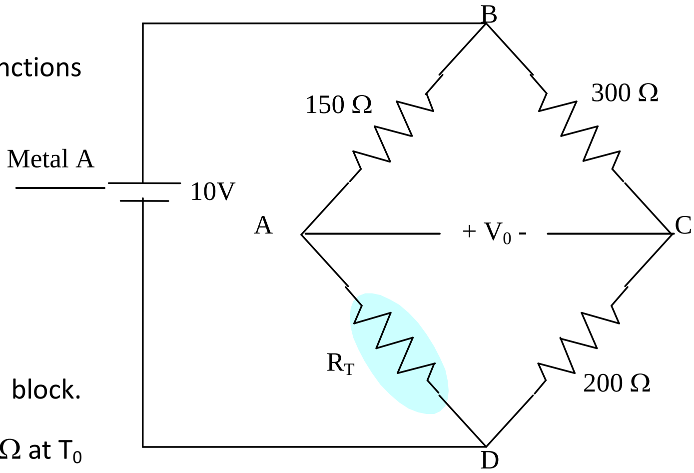


Figure for question 3-c

c. Assume R_T is placed into one arm of the Wheatstone bridge as shown in the figure. Calculate the bridge voltage at 0°C and 25°C.

Ans. For the thermopile:

$V_C = V_{Cu\text{-}Con}$ = (9.297-6.704)x30/50 + 6.704 = 8.26 mV; $V_D = V_{Cu\text{-}Cu}$ = 0 mV; T_B = 100°C and T_E = 50 + 50x(4.00-2.056)/(4.289-2.056) = 93.53 °C.

$V_A = V_{Cu\text{-}Fe} = V_{Cu\text{-}Con} + V_{Con\text{-}Fe}$ = 1.004 – 1.28 = -0.276 mV; $V_F = V_{Cu\text{-}Con}$ = 1.004 mV. E_T = 1.004 – 4.00 + 0 + 8.26 – 5.27 + 0.276 = 0.27 mV.

$R_T = R_0[1+\alpha(T-T_0)]$ where R_0 = 100Ω at T_0 = 0°C and $\alpha = 3.92 \times 10^{-4}$/°C.

R_T = 100 (1 + 0.01) = 101Ω and $\partial R_T/\partial T = \alpha R_0$ = 0.04 Ω/ °C.

$V_0 = E_b\left(\frac{R_T}{R_T+150} - \frac{200}{500}\right)$ yields 0 mV and 23.9 mV at 0°C and 25°C respectively.

General Questions

1. Discuss the problem of self-heating in resistance temperature devices.
2. For a thermocouple:
    a. State the empirical laws.
    b. Explain the cold junction and cold junction compensation briefly.
    c. What are the similarities and differences between bimetalic temperature sensors and thermocouples?
    d. It is required to measure temperature in the range 50-200 °C by means of a thermocouple having a sensitivity of 50 μV/°C ± 1.5%. The reference temperature T_0 = 0 ± 0.1 °C. The available millivoltmeter has uncertainty of ± 40 μV. Find the temperature and its uncertainty for an output of 2.5 mV and 10 mV.
3. A temperature measurement set-up using a resistance temperature sensor is shown.
    e. Write down an explicit formula relating V_0 to temperature T. $R = R_0\,[1 + \alpha\{T - T_0)]$
    f. Show that the indicated temperature is

$$T = T_0 + \frac{4V_0}{\alpha E_b}$$

if the effect of lead resistance R_l is ignored. V_0 is the bridge output voltage.

    g. Describe a resistance thermometer and explain a method for lead resistance (R_l) compensation.

h. In the shown resistance thermometer bridge, show that the actual temperature T is

$$T = T_0 + \frac{4V_0}{\alpha E_b} - \frac{2R_l}{\alpha R_0}$$

i. In a similar circuit α = 5x10^{-4}, R_0 = 100Ω, R_l = 0.020Ω, E_b = 10 V, T_0 = 0°C and V_0 = -0.1 V. Find the true and indicated temperatures and the percentage error due to lead resistance.

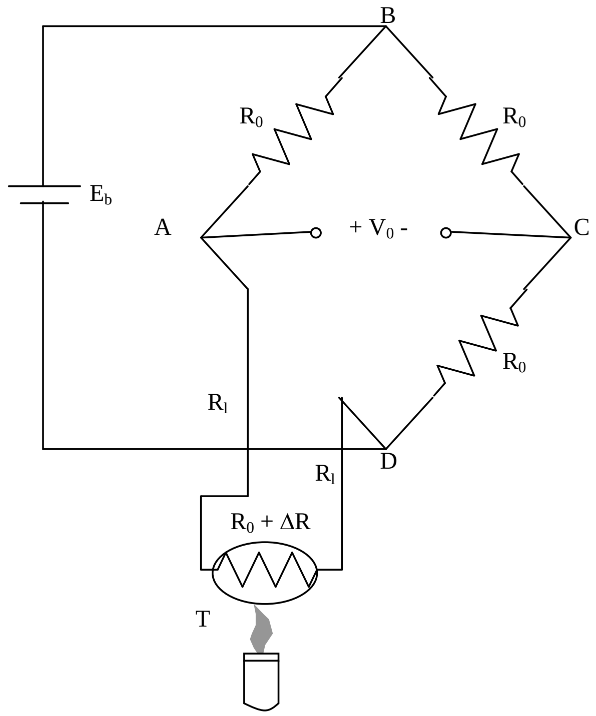


Figure for problem 3.

4. A metallic resistance thermometer has a linear variation of resistance with temperature

$$R = R_0 [1 + \alpha\{T - T_0\}]$$

The resistance R_0 at temperature T_0 = 280K ± 0.01K is found to be R_0 = 20 kΩ ± 0.1%, while at a temperature T the resistance is found to be R = 30 kΩ ± 0.1%. The coefficient α = 0.00392/K

j. Write an explicit expression for T.

k. Show that the uncertainty ΔT in T is given by:

$$(\Delta T)^2 = (\Delta T_0)^2 + \frac{1}{\alpha^2}\left(\frac{R}{R_0}\right)^2\left[\left(\frac{\Delta R_0}{R_0}\right)^2 + \left(\frac{\Delta R}{R}\right)^2\right]$$

l. Calculate the nominal value of T and its uncertainty.

m. Find the static sensitivity $\frac{\partial R}{\partial T}$ of the thermometer.

BIBLIOGRAPHY

Further Reading

Useful Websites

MEASUREMENT OF DISPLACEMENT AND MECHANICAL STRAIN

LEARNING OBJECTIVES

After completing this chapter, the students are expected to:

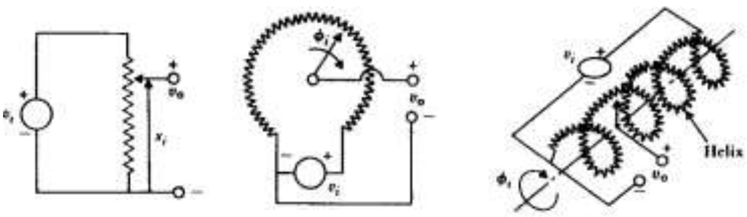
1. Describe displacement sensors.
2. Explain the resistive displacement sensors.
3. Describe inductive displacement sensors.
4. Illustrate the principles of capacitive sensors.
5. Discuss applications and limitations of piezoelectric sensors.
6. Express strain and stress as important mechanical measures.
7. Discuss mechanical principles of strain gages.
8. Explain changes in the electrical resistance of the strain gage wire.
9. Exemplify the use of strain gages.
10. Describe bonded and unbonded strain-gages.
11. Explain the effect of temperature and strain in other directions in displacement measurements.
12. Analyze the wheatstone bridge.
13. Discuss utilization of the wheatstone bridge.
14. Design circuits involving the wheatstone bridge.
15. Describe the null-mode and deflection-mode of operation of wheatstone bridges.
16. Describe mechanical connection of strain gages and arrangement of bridges for using a single, double and four active strain gages.
17. Discuss elimination of temperature and unwanted strain in the measurements using wheatstone bridges.
18. Recognize quantum tunneling composites.
19. Describe applications of novel sensors.

DISPLACEMENT SENSORS

Displacement is one of the major mechanical variables that is measured in many engineering applications. The displacement *x* is related to velocity and acceleration through differential /integral operations as velocity $v = dx/dt$ and acceleration $a = d^2x/dt^2$. It is converted into electrical current or voltage using resistive, inductive, capacitive and piezoelectric sensors and related circuitries. This chapter will brief the commonly used sensors for displacement and mechanical strain.

Resistive Sensors

Resistive sensors can be divided into two groups as potentiometers and strain gages. Potentiometers will be discussed below and strain gages will be treated in a separate section.



(a) (b) Rotational – (c) Helical

Figure 8.1 Potentiometer-type displacement transducers

Potentiometers are used for translational and rotational displacements as illustrated Figure 8.1. In a translational type potentiometer (a), the resistance between the wiper and the reference terminal

$R_i = kx_i$

and

$v_0 = v_sR_i/R = (kv_s/R)x_i$

R is the total resistance of the potentiometer and x_i is the displacement, provided that there is no instrument loading.

In the rotational type (b), the output voltage becomes proportional to the angular displacement ϕ_i. The resolution of the measurement depends upon the area covered by the wiper arm. The resolution can be improved by using helical multi-turn potentiometers as illustrated in (c).

Inductive Sensors

Inductance is defined as

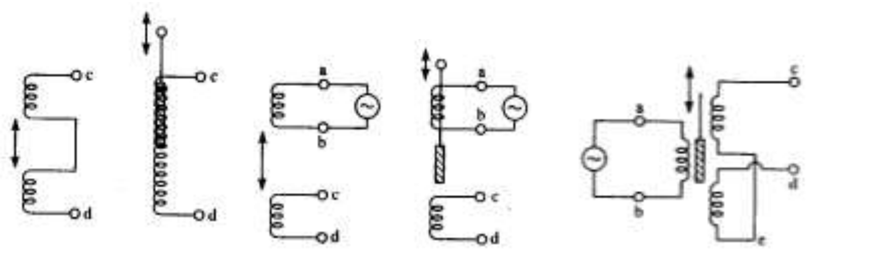
$L = n^2G\mu$

Where

n= number of turns of coil

G = geometric form factor

μ = effective permeability



(a) Self- (b) Mutual (c) Differential

Figure 8.2 Inductive-type displacement sensors

We can obtain a change in the inductance *L* by varying any one of the three defining parameters. The change can be induced as self-inductance (Figure 8.2(a)) and mutual inductance. The Linear Variable Differential Transformer (LVDT) shown in Figure 8.2(c)) is the mostly used inductive transducer. The input coil of the device is excited with an alternating voltage. The displacement of the core causes variation in the magnitude of the output voltage as illustrated in Figure 8.3. The output voltage is zero as the core is in the center. The magnitude of the output increases as the core moves away from the center. However, the increase is in phase with the input as the core travels up and out of phase as the core moves down. A phase sensitive demodulator decodes the signal and produces a voltage proportional to the displacement of the core.

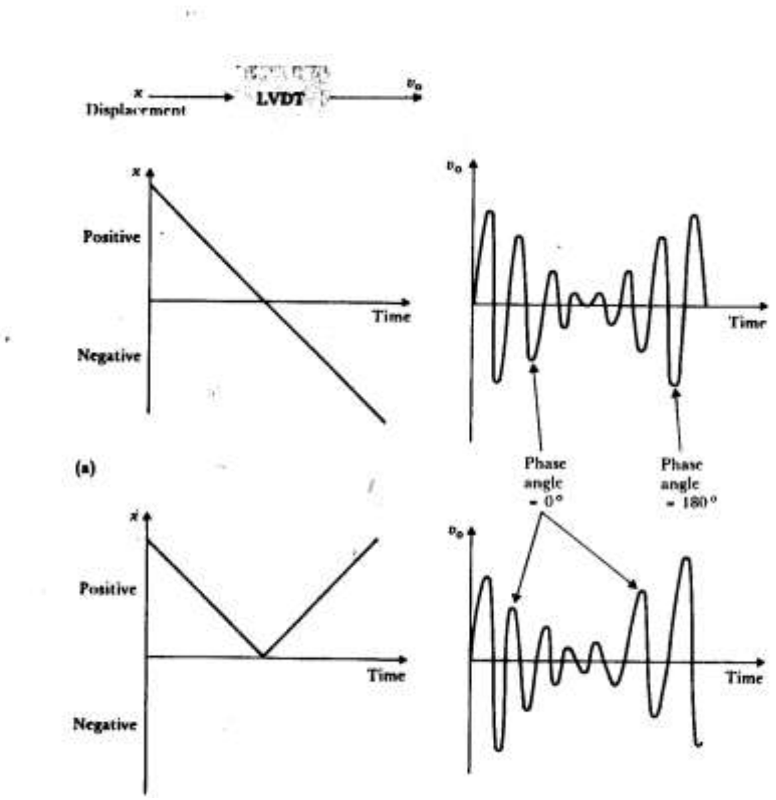
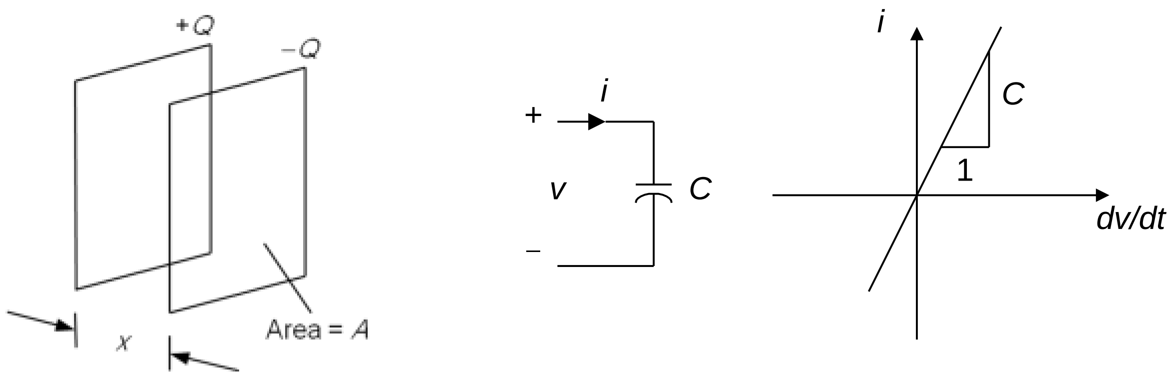


Figure 8.3 Characteristics of a LVDT

Capacitive Sensors



(a) Two parallel plates forming a capacitance

(b) Symbol of a capacitor

(c) Transfer characteristic of a capacitor

Figure 8.4 Capacitive type displacement sensor its symbol and characteristic

Capacitors store energy in the electrical field between two plates and the capacitance is defined by

$C = \varepsilon_0\varepsilon_r A/x$

Where

ε_0 = dielectric coefficient of the air

ε_r = relative dielectric coefficient of the medium between plates

A = Area common between plates

x = distance between plates

We can change the capacitance by changing any one of the defining parameters. In many applications, one of the capacitance plates is kept fixed while the other one can move. Sensitivity of the sensor for a displacement change (*x*) is defined as

$$sensitivity = K = \frac{\Delta C}{\Delta x} = -\varepsilon_0\varepsilon_r\frac{A}{x^2}$$

Yielding

$$\frac{dC}{dx} = -\frac{C}{x} \quad or \quad \frac{dC}{C} = -\frac{dx}{x}$$

The electrical charge in a capacitor is defined as

$$Q = CV$$

Where C is the capacitance in farad and V is the voltage in volt yielding the charge Q in coulomb. The current in the capacitor is the rate of change of the charge, that is

$$i = \frac{dQ}{dt} = V_1\frac{dC}{dt} + C\frac{dV_1}{dt} = -C\frac{E}{x_0}\frac{dx}{dt} + C\frac{dV_1}{dt}$$

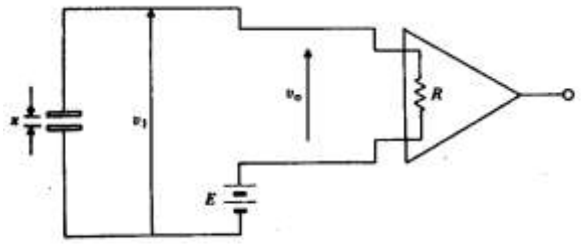


Figure 8.5 Capacitive sensor for measuring dynamic displacement changes

Figure 8.5 shows an application of the capacitive sensor in measuring dynamic displacement changes. The output voltage occurs across the input resistance of the amplifier. The sensor capacitance holds the excitation voltage *E* when there is no change in the displacement and the output voltage is zero. A current in the sensor is generated as the displacement x varies yielding an output voltage.

$$v_0 = -iR = v_1 - E \quad \text{and} \quad \frac{dv_0}{dt} = \frac{dv_1}{dt}$$

Combining the previous equations

$$v_0 = RC\frac{E}{x_0}\frac{dx}{dt} - RC\frac{dv_0}{dt}$$

Reorganizing the above yields the differential equation

$$RC\frac{dv_0}{dt} + v_0 = RC\frac{E}{x_0}\frac{dx}{dt}$$

The transfer function becomes

$$\frac{V_0(j\omega)}{X(j\omega)} = \frac{(E/x_0)\,j\omega\tau}{j\omega\tau + 1} \quad where \quad \tau = RC = R\varepsilon_0\varepsilon_r A/x_0$$

This is a characteristic of a high-pass filter. Hence, the sensor is useful at frequencies above the cut-off frequency of $\omega_C = \frac{1}{RC}$, *C* is the nominal capacitance of the sensor and *R* is the input resistance of the amplifier.

Piezoelectric Sensors

Certain crystals generate electrical charges as they are exposed to external forces as illustrated in Figure 8.6. The charge q is proportional to the applied force as

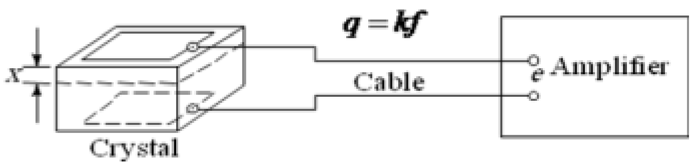


Figure 8.6 Symbolic representation of a piezoelectric crystal

$q = kf$

k being the piezoelectric constant in Coulomb/Newton. These sensors generate voltage outputs without requiring external electrical power supplies. Sensors discussed in previous sections have been passive devices that necessitate external electrical supplies for generating electrical outputs. The voltage across the opposite terminals of the device can be expressed as

$$v = \frac{kf}{C} = \frac{kfx}{\varepsilon_0 \varepsilon_r A}$$

The crystal can be modeled as a charge generator in parallel with a resistor and capacitor. The cable connecting the crystal to the amplifier behaves as a capacitor. The amplifier can be represented by an input capacitor in parallel with the input resistor. Figure 8.7 shows the overall equivalent circuit. The externally applied force causes a displacement x and the charge can be redefined in terms of this displacement as

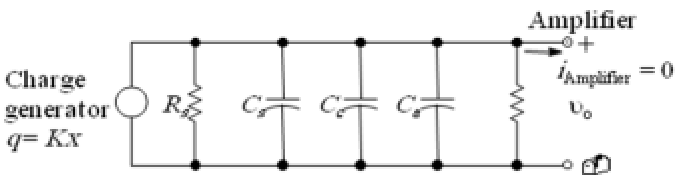


Figure 8.7 Model of the piezoelectric crystal

$q = Kx$

K being a new proportionality constant in Coulomb/meter.

The model can be simplified as shown in Figure 8.8 by combining the capacitive and resistive elements. Rate of change of the displacement is the velocity. The rate of change of the charge is the electrical current. Hence, the current coming out of the sensor is proportional to the velocity.

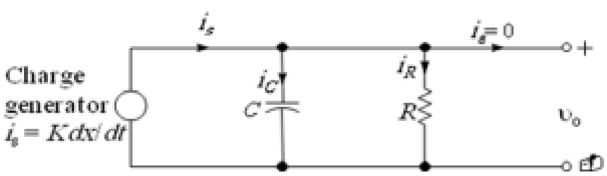


Figure 8.8 Simplified model of a piezoelectric crystal

$$i_s = \frac{dq}{dt} = K\frac{dx}{dt} = i_C + i_R$$

The voltage developed is

$$v_0 = v_C = (\frac{1}{C})\int i_C dt$$

The differential equation can be obtained from the previous two equations as

$$i_c = i_s - i_R = C(\frac{dv_0}{dt}) = K\frac{dx}{xt} - \frac{v_0}{R}$$

The equation leads to the transfer function

$$\frac{V_0(j\omega)}{X(j\omega)} = \frac{K_S j\omega\tau}{j\omega\tau + 1}$$

With $K_s = K/C$ (V/m) and $\tau = RC$ (s). This is a characteristic of a high-pass filter.

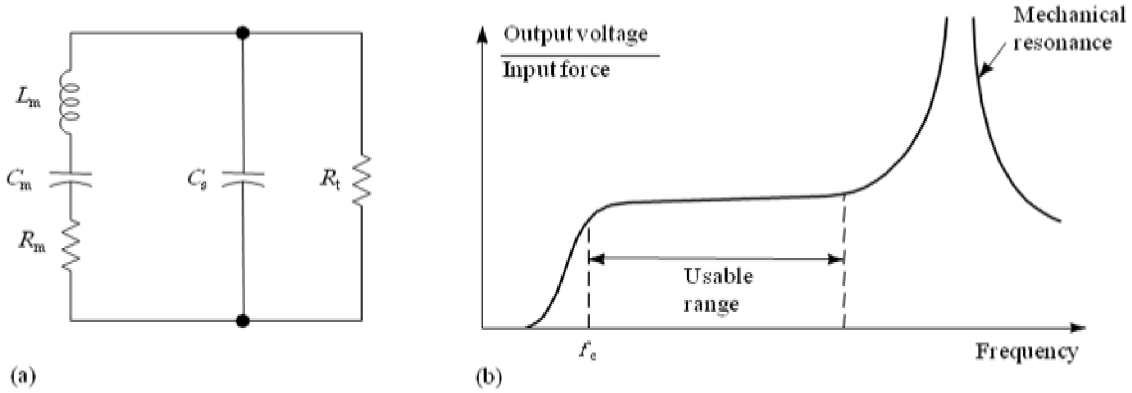


Figure 8.9 The high-frequency model (a) and frequency response (b) of a piezelectric sensor

The high frequency model and frequency response of a piezoelectric sensor is given in Figure 8.9. R_S is the sensor leakage resistance and C_S the capacitance. L_m, C_m and R_m represent the mechanical system. Mechanical resonance occurs at certain frequency that depends on the crystal material and geometry. The crystal can be used as a displacement sensor from the cut-off frequency f_s up to the onset of the resonance.

At the resonance frequency, the crystal oscillates mechanically as excited electrically and oscillates electrically as excited mechanically. The crystal is used in ultrasonic wave generation and detection. Also, due to the sharp resonance characteristics, the crystal becomes a part of oscillators.

STRAIN GAGES (GAUGES)

Mechanical Principles

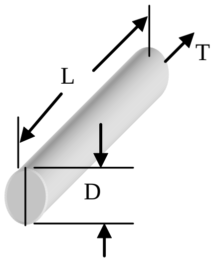


Figure 8.10 A metal bar

Tension and compression

A bar of metal as shown in Figure 8.10 is subjected to a force (T) that will elongate its dimension along the long axis that is called the *axial* direction. This force is called the *tension*. If the force acts in opposite direction and shortens the length, this called the *compression*.

Stress

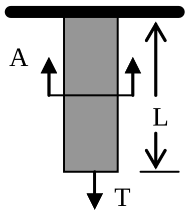


Figure 8.11 Bar with tension

Stress is defined as the force per unit area. Hence, the tension *T* produces an axial stress as illustrated in Figure 8.11,

$\sigma_a = T/A$ (N/m^2)

where A is the cross-sectional area. Dimension of stress is the same as that of the pressure.

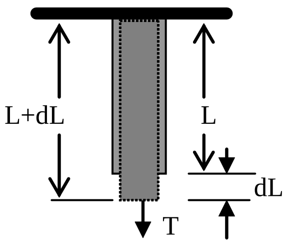


Figure 8.12 The strain

Strain

The stress generates changes in the dimensions of the bar as shown in Figure 8.12. The fractional change in length is defined as the **strain**. The change in the direction of the force is called the axial strain

$\varepsilon_a = dL/L$ (μm/m)

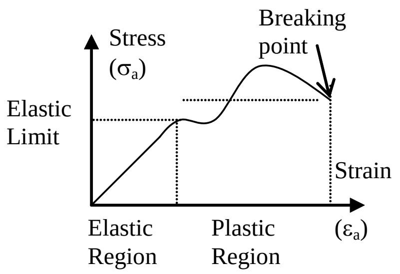
Dimension of strain is unity, i.e. strain is dimensionless.

Hooke's law

Stress is linearly related to strain for elastic materials. The Hooke's law mathematically expresses this relationship,

$\varepsilon_a = \sigma_a / E_y = (T/A)/E_y$

where E_y is called the **modulus of elasticity**, also called the **Young's modulus**. The relationship between the axial stress σ_a and axial strain ε_a is displayed in Figure 8.4. It has two distinct regions as the elastic (linear) and plastic (deformation). In the elastic range, the change is reversible, while in the plastic range the change is irreversible. Table in Figure 8.13 indicates elastic properties of some materials commonly used in engineering applications. The slope of the characteristic (ratio of change in stress to strain) is the Young's modulus and it is fairly constant if the stress remains below the



Material	E_y, N/m^2	Elastic limit σ_a N/m^2	Breaking strength σ_u N/m^2
Aluminum	$7x10^{10}$	$2.0x10^8$	$2.2x10^8$
Brass	$9x10^{10}$	$3.9x10^8$	$4.7x10^8$
Glass	$5x10^{10}$	$8x10^8$	$10x10^8$
Iron	$18x10^{10}$	$1.5x10^8$	$3.0x10^8$
Phosphor bronze	$10x10^{10}$	$4.2x10^8$	$5.6x10^8$
Steel	$20x10^{10}$	$9.0x10^8$	$11.0x10^8$

Elastic properties of some materials

Figure 8.13 The stress-strain relationship and elastic properties of some materials

elastic limit. The axial strain is in between 10^{-6} and 10^{-3} in most engineering applications. The strain is expressed in terms of micro-strain (μstrain) and

1 μstrain = 1 μm/m = 10^{-6}

Transverse strain

The tension that produces a strain in the axial direction causes another strain along the transverse axis (perpendicular to the axial axis) as

$\varepsilon_t = dD/D$

This is related to the axial strain through a coefficient known as the Poisson's ratio as

$dD/D = - \nu dL/L$

The negative sign indicates that the action is in reverse direction, that is, as the length increases, the diameter decreases and vice versa. For most metals ν is around 0.3 in the elastic region and 0.5 in the plastic region.

Electrical Resistance of the Strain Gage Wire

The resistance of the bar shown in Figure 8.10 is defined by

$R=\rho L/A$

Here, all three defining parameters, the resistivity ρ, the length L and the cross-sectional area A can change under the stress. Therefore, the change in the resistance can be obtained using the partial differential equation as follows:

$$dR = \frac{\partial R}{\partial \rho} d\rho + \frac{\partial R}{\partial L} dL + \frac{\partial R}{\partial A} dA$$

It yields;

$$dR = \frac{L}{A}d\rho + \frac{\rho}{A}dL - \frac{\rho L}{A^2}dA$$

and dividing both sides by R:

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dL}{L} - \frac{dA}{A}$$

With

$A = \pi r^2 = (\pi/4)D^2$

$dA/A = 2\ dD/D$

and

$dD/D = -\ \nu dL/L$

The relative change in resistance becomes

$$\frac{dR}{R} = \frac{d\rho}{\rho} + \frac{dL}{L}(1+2\nu)$$

The first term $d\rho/\rho$ is called "the piezoresistive effect" and the second term $\frac{dL}{L}(1+2\nu)$ is called "the dimensional effect". The ratio of the relative change in resistance to relative change in the length (axial strain) is called the gage factor K,

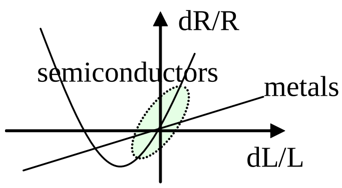


Figure 8.14 The gage factor

$K = (dR/R)/(dL/L) = (dR/R)/\varepsilon_a$

For wire type strain gages the second effect will be dominant yielding $K \approx 2$ and for heavily doped semiconductor type gages the second effect is dominant yielding K that ranges between 50 and 200. The variation of the relative change in resistance with the axial strain is shown in Figure 8.14. The metal gages have low gage factors, but linear characteristics. The semiconductor gages have parabolic characteristics that can be approximated to linear in a narrow range around the origin. The differential change dR can be replaced by the incremental change ΔR in this linear region. Then, the relative change in resistance

$\Delta R/R = K\varepsilon_a$

and it can be calculated easily if the gage factor K and strain ε_a are given.

Examples

Example 8.1

A phosphor-bronze wire, 1.0 mm^2 in cross-section area, is subjected to a tensile force of 10 N. Using the data in the table given previously;

- How much is the axial stress?

Axial stress $\sigma_a = T/A = 10N/10^{-6}m^2 = 10x10^6\ N/m^2$

- What is its elongation if the wire is 10 m long?

Axial strain $\varepsilon_a = \Delta l/l = \sigma_a/E_y = (10x10^6)/(10x10^{10}) = 10^{-4} = 100\ \mu$strain; $\Delta l = lx\varepsilon_a = 10x10^{-4}$ m = 1.0 mm (change in length is four order of magnitude smaller than the original one and most mechanical displacement measuring devices can't measure this)

- How much force is required to break the wire?

The breaking stress $=5.6x10^8\ N/m^2 = T_B/A$; $T_B = 5.6x10^8x10^{-6} = 560$ N

- How much is the change in resistance and value of the resistance under stress if K = 2 and untrained resistance of the wire is 100 Ω? Ans. $\Delta R/R = K\varepsilon_a = 2x10^{-4}$ yielding $\Delta R = 0.02\ \Omega$ and $R_{stress} = 100.02\ \Omega$ (most ohmmeters do not have this precision!)

Example 8.2

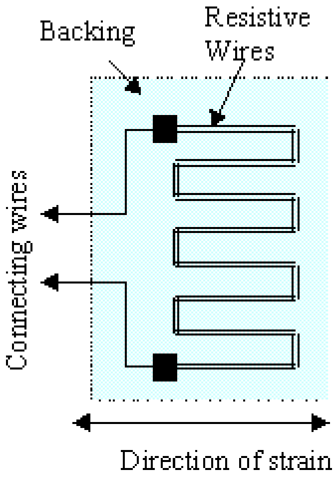
A strain gage has a gage factor 2 and exposed to an axial strain of 300 μm/m. The unstrained resistance is 350 Ω. Find the percentage and absolute changes in the resistance.

$\varepsilon_a = 300\ \mu m/m = 0.3x10^{-3}$; $\Delta R/R = K\varepsilon_a = 0.6x10^{-3}$ yielding %age change = 0.06% and $\Delta R = 350x0.6x10^{-3}$ = 0.21 Ω.

Example 8.3

A strain gage has an unstrained resistance of 1000 Ω and gage factor of 80. The change in the resistance is 1Ω when it is exposed to a strain. Find the percentage change in the resistance, the percentage change in the length and the external strain (μm/m)

$\Delta R/R$ (%) = 0.1 %; $\Delta L/L$ (%) = [$\Delta R/R$ (%)]/K = $1.25x10^{-3}$%, and ε_a = [$\Delta L/L$ (%)]/100 = $1.25x10^{-5}$ = 12.5 μm/m



Bonded and Unbonded Strain-Gages

The Bonded gage

A strain gage consists of a small diameter wire (an etched metal foil in reality), which is attached to the backing material (usually a plastic) as illustrated in Figure 8.15. The wire is looped back and forth several times to produce and effectively longer wire. The combination (elastic conductor of the strain gage and the backing) is bound to the specimen with insulating cement under no-load conditions as shown in Figure 8.16. A load is applied, which produces a deformation in both the specimen and the resistance element. This deformation is indicated through measurement of the change in resistance of the element and calculation procedures that will be described later.

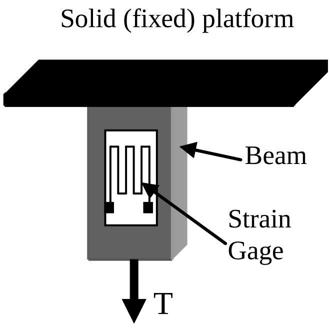


Figure 8.16 Fixing the gage

The bonded type strain gages come in different shapes and combinations to detect the strain in various applications. Gage factor is typically around 2.0. The electrical resistance of the unstrained gage is typically 120 Ω or 350 Ω. 600 Ω and 700 Ω gages are also available.

The Unbonded gage

Unbonded strain gages are formed of pre-strained resistive wires fixed between two poles as shown in Figure 8.17. The change in position of one of the poles increases and decreases the strain that is indicated through the measurement of the resistance as in the case of the bonded type.

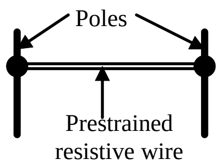


Figure 8.17

Effect of Temperature and Strain in other Directions

The temperature affects all resistive elements as

$$R = R_0[1+\alpha(T-T_0)]$$

R_0 is the resistance at T_0 and α is the temperature coefficient. This is very much pronounced in case of semiconductor gages due to high temperature coefficient.

The strain gage has the highest sensitivity against the strain in certain direction. However, it also has sensitivity to strains from other directions. The strain gage manufacturers generally specify this. Eventually, the change in the resistance can be expressed as the sum of resistance changes imposed by the wanted strain (sw), unwanted strain (su) and temperature (T).

$\Delta R = \Delta R_{sw} + \Delta R_{su} + \Delta R_T$

The effect of unwanted strain and temperature must be eliminated before the resistance change is used to indicate the strain.

THE WHEATSTONE BRIDGE

Utilization

The conventional methods for measuring the resistance involves application of a fixed current and measuring the voltage developed, or application of a fixed voltage and measure the resultant current. The relative change in the resistance of the strain gage $\Delta R/R$ is so small that the variation in the measured voltage or current remains within the uncertainty range. Hence, conventional methods cannot be used directly. The Wheatstone bridge shown in Figure 8.18 is a technique commonly used to measure changes in resistances accurately.

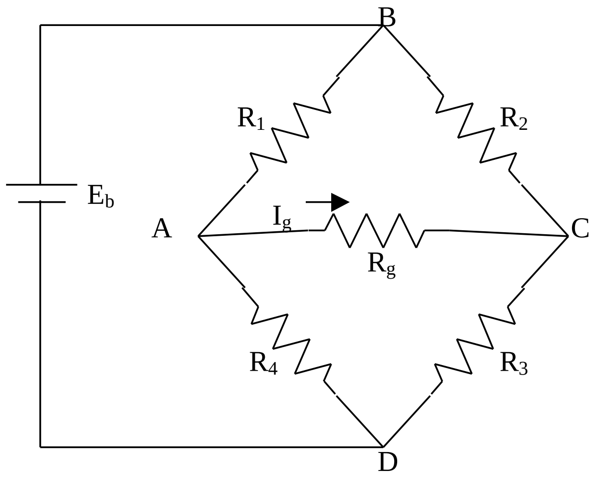
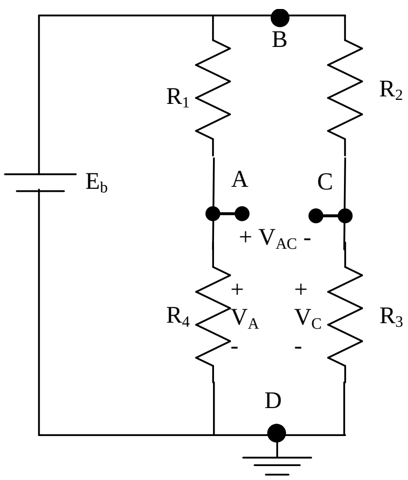


Figure 8.18 The Wheatstone bridge

Circuit Configuration

The bridge has a voltage source E_b and four arms with resistances as shown in Figure 8.18. The voltage source is connected between B and D to supply the bridge. The output is taken between A and C. The output may drive a moving coil meter or applied to a voltmeter.



The output voltage $E_0 = V_{AC}$. The circuit can be redrawn as shown in Figure 8.19 assuming the open circuit case at the moment ($R_g \rightarrow \infty$). Voltage across R_4 is

$V_A = E_b x R_4/(R_1+R_4)$

similarly

$V_C = E_b x R_3/(R_2+R_3)$

yielding

$E_0 = V_{AC} = V_A - V_C = E_b\left(\frac{R_4}{R_1+R_4} - \frac{R_3}{R_2+R_3}\right) = E_b\frac{R_2R_4 - R_1R_3}{(R_1+R_4)(R_2+R_3)}$

Null-mode of Operation

At balance

$R_2R_4 = R_1R_3$ or $R_1/R_4 = R_2/R_3$

and the output voltage is zero. This condition can be used to determine the exact value of an unknown resistor. It is placed into one the arms and others are adjusted until a zero volt is obtained at the output. This is called "the null mode of operation" as illustrated in Figure 8.20.

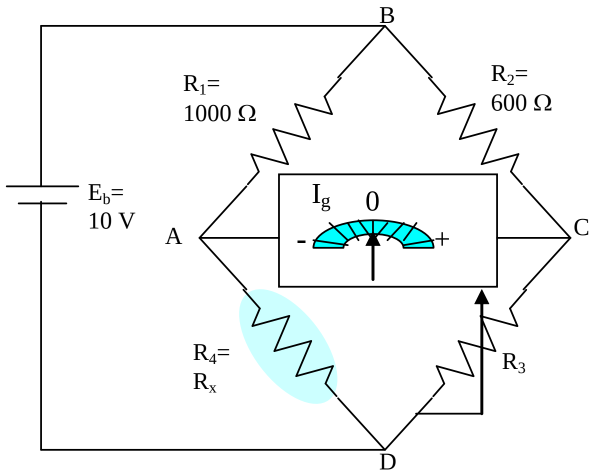


Figure 8.20 Circuit for null-mode

Example 8.4

Assume that the bridge shown in Figure 8.20 is used to determine the resistance of an unknown resistance R_x. The variable resistance is the resistance box that allows selection of several resistors in series to obtain the total resistance and it is set until null position in the meter observed. Calculate the unknown resistance if the variable resistance setting indicates 625.4Ω.

According to formula stated above, the bridge will be balanced if $R_1/R_4 = R_2/R_3$. Hence, $R_4 = R_x = R_1/(R_2/R_3)$ = 1000x625.4/600 = 1042.3 Ω.

Deflection-mode of Operation

All resistors can very around their nominal values as $R_1 + \Delta R_1$, $R_2 + \Delta R_2$, $R_3 + \Delta R_3$ and $R_4 + \Delta R_4$. Sensitivity of the output voltage to either one of the resistances can be found using the sensitivity analysis as follows:

$$S_{R_1} = \frac{\partial E_0}{\partial R_1} = E_b\frac{-R_3(R_1+R_4)(R_2+R_3)-(R_2+R_3)((R_2R_4+R_1R_3)}{(R_1+R_4)^2(R_2+R_3)^2} = -E_b\frac{R_4}{(R_1+R_4)^2}$$

Similarly,

$$S_{R_2} = \frac{\partial E_0}{\partial R_2} = E_b\frac{R_3}{(R_2+R_3)^2},\quad S_{R_3} = \frac{\partial E_0}{\partial R_3} = -E_b\frac{R_2}{(R_2+R_3)^2},\quad S_{R_4} = \frac{\partial E_0}{\partial R_4} = E_b\frac{R_1}{(R_1+R_4)^2}$$

If more than one element changes together, the output can be computed through superposition. The sensitivity is not constant indicating that the output-input relationship is not linear. It can be approximated by a linear characteristic only in a narrow range around the balance condition. Hence, the sensitivity analysis assumes small disturbances around the nominal value and it may yield large errors if the disturbance is large enough and becomes comparable to the nominal value.

Its Thevenin equivalent circuit shown in Figure 8.21 can replace the bridge. The equivalent Thevenin voltage is

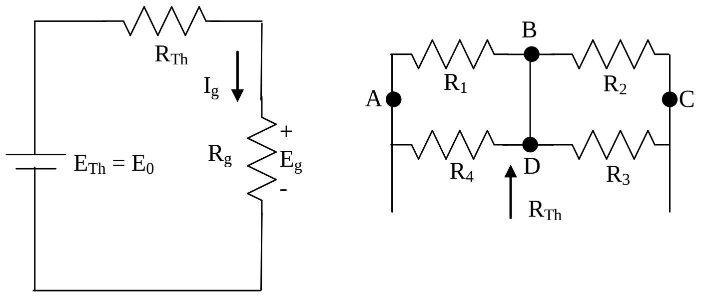


Figure 8.21 The equivalent circuit of the bridge

$E_{Th} = E_0 = V_{AC}$ (open circuit)

The equivalent resistance; $R_{Th} = R_1//R_4 + R_2//R_3$

The current through R_g; $I_g = E_0/(R_{Th} + R_g)$

and voltage across R_g; $E_g = E_0R_g/(R_{Th} + R_g)$

In case of open-circuit ($R_g \to \infty$) $E_g = E_0$. This output voltage causes **deflection** of the needle in a moving coil meter when applied.

Initially $R_1R_3 = R_2R_4$ and the bridge at balance yielding $E_g = 0$ and $I_g = 0$. At a slight unbalance $R_l \to R_l + \Delta R_l$ where $/\Delta R_l/ << R_l$, the resistance is slightly changed while E_g is drastically changed (from 0 to some finite value).

Example 8.5

Given E_b = 10 V, R_g = 50 Ω, the bridge is initially balanced with $R_1 = R_2 = R_3 = R_4 = R = 1000\Omega$. The bridge is unbalanced by $\Delta R_1 = 1\Omega$, $\Delta R_2 = -1\Omega$, $\Delta R_3 = 2\Omega$, $\Delta R_4 = -1\Omega$. Find the current I_g through exact and approximate methods and determine the percentage error in the current when the approximate method is used. Assume that the measurement is ideal and no measurement error is made.

<u>Exact solution</u>

$$E_0 = E_b \frac{R_2R_4 - R_1R_3}{(R_1 + R_4)(R_2 + R_3)} = E_b \frac{(R + \Delta R_2)(R + \Delta R_4) - (R + \Delta R_1)(R + \Delta R_3)}{(R + \Delta R_1 + R + \Delta R_4)(R + \Delta R_2 + R + \Delta R_3)}$$ that gives

$$E_0 = 10\frac{999x999 - 1001x1002}{(1001+999)(999+1002)} = -12.49625\, mV$$

R_{Th} = $R_1//R_4$ + $R_2//R_3$ = $\dfrac{(R+\Delta R_1)(R+\Delta R_4)}{(2R+\Delta R_1+\Delta R_4)} + \dfrac{(R+\Delta R_2)(R+\Delta R_3)}{(2R+\Delta R_2+\Delta R_3)}$ =1001x999/2000 + 999x1002/2001 = 499.9995 + 500.2489 = 1000.2484 Ω

$I_g = E_0/(R_{Th} + R_g)$ = -12.49625/1050.2484 = -11.8988 μA

Approximate solution-1

$$E_0 = E_b\frac{(R+\Delta R_2)(R+\Delta R_4) - (R+\Delta R_1)(R+\Delta R_3)}{(R+\Delta R_1+R+\Delta R_4)(R+\Delta R_2+R+\Delta R_3)} \cong E_b\frac{\Delta R_4+\Delta R_2-\Delta R_1-\Delta R_3}{4R} = -12.5\ mV$$

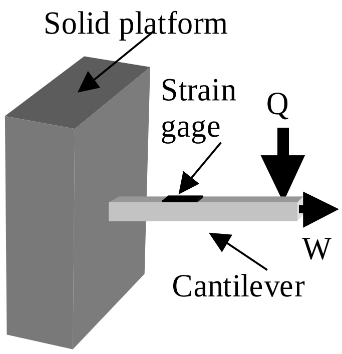
This is obtained after ignoring the cross terms in ΔR's in the numerator and also ignoring ΔR's in the denominator as compared to $2R$.

$R_{Th} \cong R/2 + R/2$ = 1000Ω yielding I_g = -12.5/1050 = -11.9048 μA

%age error in I_g = 100x(-11.9048 + 11.8988)/(-11.8988) = 0.05398%

Approximate solution-2

Using the sensitivity analysis, $\Delta E_0 = E_0 = (\partial E_0/\partial R_1)\Delta R_1 + (\partial E_0/\partial R_2)\Delta R_2 + (\partial E_0/\partial R_3)\Delta R_3 + (\partial E_0/\partial R_4)\Delta R_4$ = (E_b/4000)[-1-1-2-1] = -12.5 mV; R_{Th} = 1000 Ω as above yielding the same result as the approximate solution-1.



BRIDGE CONFIGURATIONS FOR STRAIN GAGE MEASUREMENTS

Bridge with a Single Active Element (Quarter Bridge)

Physical Connection

The strain gage is exposed to the force that causes the stress in variety of ways. The cantilever type shown in Figure 8.22 is one the most famous. The lever is fixed to a solid platform and a force *Q* is applied to its free end. This force causes tension in the gage when applied in the direction shown and causes an increase in its resistance (ΔR positive). As the force is applied in the opposite direction, it produces compression in the gage that produces a decrease in its resistance (ΔR negative).

Configuring the Bridge

The strain gage is placed into one of the bridge arms and other three arms are completed with fixed resistors as shown in Figure 8.23. R_4 is taken as the strain gage. R_3 is made variable to balance (null) the bridge when there is no force applied (silent condition). This is needed since the resistors used have tolerances and exact matching is very difficult.

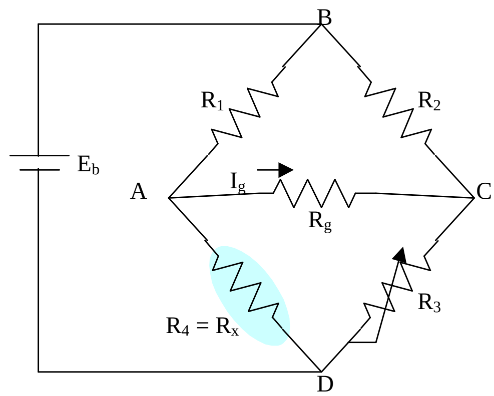


Figure 8.23 A quarter bridge

Analysis of the Circuit

Let $R_1 = R_2 = R_3 = R$ and $R_4 = R_x = R + \Delta R = R(1 + \Delta R/R)$,

and let $x = \Delta R/R$. The open circuit voltage $E_0 = 0$ at balance ($\Delta R = 0$). At slight unbalance ($\Delta R \neq 0$)

$$E_0 = E_b \frac{R_2 R_4 - R_1 R_3}{(R_1 + R_4)(R_2 + R_3)} = E_b \frac{R(R + \Delta R) - R^2}{(R + R + \Delta R)(R + R)} = E_b \frac{\Delta R}{2(2R + \Delta R)}$$

After replacing $x = \Delta R/R$,

$$E_0 = E_b \frac{x}{2(2 + x)} = E_b \frac{x}{4(1 + \frac{x}{2})}$$

The denominator can be expended using Taylor series as

$$(1 + \frac{x}{2})^{-1} = 1 - \frac{x}{2} + \frac{x^2}{4} - \dots$$

Then

$$E_0 = \frac{E_b}{4}(x - \frac{x^2}{2} + \frac{x^3}{4} - ...)$$

Since $x<<1$, higher order terms can be neglected yielding

$$E_0 \cong \frac{E_b}{4}x = \frac{E_b}{4}\frac{\Delta R}{R}$$

Sensitivity analysis can also be used. $S_{R_4} = \frac{\partial E_0}{\partial R_4} = E_b \frac{R_1}{(R_1 + R_4)^2}$ was given previously. Hence,

$$E_0 = \Delta R_4 S_{R_4} = E_b \frac{R}{(R+R)^2}\Delta R = \frac{E_b}{4}\frac{\Delta R}{R}$$

Effect of Temperature and Tensile Strain

The change in the resistance can be expressed as the sum of resistance changes imposed by the wanted strain (sw), unwanted strain (su) and temperature (*T*) as stated before. *Q* is the wanted strain and *W* (tensile) is the unwanted strain in this case. Hence, $\Delta R = \Delta R_Q + \Delta R_W + \Delta R_T$ as already stated. The effect of unwanted strain and temperature must be eliminated. The circuit as it is provides no compensation. Using a second strain gage of the same type for R_1 can compensate effect of temperature. This second gage can be placed at a silent location within the sensor housing, hence kept at the same temperature as the first one. As a result, both R_1 and R_4 have the same amount of changes due to temperature that cancel each other in the equation yielding perfect temperature compensation.

Example 8.6

E_b = 4V, R_g = 50 Ω , $R_1 = R_4$ = 120 Ω, $R_2 = R_3$ = 100 Ω at balance (no load). R_4 is used as the strain gage with gage factor K = 2 . Find the galvanometer current I_g for ε_a = 400 μm/m.

Solution: $\Delta R/R = \varepsilon K$ = 2x4x10^{-4} = 8x10^{-4} ; $\Delta R = R\varepsilon K$ = 120x8x10^{-4} = 0.096 Ω.

<u>Exact calculation</u>:

R_{Th} = (120x120.096)/240.096 + 50 = 110.02399 Ω

$$E_0 = 4\{\frac{120.096x100 - 120x100}{(120.096+120)(100+100)} = 0.7996mV\ ;\ I_g = \frac{0.7996x10^{-3}}{110.02399+50} = 4.9967\mu A$$

<u>Approximate calculation</u>:

R_{Th} = (120x120)/240+ 50 = 110 Ω $E_0 = \frac{E_b}{4}\frac{\Delta R}{R} = 8x10^{-4}V = 0.8mV$; $I_g = \frac{0.8x10^{-3}}{110+50} = 5\mu A$,

The percentage error in $I_g = \frac{5-4.9967}{4.9967}x100 < 0.065\%$

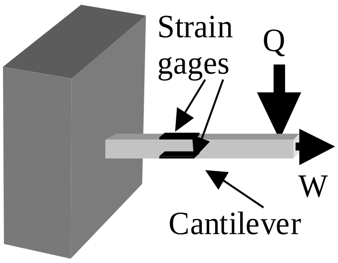


Figure 8.24

Bridge with Two Active Elements (Half Bridge)

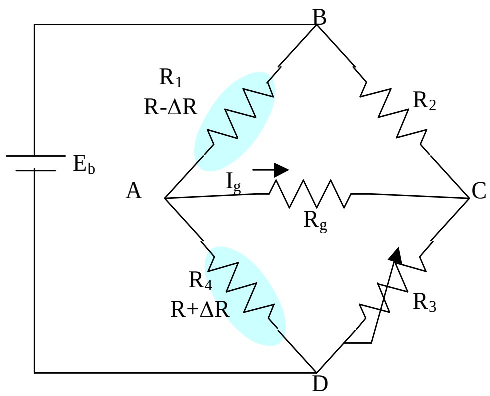
Physical Connection

Two strain gages are fixed to opposite surfaces of the cantilever as shown in Figure 8.24. The force, when applied in the direction shown, causes tension in the gage on the top surface ($R + \Delta R_Q$) and compression on the gage at the bottom surface ($R - \Delta R_Q$). The tensile force W causes ($R + \Delta R_W$) on both gages. The temperature also produces ($R + \Delta R_T$) on both gages.

Configuring the Bridge

The strain gages are placed into two neighboring arms of one branch of the bridge as shown in Figure 8.25. The other branch is compensated by two equal-value fixed resistors. In the Figure R_1 and R_4 are taken as the strain gages. R_3 is made variable to balance (null) the bridge when there is no force applied (silent condition).

Analysis of the Circuit



Let $R_2 = R_3 = R$; $R_1 = R - \Delta R$; $R_4 = R + \Delta R$, the open circuit voltage $E_0 = 0$ at balance ($\Delta R = 0$). At slight unbalance ($\Delta R \neq 0$)

$$E_0 = E_b \frac{R_2 R_4 - R_1 R_3}{(R_1 + R_4)(R_2 + R_3)}$$

$$= E_b \frac{R(R+\Delta R) - R(R-\Delta R)}{(R-\Delta R+R+\Delta R)(R+R)}$$

$$= E_b \frac{2\Delta R}{4R} = \frac{E_b}{2}\frac{\Delta R}{R}$$

The expression yields exact result without any approximation. The output voltage is doubled compared to the case of single element.

$$R_{Th} = \frac{(R-\Delta R)(R+\Delta R)}{2R} + \frac{R}{2} = R(1 - \frac{\Delta R^2}{2R^2})$$

with $\Delta R << R$, $R_{Th} \cong R$. Hence, the error in accepting the approximate solution (only for R_{Th}, since E_0 is exact) is negligible. Effects of wanted and unwanted strains and temperature are illustrated in Figure 8.26 for the measuring gages of Figure 8.25. Temperature and unwanted tensile strain will have no effect on the output voltage since they are completely compensated as follows:

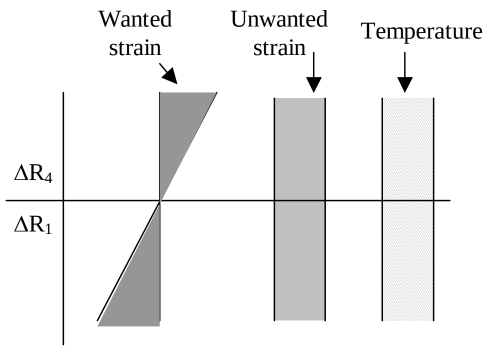


Figure 8.26 Effects of wanted and unwanted strains and temperature on measuring gages

$E_0 = \frac{E_b}{4R}(\Delta R_4 - \Delta R_3)$ (from the sensitivity analysis)

$\Delta R_4 = \Delta R_Q + \Delta R_W + \Delta R_T$, and $\Delta R_3 = -\Delta R_Q + \Delta R_W + \Delta R_T$ yielding the bridge equation for the half-bridge configuration as

$$E_0 = \frac{E_b}{4R}(\Delta R_4 - \Delta R_3) = \frac{E_b}{2}\frac{\Delta R}{R}$$

Bridge with Four Active Elements (Full Bridge)

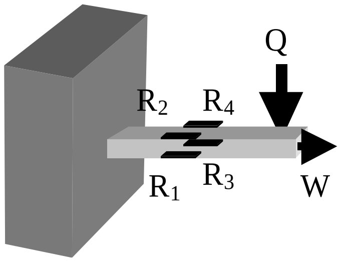
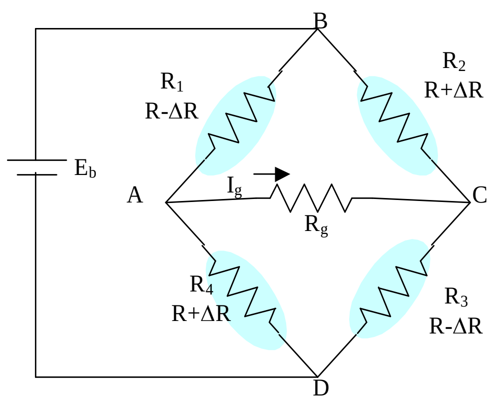


Figure 8.27

All four arms of the bridge are made up of strain gages that are affected by the external strain. Two gages are fixed on either of the opposite surfaces of the cantilever as shown in Figure 8.27. The force, when applied in the direction shown, causes tension on gages at the top surface ($R + \Delta R_Q$) and compression on gages at the bottom surface ($R - \Delta R_Q$). The tensile force W causes ($R + \Delta R_W$) on all gages. The temperature also produces ($R + \Delta R_T$) on all gages.



The strain gages that are working together are placed into opposite (non-neighboring) arms of the bridge as shown in Figure 8.28. The strain gage resistors are manufactured for a perfect match to have the open circuit voltage $E_0 = 0$ at balance ($\Delta R = 0$). At slight unbalance ($\Delta R \neq 0$) with $R_1 = R_3 = R - \Delta R$; $R_2 = R_4 = R + \Delta R$.

$$E_0 = E_b \frac{R_2 R_4 - R_1 R_3}{(R_1 + R_4)(R_2 + R_3)}$$

$$= E_b \frac{(R+\Delta R)(R+\Delta R)-(R-\Delta R)(R-\Delta R)}{(R-\Delta R+R+\Delta R)(R+\Delta R+R-\Delta R)} = E_b \frac{\Delta R}{R}$$

The expression also yields exact result without any approximation. The output voltage is quadrupled compared to the case of single element.

$$R_{Th} = 2\frac{(R-\Delta R)(R+\Delta R)}{2R} = R(1-\frac{\Delta R^2}{R^2})$$

with $\Delta R << R$, $R_{Th} \cong R$. Hence, the error in accepting the approximate solution (only for R_{Th}, since E_0 is exact) is negligible. Temperature and unwanted tensile strain will have no effect on the output voltage since they are completely compensated as in the case of the half-bridge.

Generalized Instrumentation System

A cantilevered beam and the Wheatstone bridge can be used to determine the strain and/or the bending force as illustrated in Figure 8.29. The cantilever converts the bending force into a bending stress and a bending strain provided that the metal stays within its elastic limits. The strain gages

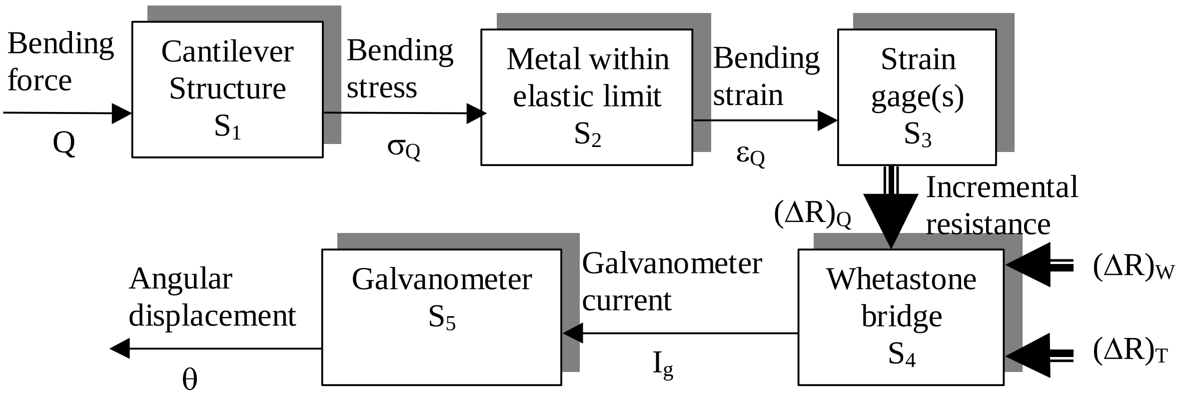


Figure 8.29 A functional block diagram to Illustrate measurement of strain using a cantilevered beam and Wheatstone bridge

that are placed over the beam encounter incremental changes in their resistance. The Wheatstone bridge provides the environment for determining small changes in resistances and generates an output voltage that can be displayed using a galvanometer. At the end, the angular produces a displacement of its pointer that is proportional to the input force.

NOVEL PRESSURE SENSORS

Quantum Tunneling Composites

First produced in 1996, the Quantum Tunneling Composite (QTC) is a composite material made from micron-sized particles conductive filler particles combined with a non-conducting elastomeric binder, typically silicone rubber. The unique method of combining these raw materials results in a composite which exhibits significantly different electrical properties when compared with any other electrically conductive material. Hence it is a flexible polymer that exhibits extraordinary electrical properties as illustrated in Figure 8.30. QTC usually comes in the form of pills or sheet. QTC pills are just tiny little

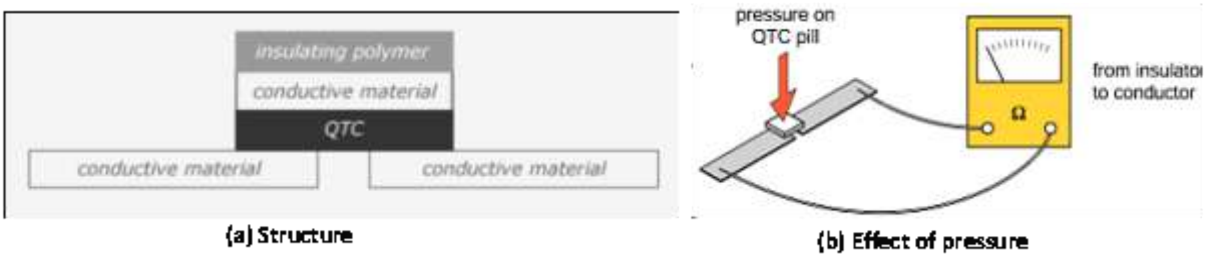


Figure 8.30 Structure and effect of pressure for QTC

pieces of the material. The sheets are composed of one layer of QTC, one layer of a conductive material, and a third layer of a plastic insulator. While QTC sheets switch quickly between high and low resistances, QTC pills are pressure sensitive variable resistors.

QTC is used as a pressure sensor; in its normal state it is a perfect insulator, but when compressed it becomes a more or less perfect conductor and able to pass very high currents. It utilizes quantum tunneling: without pressure, the conductive elements are too far apart to conduct electricity; when pressure is applied, they move closer and electrons can tunnel through the insulator. The effect is far more pronounced than would be expected from classical (non-quantum) effects alone, as classical electrical resistance is linear (proportional to distance), while quantum

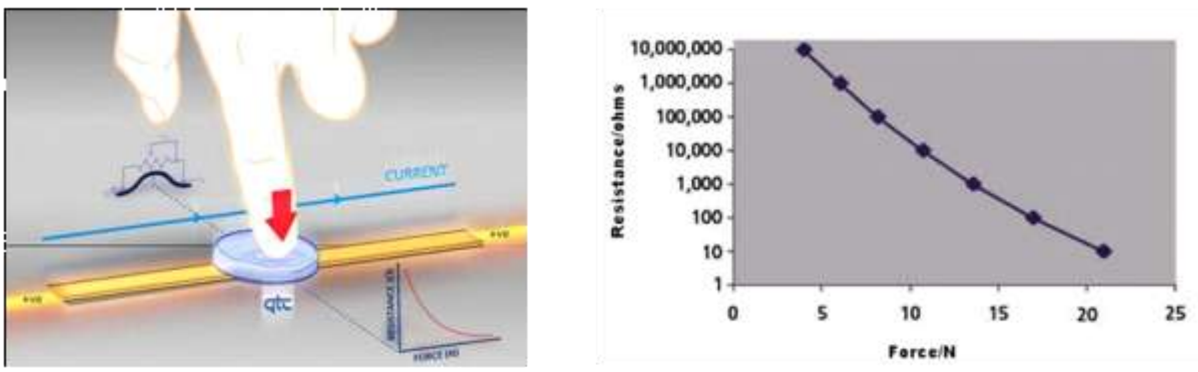


Figure 8.31 Effect of pressure on a QTC pill

tunneling is exponential with decreasing distance, allowing the resistance to change by a factor of up to 10^{12} between pressured and unpressured states as shown in Figure 8.31.

Applications

QTC has the unique ability to smoothly change from an electrical insulator to a metal-like conductor when placed under pressure. While in an unstressed state the QTC material is a near-perfect insulator; with any form of deformation the material starts to conduct and with sufficient pressure metallic conductivity levels can be achieved. This property can be utilized to convert pressure or force into an electrical signal as illustrated in Figure 8.32.

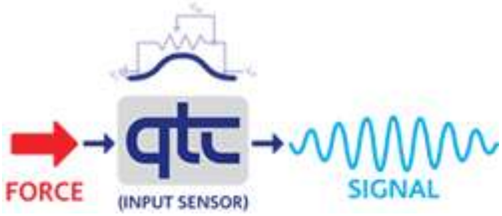


Figure 8.32 QTC as a force sensor

QTC can be tailored to suit different force, pressure or touch sensing applications – from sensing feather-light or finger operation to heavy pressure applications. Figure 8.33 shows various application examples of sensing capabilities of QTC material.

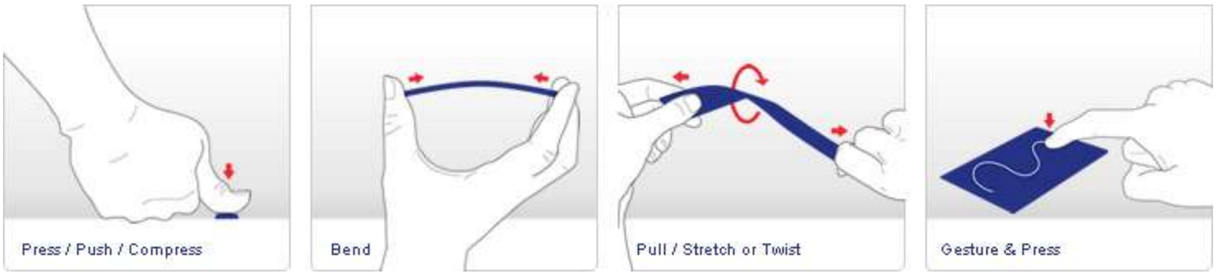


Figure 8.33 Examples of potential sensing capabilities of QTC material

QTC has been implemented within clothing to make “smart”, touchable membrane control panels to control electronic devices within the clothing, e.g. mp3 players or mobile phones. This allows equipment to be operated without removing clothing layers or opening fastenings and makes standard equipment usable in extreme weather or environmental conditions such as Arctic/Antarctic exploration or spacesuits. However, eventually, due to the low cost of QTC, this technology will become available to the general user.

PROBLEMS ON MEASUREMENT OF MECHANICAL QUANTITIES

Review Questions

1. How a mechanical displacement can be sensed?
2. What are the resistive displacement sensors and how they are used?
3. What are the inductive displacement sensors?
4. What is the LVDT with advantages and limitations?
5. What are the principles of operation of capacitive sensors?
6. What are the applications and limitations of piezoelectric sensors?
7. What are the tension and compression?
8. What are the stress and strain?
9. What is the transverse strain?
10. How the strain and stress are related to each other?
11. Why the strain is used as an important mechanical measure?
12. What is the piezoresistance?
13. What is the gage factor?
14. What are the mechanical principles for strain gages?
15. What causes the changes in the electrical resistance of the strain gage wire?
16. How can strain gages can be used in practice?
17. What are bonded and unbonded strain-gages?
18. How do the temperature and strain in other directions affect the displacement measurements?
19. What is the wheatstone bridge?
20. Where and how a wheatstone bridge is used in practice?
21. How can you design a measurement circuit that uses a wheatstone bridge?
22. How can you compare the null-mode and deflection-mode of operation of wheatstone bridges?
23. How can you connect strain gages mechanically to measure force?
24. How can you relate the output voltage to the strain in a quarter bridge configuration?
25. What are the limitations of the quarter bridge?
26. Why the half bridge is the mostly preferred configuration?
27. How can you eliminate the effects of temperature and unwanted strain in the measurements using wheatstone bridges?
28. What are the quantum tunneling composites and how they can be used in sensing the strain?
29. What are the advantages of quantum tunneling composites in touch screen displays?

Multiple-Choice Questions

1. Which one of the following transducers needs a phase sensitive demodulator?
    a. Thermistor;
    b. Strain gage;
    c. LVDT;
    d. Piezoelectric crystal.
2. The effect of temperature on a strain gage displacement transducer can be best compensated by using a:
    a. Thermistor;
    b. Second strain gage at a reference temperature;
    c. Thermocouple;
    d. Second strain gage at the same temperature as the measuring one.
3. In piezoelectric transducers which of the following is primarily related to the velocity?
    a. Current coming out of it;
    b. Voltage across it;
    c. Impedance of it;
    d. Charge on it.
4. A strain gage type displacement transducer has a gage factor of 40 and unstrained resistance of 120Ω.The Poisson coefficient is 0.4. How much of the gage factor is due to piezoresistive effect?
    a. 38;
    b. 39;
    c. 35;
    d. None. It is....
5. An external strain causes 6Ω change in the resistance in the above question. The percentage change in dimension is:
    a. 1;
    b. 0.5;
    c. 0.125;
    d. 2.
6. The LVDT requires
    a. Wheatstone bridge
    b. DC power supply
    c. Phase-sensitive demodulator
    d. A thermistor to compensate for temperature
    e. Balancing resistor

7. A strain gage with gage factor of 40 and unstrained resistance of 100 Ω is connected to an arm of a Wheatstone bridge. The change in the resistance for 0.1% dimensional change is:
    a. 2 Ω
    b. 6Ω
    c. 5Ω
    d. 4Ω
    e. None, it is ….
8. The bridge is supplied with 10 V DC. The output voltage across the measuring arms of the bridge is:
    a. 1 V
    b. 0.1 V
    c. 0.01 V
    d. 0.4 V
    e. None, it is …
9. The effect of temperature on a strain gage displacement transducer can be best compensated by using a:
    a. Thermistor;
    b. Second strain gage at the same temperature as the measuring one.
    c. Second strain gage at a reference temperature;
    d. Thermocouple;

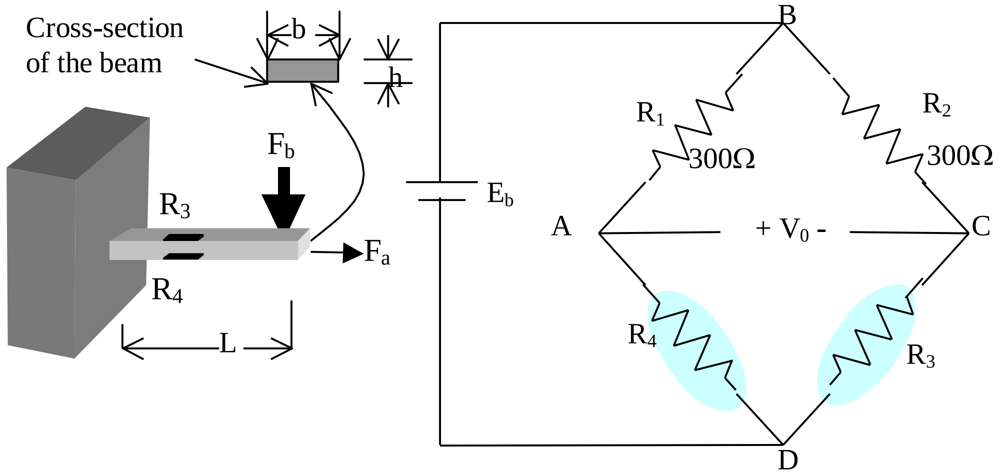
Questions with Solutions

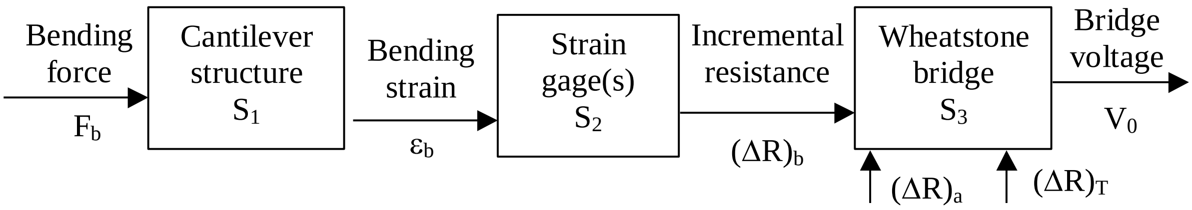
1. Two identical strain gages are placed on opposite surfaces of a cantilevered beam as shown in the figure. The unstrained resistance is 120 Ω; gage factor K = 1.8± 0.2%; E_b = 10.0 V± 0.5%. The bridge is compensated by two 300Ω resistors.
    a. **Write down** the mathematical relationship between the bending force F_b and strain ε_b given that the bending stress $\sigma_b = \frac{6L}{bh^2}F_b$. Also write down the mathematical expression for the sensitivity $S_1=\varepsilon_b/F_b$.

Ans. $\varepsilon_b = {}^{\sigma_b}/_{E_Y}$; given σ_b, $\varepsilon_b = \frac{6L}{bh^2 E_Y}F_b$ and $S_1 = \frac{6L}{bh^2 E_{Y\ b}}$

    b. **Write down** mathematical expressions for sensitivities S_2 and S_3.

Ans. $K = \frac{\Delta R/R}{\Delta l/l}$ yielding $\Delta R_b = RK\frac{\Delta l}{l} = \varepsilon_b RK$; hence $S_2 = RK$





It looks like a half bridge but sensors are placed differently. For this configuration: $V_0 = E_b\left(\frac{R+\Delta R}{R_1+R+\Delta R} - \frac{R-\Delta R}{R_1+R-\Delta R}\right) = E_b\left(\frac{2R_1\Delta R}{(R_1+R+\Delta R)(R_1+R-\Delta R)}\right) = E_b\left(\frac{2R_1\Delta R}{(R_1+R)^2-\Delta R^2}\right)$; hence, the sensitivity $S_3 = \frac{\partial V_0}{\partial \Delta R} = E_b\left(\frac{2R_1((R_1+R)^2-\Delta R^2)-4R_1\Delta R^2}{((R_1+R)^2-\Delta R^2)^2}\right)$, as S_3 is evaluated around $\Delta R = 0$, it's nominal value becomes: $S_3 = E_b\left(\frac{2R_1(R_1+R)^2}{(R_1+R)^4}\right) = E_b\left(\frac{2R_1}{(R_1+R)^2}\right)$

c. **Prove** that the given arrangement compensates the effects of temperature and axial forces (F_a).

d. **Calculate** ε_b , and V_0 and their **uncertainties** given the following data:

F_b = 15 N $\pm$ 1% ; Young's modulus of elasticity = $E_Y = 20x10^{10}\ Nm^{-2}$

L =100 mm; h = 4 mm $\pm$ 1%; b = 20 mm.

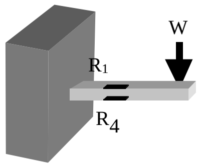
Ans. $\varepsilon_b = \frac{6L}{bh^2E_Y}F_b = \frac{0.6x15}{0.004^2x0.02x20x10^{10}} = 140.63\mu m/m$, maximum error is $\pm$3%, the expected error =$\sqrt{1+4} = 2.24\%$

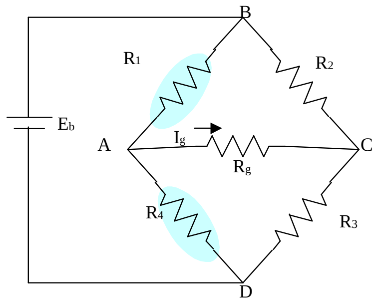
$V_0 = S_2 x S_3 x \varepsilon_b = RKE_b\left(\frac{2R_1}{(R_1+R)^2}\right)\varepsilon_b = 1.033mV$ with maximum uncertainty of $\pm$(0.2+0.5+3)% = $\pm$3.7%, expected uncertainty is =$\sqrt{0.04+0.25+5} = 2.3\%$

2. Assume that a Wheatstone bridge is used for measuring the strain using a galvanometer with coil resistance of 50 Ω and full-scale deflection current 50μA. The galvanometer is calibrated to read the strain directly. The bridge supply is 10 V, the unstrained resistance of strain gages is 110 Ω and compensating bridge resistors are also taken as 110 Ω. The gage factor of the strain gage is 2. Find the value of strain for full-scale deflection.

Ans. Assuming that we use a half bridge, $V_{th} = \frac{E_b}{2}\frac{\Delta R}{R} = 10\varepsilon_b$, R_{th} = 110 Ω, $I_g = \frac{V_{th}}{R_{th}+R_g} = \frac{\varepsilon_b}{16}$ yielding the full-scale $(\varepsilon_b)_{max}$ = 800 μm/m

3. In the bridge shown, two strain gages R_1 and R_4 are placed onto the opposite faces of a cantilever. The unstrained resistance is 120 Ω; gage factor F_G = 50; E_b = 9 V. The bridge is compensated by R_2=R_3= 360 Ω; R_g= 100 Ω;,. Calculate the value of the current through the galvanometer (I_g) for an applied strain ε=0.15x10^{-3}.





Ans. $\frac{\Delta R}{R} = F_G\frac{\Delta l}{l} = 7.5x10^{-3}$, V_T = 4.5x7.5x10^{-3} = 33.8 mV, R_T = 60 + 180 = 240 Ω; $I_g = \frac{V_T}{R_T+R_m} = 99.3\mu A$

General Questions

1. Table for Pr 8.1 shows elastic properties of some engineering materials. Assuming that a wire 10 m long and 1.0 mm2 in cross-section area is made of each material.
   a. Which element has the largest strain?
   b. Which element can carry the largest weight without breaking?
   c. Assume that the wire is made of aluminum and subjected to a tensile force of 10 N. How much is the axial stress? What is its elongation?

Table for Pr 8.1

Material	E_v, N/m^2	Elastic limit N/m^2	Breaking strength N/m^2
Aluminum	7x10^{10}	2.0x10^8	2.2x10^8
Brass	9x10^{10}	3.9x10^8	4.7x10^8
Glass	5x10^{10}	8x10^8	10x10^8
Iron	18x10^{10}	1.5x10^8	3.0x10^8
Phosphor bronze	10x10^{10}	4.2x10^8	5.6x10^8
Steel	20x10^{10}	9.0x10^8	11.0x10^8

Elastic properties of materials

d. Assume the gage factor is 2.0 and unstrained resistance is 100 Ω. How much is the resistance with the strain due to the tensile force of 10 N?

2. A stainless-steel beam shown in Figure for Pr 8.2 has 10.0 mm^2 in cross-section area. It is subjected to a tensile force of 100 N. Modulus of elasticity for stainless steel, $E_Y = 2.0 \times 10^{11}$ N/m^2.
   a. How much is the axial stress?
   b. What is its elongation if the beam has the unstrained length 10 cm?

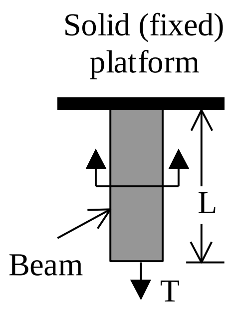


Figure for Pr 8.2

3. Assume that a Wheatstone bridge is used in quarter bridge configuration for measuring the strain. A galvanometer with coil resistance of 50 Ω and full-scale deflection current 50μA is calibrated to read the μstrain directly. The bridge supply is 5.0 V, the unstrained resistance of the strain gage is 120 Ω and compensating bridge resistors are also taken as 120 Ω. The gage factor of the strain gage is 2.0. Find the value of strain for full-scale deflection and error in accepting the approximate solution.

4. A strain gage has a gage factor 2.1 and unstrained resistance is 600 Ω at 25°C. The temperature coefficient is 2 μstrain/°C (i.e. $\Delta R_{Temp}/R = 4.2 \times 10^{-6}/°C$)
   a. Find the percentage and absolute changes in the resistance if ε_a = 500 μm/m at 25°C.
   b. Find the percentage change in resistance for ε_a = 0 μm/m at 75°C
   c. Find the percentage change in resistance for ε_a = 500 μm/m at 75°C

5. In a strain gage:
   a. Show the mathematical relationship between stress and strain.
   b. Find the percentage change in the length and in the resistance of the strain gage with gage factor = 25, unstrained resistance is 350 Ω and strain $\varepsilon = 400 \times 10^{-6}$ = 400 μstrain.

6. For the Wheatstone bridge shown in Figure for Pr 8.6;
   a. Determine the condition for $V_{AC} = 0$
   b. Assume that R_1 is an unknown resistance. R_4 = 210 Ω, R_2 = 125 Ω. R_3 is a variable resistor and when it set to 350 Ω, $V_{Ac} = 0$. Calculate the value of R_1.

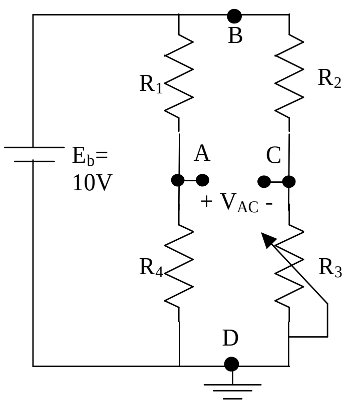


Figure for Pr 8.6

7. In the bridge shown in Figure for Pr 8.7, two strain gages R_1 and R_4 are placed onto the opposite faces of a cantilever.

a. Determine the condition for the current $I_g = 0$.

b. Derive the equation relating I_g to the strain.

8. The unstrained resistance of strain gages is 120 Ω; gage factor K = 50; E_b = 10.0 V. The bridge is compensated by R_2=R_3= 220 Ω; R_g= 80 Ω. Calculate the value of the current I_g for an applied strain ε=200 μstrain using both the approximate formula and exact formula. Calculate the error in accepting the approximate formula.

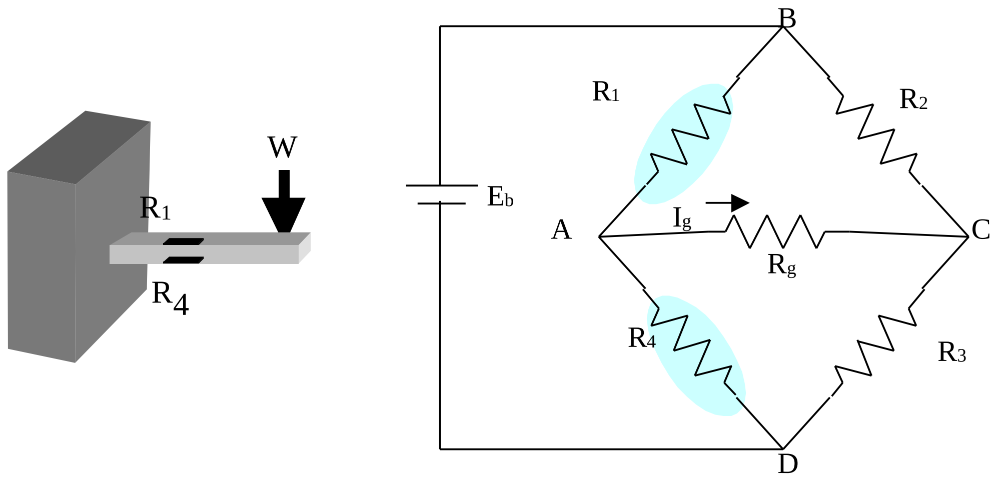


Figure for Pr 8.7 Two strain gages on a cantilever forming a half-bridge

9. In Figure for Pr 8.7, two strain gages R_1 and R_4 are placed onto the opposite faces of a cantilever. The unstrained resistance is 120 Ω; gage factor F_G = 50; E_b = 9.0 V. The bridge is compensated by R_2=R_3= 360 Ω; R_g= 100 Ω. Calculate the value of the current through the galvanometer (I_g) for an applied strain ε=0.15x10^{-3}.

10. Two identical strain gages are placed on opposite surfaces of a cantilevered beam as shown in Figure for Pr 8.10. The unstrained resistance is 120 Ω; gage factor K = 2.0± 0.2%; E_b = 9.0 V± 0.5%. The bridge is compensated by two 330Ω resistors.

    a. **Write down** the mathematical relationship between the bending force F_b and strain ε_b given that the bending stress $\sigma_b = \frac{6L}{bh^2} F_b$. Also write down the mathematical expression for the sensitivity S_1=ε_b /F_b.

    b. **Write down** mathematical expressions for sensitivities S_2 and S_3.

    c. **Prove** that the given arrangement compensates the effects of temperature and axial forces (F_a).

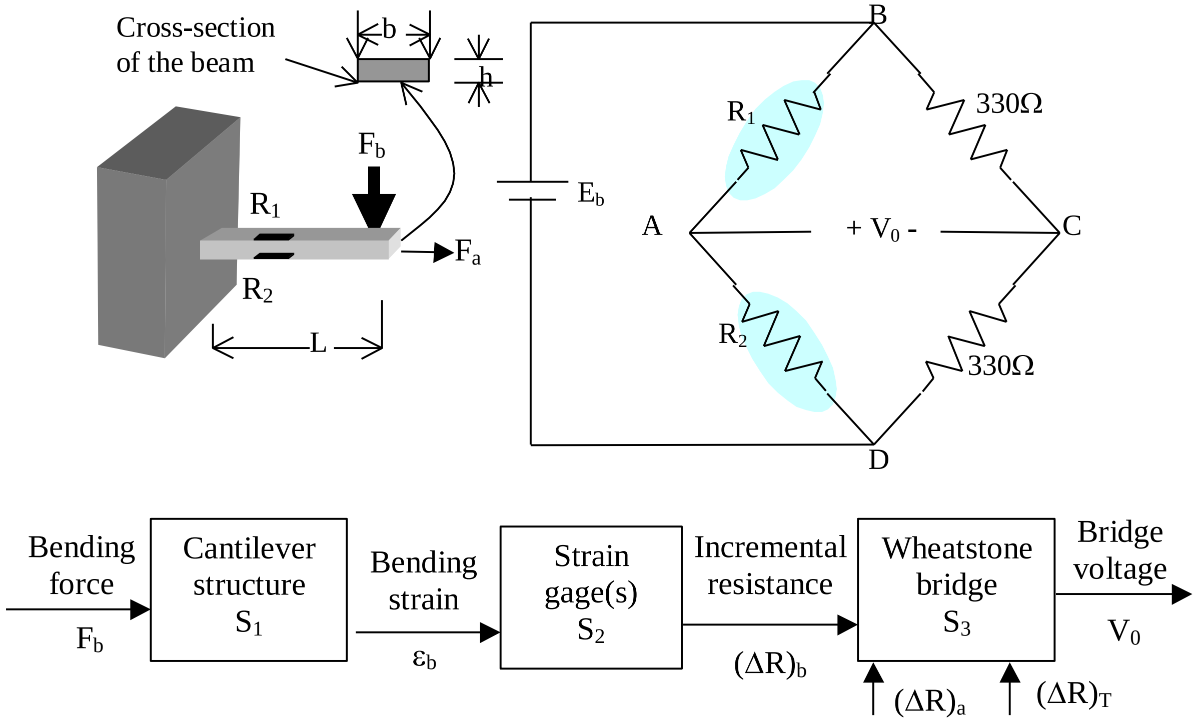


Figure for Pr 8.10

d. **Calculate** ε_b , and V_0 and their **uncertainties** given the following data:

F_b = 10 N ± 1% ; Young's modulus of elasticity = E_Y = 20x10^{10} Nm^{-2}

L = 50 mm; h = 2 mm ± 1%; b = 40 mm.

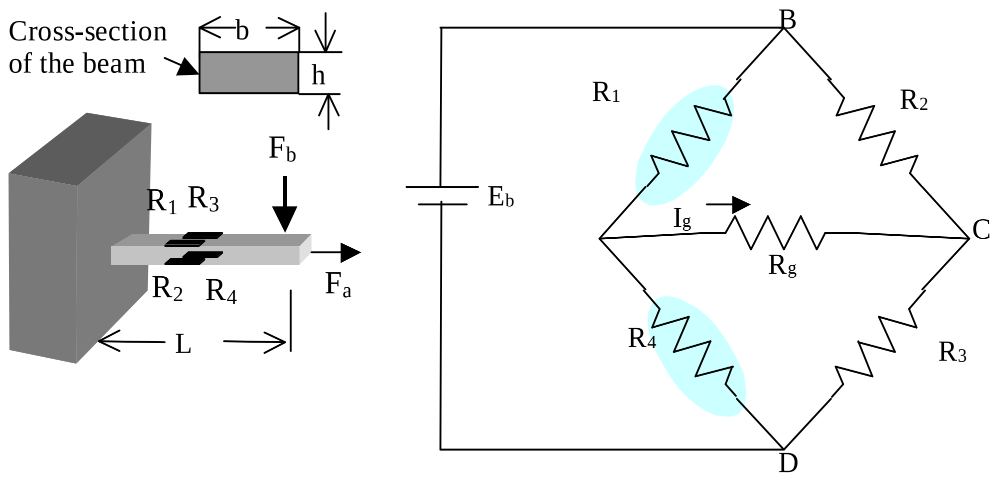


Figure for Pr 8.11

11. Four resistances in a Wheatstone bridge are made up of strain gage elements placed in a cantilevered beam as shown in Figure for Pr 8.11. At no load, $R_1 = R_2 = R_0$, $R_3 = 4R_0$.

e. What is the no-load value for R_4? The functional elements for the arrangement are as shown.

f. **Prove** that the given arrangement compensates the effects of temperature and axial forces (F_a).

12. The block diagram presentation of the system as a general measuring set-up was given in Figure 8.29.

g. Show that the sensitivity S_1 is

$$S_1 = \frac{\sigma_b}{F_b} = \frac{6L}{bh^2}$$

h. **Write** similar expressions (without proof) for sensitivities S_2 , S_3 , S_4 and S_5 .

i. **Calculate** ε_b , σ_b and F_b and their **uncertainties** given the following data

$\theta = 45° \pm 0.2°$; $S_5 = 10\ °/\mu A \pm 0.1\%$; $E_b = 5V \pm 0.4\%$

$R_0 = 50\ \Omega$; $R_g = 100\ \Omega$; gage factor K = 2.0

Young's modulus of elasticity = $E_Y = 2x10^{11}\ Nm^{-2} \pm 1\%$

$L = 480\ mm \pm 1\%$; $h = 3\ mm \pm 1\%$; $b = 60\ mm \pm 1\%$

BIBLIOGRAPHY

Further Reading

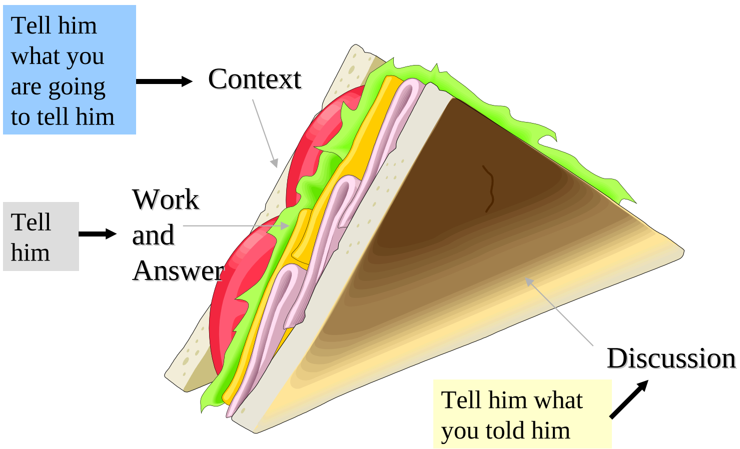
Useful Websites

PRACTICAL AND REPORTING

LABORATORY NOTES AND SHEETS

General Guidelines in Presenting Technical Work

All technical work must be completed with a written report and/or oral presentation. A poorly presented work may be undermined by the reader and ignored. If it is a lab report, it will lead to a poor lab grade that damages your records. The work must be presented in a simple format that describes what we did, informs why we did, reports results of what we did and explains what results mean to us as illustrated in the picture. Hence, it prepares the reader for what is coming, it describes the work, and finally it concludes the work with our interpretations.



Dr. Bellamy illustrates the practice of technical presentation in his book used for IE 201 in form of a presentation sandwich as shown in the figure. The sandwich is made up of three parts as the first slice, filling and the second slice. The first slice is the context gets the reader interested in the work. It is written before the work but with begin with the end in mind. The filling is the actual work. However, it can't stand-alone by itself. The second slice is written after completing the work and it reveals the work we have done.

The Formal Laboratory Report

Introduction

Necessarily, there is a considerable difference between what one does in a report intended for publication in a journal and what one does in a lab report based on one or two weeks of work in a lab course. Some general principles apply to both, however. Failure to write down the results of an experiment and the procedure in which the experiment was done is equivalent to not having done the experiment at all. If the results noted in the lab notebook do not appear in a formal report, your instructor or boss at work will never find out about them or what a fine job you did. It is essential that your work be written up carefully so that your organization and others may be aware of what was done and that you did it. The formal report is the one upon which you and the quality of your work will be judged; it is worth doing well.

What you have learned in other courses, for example in technical writing courses and in physics lab, should be helpful in deciding how to write the formal report. In addition, the instructors of those courses will detail specific requirements for individual courses. The following paragraphs give some of the things that should be included in most reports.

General Requirements

SPACING

The report must be typed on one side of the paper only. Dot matrix output is acceptable. There must be at least a one-inch margin on all sides. Reports that are not clearly legible can result in lesser grades.

FORMAT

The report should have numbered pages, figures and tables. References should be indicated, either as footnotes or as a numbered list (square bracket) at the end of the report, using the format indicated in the instructions given for the laboratory notebook.

SYMBOLS AND ABBREVIATIONS

Try to avoid using abbreviations in the text. Use the full word, such as "ampere," "megahertz," etc. You can use abbreviations in the figures and should use them in tables. For abbreviations, use the recommended IEEE symbols.

UNITS

Use the metric system, specifically the MKSA subsystem of the SI (System Internationale) as recommended by the IEEE.

MATHEMATICAL NOTATION AND FORMULAS

Unless you are an accomplished typist and you have a word processor capable of using math symbols and Greek letters, write formulas neatly by hand, in black ink. Pay special attention to Greek letters, and make a clear difference between small case "e1" (1) and the numeral "one" (1).

Specific Contents of the Report

COVER PAGE

The cover page must include course name and number, experiment title (no more than nine words), your name and the date the report is submitted. Pages must be stapled at the upper left hand corner or held together in a transparent plastic cover of the type sold at bookstores. All pages are to be A4 (21cm by 29.7 cm) in size. Larger pages should be folded to fit within the above-mentioned bounds.

SHORT INTRODUCTION

A concise explanation of the nature of the problem at hand and the purpose of the experiment.

MAIN BODY

This must include circuit diagram(s), formulas used in the calculations, sample calculations (see notebook instructions), tables of data and graphs (see notebook instructions) and relevant procedures.

CONCLUSIONS

Perhaps this is the most important part of the report. It should include what was achieved in the experiment, limitations and advantages of the tested circuit or device, and any other piece of information obtained as a direct result of the experiment, which you think, is relevant. Be concise; do not fill the paper with long irrelevant explanations. Above all, do not quote material published elsewhere, unless it is essential. Just give the appropriate references.

More On Graphs

Graphs should be neatly done on the appropriate type of graph paper. In addition to the instructions given in the notebook instructions, remember that for purposes of reproduction, it is better to avoid using colors. If more than one curve is plotted in a graph, use different types of segmented lines, and identify each by means of a short label with an arrow pointing to the corresponding curve. Schematics must be done using templates in a neat, professional way. Black ink is preferred. Alternatively the report can be prepared using appropriate word processing and graphical software.

One-Page Lab Report

In many experiments, the students are asked to submit a one-page report. Again the sandwich phenomenon is observed. The box below illustrates such a sample report.

BME310, sample report, oxygen measurement.

John Webster (lab partners Tom Edison and Carrie Nation) 8/29/96

Abstract We constructed a PO_2 sensor from parts and used it to measure PO_2. PO_2 of inhaled air was 152 mm Hg whereas that of exhaled air was 114 mm Hg. The sensor time constant was 20 s. The PO_2 of tap water was 76 mm Hg, but increased to 85 mm Hg when stirred.

Introduction and purpose Our metabolism requires oxygen. We need to know the partial pressure of oxygen in the arterial and venous blood to assess how well the lungs and heart are oxygenating the tissues. We will construct a PO_2 sensor from parts, measure its time constant, and use it to measure PO_2 in liquids and gases. We will measure how much the PO_2 changes between inhaled and exhaled air.

Theory When a Pt electrode is biased –0.7 V with reference to a Ag/AgCl electrode, oxygen is reduced at a rate proportional to its partial pressure PO_2,

$O_2 + 2\ H_2O + 4\ e^- \rightarrow 4\ OH^-$. The current from the resulting electrode is linearly proportional to PO_2 in an electrolyte. Since contaminants in blood cause error, the electrode is covered with a plastic membrane impermeable to liquid but permeable to gas. The O_2 from the blood diffuses through the membrane to reach the Pt electrode. Because the electrode consumes O_2, there is a gradient of PO_2 from maximum in the blood to zero at the Pt tip. Therefore we achieve more stable results if we stir the blood to maintain maximal PO_2 as near the Pt tip as possible.

Experimental procedure Following instructions in the notes, we assembled the Clark electrode and connected it to the variable voltage circuit shown in the notes. We used DMMs to obtain current vs. polarizing voltage. We calibrated the differential O_2 analyzer.

Results Note results on the data sheet. Numbered answers to numbered questions follow:

1 Current vs. polarizing voltage is plotted.

2 PO_2 of our exhaled air was 114 mm Hg.

3 We excluded the first portion of the breath because the dead space has room air which does not participate in gas exchange.

4 Difference in O_2 concentration was 38 mm Hg.

One-Page Lab Report

GENERAL GUIDELINES FOR EXPERIMENTS

Preparation for Experiments

The experiments and lab projects in this course intend to develop abilities of students to design and conduct experiments, analyze and interpret data. Students will design and conduct lab experiments and prepare laboratory reports that include the following as a minimum:

- A title page, objectives, preliminary work, apparatus and electronic components used, procedure, data tables, and graphs, discussion of the experimental procedure and results, as well as analysis and interpretation of data with appropriate comments.

The performance will evaluated based the skills and ability of the student to:

- Follow specified experimental procedures to illustrate scientific and engineering principles.
- Operate instruments and electrical engineering equipments.
- Develop experimental procedures, identify operating conditions, configure equipments, and conduct measurements to acquire useful electrical engineering design.
- Examine laboratory data for reliability and accuracy.
- Interpret results.

Before starting any lab work:

- Prepare the necessary theoretical background and the preliminary work before attempting the experiment.
- Discuss your design with the instructor.
- Make a list of electronic components needed for the experiment and gather them with the help of the lab engineer.
- Set your circuit up on the bread-board and nave your friend and/or the lab engineer check the connections before applying power to it.
- Make an estimate of the normal current level before you apply the power. Set the current limits to their minimum values on the D.C. supplies, connect an ammeter to the line and apply power. If you see an abnormal amount of current drawn by the circuit, immediately interrupt the power and recheck your connections and if necessary your design.

A current significantly larger than the expected would mean the existence of one or more of the following:

- A design error;

- A wrong connection;
- A damaged component (short or open). The damage could be there before or could have occurred immediately after the application of the power.

Always remember to turn the D.C. supply off before making any modifications in the circuit to protect semiconductor devices from getting damaged.

Summary of Operation of Oscilloscopes

Since the oscilloscope is probably the most important tool of a circuit designer, trouble shooter, or instrumentation engineer and since many students lack enough understanding of its actual operation to be able to use effectively use its controls, a brief explanation will be presented here. Further information about the oscilloscope is provided in the course material.

As evident, the utility of the scope lies in the fact that it displays graphically voltage functions in a circuit. This effect results from a beam of electrons being swept across the CRT face by plates with varying potentials. The face of the CRT is coated with phosphors which glow when struck by electrons. At higher sweep rates, the beam of electrons will strike the phosphors again before they have stopped glowing from this first contact. The result is a continuous history of the voltage fluctuation in the circuit.

When the beam reaches the right side of the screen it is quickly brought back to the left side where another trace starts. If no attempt at synchronization is made, the beam will not begin at the same point during each cycle and the display will appear to float across the screen (this is analogous to the vertical rolling of the picture on a TV set). To eliminate this difficulty, internal circuitry is provided which holds the beam until a preset threshold voltage is reached (at which time the trigger, which may be a type of multivibrator, is tripped and the beam again sweeps across the screen). This is the essence of what is meant when someone uses the word "**trigger**" in reference to the oscilloscope. With this, the problem with floating is eliminated and the display will appear to be stable.

For reasons which are to be explained in the chapter on display devices, it is often desirable to "trigger" the sweep on waveforms which are external to the scope circuitry. This is accomplished using the front panel switch labeled Triggering Source. When this switch is in the internal (**INT**) position, a sample of the signal being displayed on **CH1** is applied to the trigger circuit, and used for synchronizing the sweep generator. When in the external (**EXT**) position, a signal from some outside source is applied to the trigger circuit, and the sweep is synced to this signal. Note that the sync signal from an outside source must be applied through the external (**EXT**) connector. The **LINE** position of the switch synchronizes the sweep with the **60 Hz** line voltage.

The remainder of the controls, along with the other pieces of lab equipment, are fairly self explanatory and will be illustrated through experimentation.

IMPORTANT NOTE: In all exercises and all others accurately record all data and observations, include them in your report.

WARNING: Only turn the intensity as high as required for a legible display. Too high of an intensity will decrease CRT life and may cause irreparable damage to the phosphors. It will also harm your eyes.

MEASUREMENT AND ERROR

Objective: This exercise is designed to:

1. Familiarize the student with the equipment to be used throughout the semester;
2. Develop specific techniques which are frequently required in subsequent laboratory exercises;
3. Refresh the student`s knowledge on some important fundamental topics.

Preliminary Work

1. Make a detailed study of the knobs and controls on the oscilloscope panel the frequency meter, voltmeter and function/signal generator.
2. List and memorize color codes for identification of the resistors and capacitors.
3. Design experimental procedures to measure:
    a. Input impedance of the oscilloscope;
    b. Output impedance of the signal generator.

Experimental Procedure

1. Apply a 1 kHz - 5 Volts peak to peak sine wave from the signal generator into a digital voltmeter and frequency counter. Wait for 5 minutes for the system to warm-up and stabilize. Then, measure the values for the voltage and frequency at every 30 seconds. Take 10 readings for each and record them with the highest precision possible.
2. Apply a 1 kHz - 5 Volts peak to peak sine wave into channel 1 of the CRO:
3. Measure the amplitude using X1 probe including the error in the measurement;
4. Repeat (a) using X10 & X100 probes;
5. Measure the period and frequency using CRO and include the uncertainties;
6. Study functions of triggering level control and slope control knobs.
7. Determine the input impedance of the oscilloscope and output impedance of the function generator using the procedures you have developed.
8. Identify the values of resistors and capacitors provided by using the color codes. Measure the values of 10 resistors from the same type and find the magnitude of error in your identifications using the color codes.

Results and Discussions:

1. Record all measurements you have made. Compare them with your expectations. Discuss any difference you have encountered.
2. Discuss the accuracy of your measurements referring to the sources of errors. Study the accuracy and resolution of the measurement for the amplitude and frequency using:
   a. Oscilloscope only for the measurements;
   b. A digital multimeter and frequency counter for the measurements. Refer to manuals for the oscilloscope, multimeter and frequency counter if necessary (you can ask the help of the lab engineer). How many significant digits you can use to express the results in each case?
3. Discuss the relevance of the output impedance of the signal generator and input impedance of the oscilloscope. What limitations they introduce into your measurements? What is the significance of the input capacitance?
4. What can you say about the reliability of the measurement using oscilloscope?
5. Using statistical analysis find out the errors in amplitude and frequency of the signal delivered by the signal generator.
6. Determine the error in the resistance value and compare it to the tolerance. Express the resistance using significant digits only.

DETERMINING THE CHARACTERISTIC OF AN INCANDESCENT LAMP

Objective: This experiment is intended to:

1. Let students develop experimental protocols, set and conduct experiments;
2. Develop students' skills in data collection and analyzes.
3. Apply statistical techniques in data analyzes and presentation of experimental results.

Preliminary Work

1. Draw the circuit diagram of a measurement setup that you will use to determine the variation of the lamp resistance as it heats up from no power to full power level.
2. Write down the formula that express the change of resistance with temperature for metals.
3. Determine temperature coefficient of the resistance for tungsten from reference books or from the web.

Preparations Before the Experiment

1. Read and record the values written on the lamp (voltage, current, power etc).
2. Make sure that the lamp is sitting in its socket.
3. Determine the accuracy and precision of the ammeter and voltmeter that you will use in the experiment.

Experimental Procedure

1. Make your connections properly to the power supply and care for safety.
2. Measure the resistance of the lamp using an ohmmeter before you connect it to your circuit. Set the DC power supply to the nominal voltage of the lamp.
3. Connect the lamp to the DC power supply (with power switch off!), with an ammeter in series. Switch the power supply on and watch the ammeter and try to estimate the time taken for the lamp to reach into thermal stability.
4. Set the measurement circuit with the power supply, the resistance box, an ammeter, voltmeter and the lamp. Make a table of voltage and current readings for 10 settings of the resistance box from almost no current to full current into the lamp. Record all your readings to the highest precision possible. Wait for sufficient time between steps so that the steady state value is reached. Record also the waiting time between steps.

5. Repeat the measurement in the reverse order (as the current decreases from full current to almost no current).

6. Repeat (4) and (5) at least 5 times.

Results

1. Transfer the data you recorded during the experiment into an EXCEL Sheet.

2. Determine the average voltage and current values and errors in them for experiment steps as you increase the current and you decrease the current separately.

3. Calculate the resistance of the lamp and the power dissipated for each step of the experiment including the errors.

4. Draw the scatter diagram of the resistance against the power for increasing and decreasing lamp current.

5. Obtain the linear regression lines (best fit) for your scatter plots and obtain the equations of the lines.

Discussions and Conclusions

1. Comment on the errors in your measurements and their effects on the results obtained.

2. Compare the changes in the resistance as the lamp current increases and decreases. How good your straight line fit? How much is the maximum deviation? Is it within the expected error range?

3. Do you have differences between the regression lines for increasing and decreasing lamp currents? If so, explain the reasons.

4. Determine the coefficient of variation of the resistance with the power in the lamp. Estimate the temperature of the lamp using the coefficient you have found and the temperature coefficient of resistance of tungsten you obtained in the preliminary work.

5. Submit 1-page lab report to Eng. Abdulmuttalib in paper and your EXCEL file to Dr. Baha electronically.

DETERMINING THE CHARACTERISTIC OF A CAPACITOR

This experiment is intended to let students design their own experiments to determine the characteristic of an electronic component and verify their results using different methods including technical libraries. Students will submit full report in accordance with the rubric for ABET Program Outcome 3(b).

Capacitors to be used

1. Aluminum electrolytic types 1000 to 2200 μF, 100 μF and 4.7 μF
2. Tantalum 4.7 μF
3. Non-electrolytic types polyester, polystyrene, and mica at various values available in the lab.

Reminder for the experimental procedures

1. Before the lab, determine the equivalent circuit of the lab and using the library/web resources identify the components of the equivalent circuit for each category of capacitor.
2. Design a test circuit and carefully select the test equipment. Write down the model and serial number for each equipment that you use into your report sheet.
3. Be careful about the polarity of electrolytic capacitors in connecting them to the circuit.
4. Prepare the experimental protocol, set and conduct experiments; make sure that you repeat each step until you achieve statistical stability of the results.
5. Make sure that you use all measurement techniques available (i.e. square wave testing, sine wave testing etc).
6. Transfer the data you recorded during the experiment into an EXCEL Sheet.
7. Determine the average voltage and current values and errors in them for each step of the experiment.
8. Calculate the components of the equivalent circuit for each category of the capacitor, compare and contrast results obtained using different experimental approaches.
9. Compare the most reliable experimental results with your expectations in step 1.
10. Discuss the deviation of the device characteristic from the ideal one due to non-ideal components and its effect in selecting the capacitance in specific applications.

REGULATED POWER SUPPLY

The purpose of this exercise is to make the student familiar with different power sources that are used in medical equipment. In this respect:

1. The behavior of capacitors and batteries will be studied as energy as energy storage elements.
2. The performance of electronic power supplies will be determined. Every student is going to by his own power supply for this experiment. An unregulated one is preferred and a voltage regulator will be added to it.
3. The switching power supply will also be discussed.
4. Fuses used in electronic circuits will be studied.

Preliminary Work

1. Study the capacitors from the notes and from other sources. Design a circuit to determine the equivalent resistance of the capacitor.
2. Make a table of comparison for different types of batteries that are used in the medical instruments.
3. Study regulated power supplies from books on electronics and power supplies. Design a regulated power supply that will deliver 5 volts and 0.5 amp. Use an electronic components catalog (i.e. an RS catalog) to select electronic components you need for your design.
4. Draw the circuit diagram for the power supply you brought. Calculate the filter capacitance for 500 mA load and ripple voltage of 1 V. Decide whether the existing capacitance is: a) just right? b) Insufficient c) More than enough.
5. Make a web search on fuses and study the fuse parameters. Make a table of comparison between fuses used in protecting semiconductor devices, small electrical motors and electric blankets.

Experiment

1. Build the circuit you designed above for testing the resistance of a capacitor. Measure the resistance on three different types of capacitors and compare your findings to values given in component catalogs.
2. Measure the ripple voltage on the unmodified power supply for the following cases:

   a) No load. b) 50mA load. c) 500 mA load.
3. Measure the output voltage on the unmodified supply under:

   a) No load. b) 50 mA load. c) 500 mA load.

4. Modify your power supply and add a series regulator to it to obtain a regulated output voltage of 5 volts with current capacity of 500 mA. Repeat steps 2 & 3 for your regulated power supply.

5. Study a switching type power supply that will be provided to you in the lab. Identify its components and determine its principle of operation.

6. Examine batteries that you collect yourself and given in the lab. Try to determine some salient characteristics of two types that you choose.

7. Collect samples of fuses used in electronics. Classify them with respect to size, tripping time and other relevant parameters. Indicate a few application areas for each type.

TERM PROJECT

Assignment: When I need to change the battery of the wall clock?

Team size: 3 to 4 students,

Due date: 21 May 2011 (18 Jumada II 1432)

Student Outcomes to be satisfied by the assignment: b, d, f, k, l

Important Questions to Answer

- What is a wall clock, what are the easily available models in the market?
- What is the electrical characteristic of the clock you have?
- What are the batteries available to power your clock?
- How the batteries are found in the market and what is the price range for each type?
- What is the shelf and expected service life for the battery you choose for your clock?

Duties

1. Establish your teams and distribute team roles
2. Determine tasks to be done and expected time needed for each one
3. Distribute responsibilities of fulfilling tasks to team members
4. Make a time plan with definite deadlines
5. Plan and hold regular follow up meetings and take meeting minutes
6. Collect the information from the members about the results achieved and analyze the results
7. Write down the final report and submit.

Elements of the Report

1. Cover page and overall organization of the report
2. What is a clock, when and why it is used?
3. Wall clocks available in the market and their price ranges
4. Experimental procedures to determine the clock characteristics
5. Tool selection and use
6. Clock characteristics – data tables
7. Statistical analyses of the experimental data
8. Graphical presentation of the clock characteristics
9. Interpretation of the clock characteristics and modeling

10. What is a battery?
11. Types of batteries and battery manufacturers
12. Characteristics of batteries
13. Availability, market survey, shelf life
14. Matching the clock characteristics into battery characteristics
15. Selection of a proper battery for your clock
16. Estimated lifetime for the selected battery
17. Team setting
18. Time plan
19. Work sharing and responsibilities
20. What has been learnt from the project

Each element takes 5 marks totaling to 100 for the project.

REFERENCES

J.G. Webster, *Electrical Measurement, Signal Processing and Displays*, CRC Press

R.C. Dorf, *The Electrical Engineering Handbook*, CRC Press, 2000

Author, *Principles and Applications of Electrical Engineering,* McGraw-Hill,

M.W. Earley, J.S. Sargent, J.V. Sheehan and J.M. Caloggero (eds), *National Electricity Code Handbook*, 10th ed., NFPA, 2005

R.B. Northrop, *Introduction to Instrumentation and Measurements*, 2nd ed., CRC Press, 2005.

W.D. Stanley, J.R. Hackworth and R.L. Jones, *Fundamentals of Electrical Engineering and Technology,* Delmar Cengage Learning, 2006

Shultz, Grob's Introduction to Electronics, McGraw-Hill, 2007

A.D. Helfrick and W.D. Cooper, *Modern Electronic Instrumentation and Measurement Techniques*, Prentice-Hall, 1990

APPENDICES

A – QUANTITIES, UNITS AND STANDARDS

Basic and Derived Units

In all conversations, the physical quantities are presented with their proper values compared to the standard, the units. The internationally established (SI) units are the meter for length, the kilogram for mass, and the second for time, abbreviated as the mks system of units. Although the mks system is commonly used in engineering, the cgs system of units is an absolute system of units that is widely used in science. This system is based on the centimeter, gram mass, and second as basic units. Disadvantages include the fact that the derived units for force and energy are too small for practical purposes and that the system does not combine with the practical electrical units to form a comprehensive unit system.

The British engineering system of units is a gravitational system of units and is based on the foot, pound-force, and second as basic units. The system is the one that has been used in the United States. The derived unit of mass is lbf-s2/ft and is called a slug. Table A.1 list the basic and auxiliary units used in the mks system.

Table A.1 Basic and auxiliary units in the mks system

Basic Units			
Quantity	Unit	Symbol	Dimension
Length	Meter	M	
Mass	kilogram	Kg	
Time	second	S	
Electric current	ampere	A	
Temperature	Kelvin	K	
Luminous intensity	candela	Cd	
Amount of substance	mole	Mol	
Auxiliary Units			
Plane angle	radian	rad	$m \cdot m^{-1} = 1$
Solid angle	steradian	sr	$m^2 \cdot m^{-2} = 1$

There are many units used in engineering driven from the base units. Table A.2 lists the derived units mostly used electrical engineering applications.

Table A.2 Derived units

Derived Quantity	Unit	Symbol	Dimension
Acceleration	Meter per second squared	m/s^2	
Angular acceleration	Radian per second squared	rad/s^2	
Angular velocity	Radian per second	rad/s	
Area	Square meter	m^2	
Density	Kilogram per cubic meter	kg/m^3	
Dynamic viscosity	Newton second per square meter	Ns/m^2	$m^{-1}kgs^{-1}$
Electric capacitance	Farad	F, C/V	$m^{-2}kg^{-1}s^4A^2$
Electric charge, quantity of electricity	Coulomb	C	As
Electric field strength	Volt per meter	V/m	$mkgs^{-3}A^{-1}$
Electric resistance	Ohm	Ω, V/A	$m^2kgs^{-3}A^{-2}$
electric conductance	siemens	S, A/V	$m^{-2}\cdot kg^{-1}\cdot s^3\cdot A^2$
Entropy	Joule per Kelvin	J/K	$m^2kgs^{-2}K^{-1}$
Force	Newton	N	$mkgs^{-2}$
Frequency	Hertz	Hz	s^{-1}
Illumination	Lux	Lx	$m^{-2}cdsr$
Inductance	Henry	H, Wb/A	$m^2kgs^{-2}A^{-2}$
Kinematic viscosity	square meter per second	m^2/s	
Luminance	Candela per square meter	cd/m^2	
Luminous flux	Lumen	Lm	cdsr
Magnetic field strength	Ampere per meter	A/m	
Magnetic flux	Weber	Wb, V.s	$m^2kgs^{-2}A^{-1}$
Magnetic flux density	Tesla	T, Wb/m^2	$kgs^{-2}A^{-1}$

Derived Quantity	Unit	Symbol	Dimension
Magnetomotive force	Ampere turn	A	A
Power, radiant flux	Watt	W, J/s	m^2kgs^{-3}
Pressure, stress	Pascal	Pa (N/m^2)	$m^{-1}kgs^{-2}$
Radiant intensity	Watt per steradian	W/sr	$m^2kgs^{-3}sr^{-1}$
Specific heat	Joule per kilogram Kelvin	J/kg K	$m^2s^{-2}K^{-1}$
Thermal conductivity	Watt per meter Kelvin	W/m K	$mkgs^{-3}K^{-1}$
Velocity	Meter per second	m/s	
Voltage, electric potential difference, electromotive force	Volt	V, W/A	$m^2kgs^{-3}A^{-1}$
Volume	Cubic meter	m^3	
Wave number	1 per meter	m^{-1}	
Work, energy, quantity of heat	Joule	J	m^2kgs^{-2}

There are other SI derived units whose names and symbols include SI derived units with special names and symbols. Examples of them are given in Table A.3.

Table A.3. Examples of SI derived units with special names and symbols

Derived quantity	**Name**	**Symbol**
moment of force	newton meter	N·m
surface tension	newton per meter	N/m
angular velocity	radian per second	rad/s
angular acceleration	radian per second squared	rad/s^2
heat flux density, irradiance	watt per square meter	W/m^2
heat capacity, entropy	joule per Kelvin	J/K
specific heat capacity, specific entropy	joule per kilogram kelvin	J/(kg·K)
specific energy	joule per kilogram	J/kg
thermal conductivity	watt per meter kelvin	W/(m·K)

energy density	joule per cubic meter	J/m^3
electric field strength	volt per meter	V/m
electric charge density	coulomb per cubic meter	C/m^3
electric flux density	coulomb per square meter	C/m^2
Permittivity	farad per meter	F/m
Permeability	henry per meter	H/m
molar energy	joule per mole	J/mol
molar entropy, molar heat capacity	joule per mole kelvin	J/(mol·K)
exposure (x and γrays)	coulomb per kilogram	C/kg
absorbed dose rate	gray per second	Gy/s
radiant intensity	watt per steradian	W/sr
Radiance	watt per square meter steradian	$W/(m^2 \cdot sr)$
catalytic (activity) concentration	katal per cubic meter	kat/m^3

Relationships of the SI derived units with special names and symbols and the SI base units are schematically illustrated in Figure.A.12. In the first column, the symbols of the SI base units are shown in rectangles, with the name of the unit shown toward the upper left of the rectangle and the name of the associated base quantity shown in italic type below the rectangle. In the third column the symbols of the derived units with special names are shown in solid circles, with the name of the unit shown toward the upper left of the circle, the name of the associated derived quantity shown in italic type below the circle, and an expression for the derived unit in terms of other units shown toward the upper right in parenthesis. In the second column are shown those derived units without special names [the cubic meter (m3) excepted] that are used in the derivation of the derived units with special names. In the diagram, the derivation of each derived unit is indicated by arrows that bring in units in the numerator (solid lines) and units in the denominator (broken lines), as appropriate.

Two SI derived units with special names and symbols, the radian, symbol rad, and the steradian, symbol sr (bottom of the third column of the diagram), are shown without any connections to SI base units – either direct or through other SI derived units. The reason is that in the

[2] From http://physics.nist.gov/cuu/units

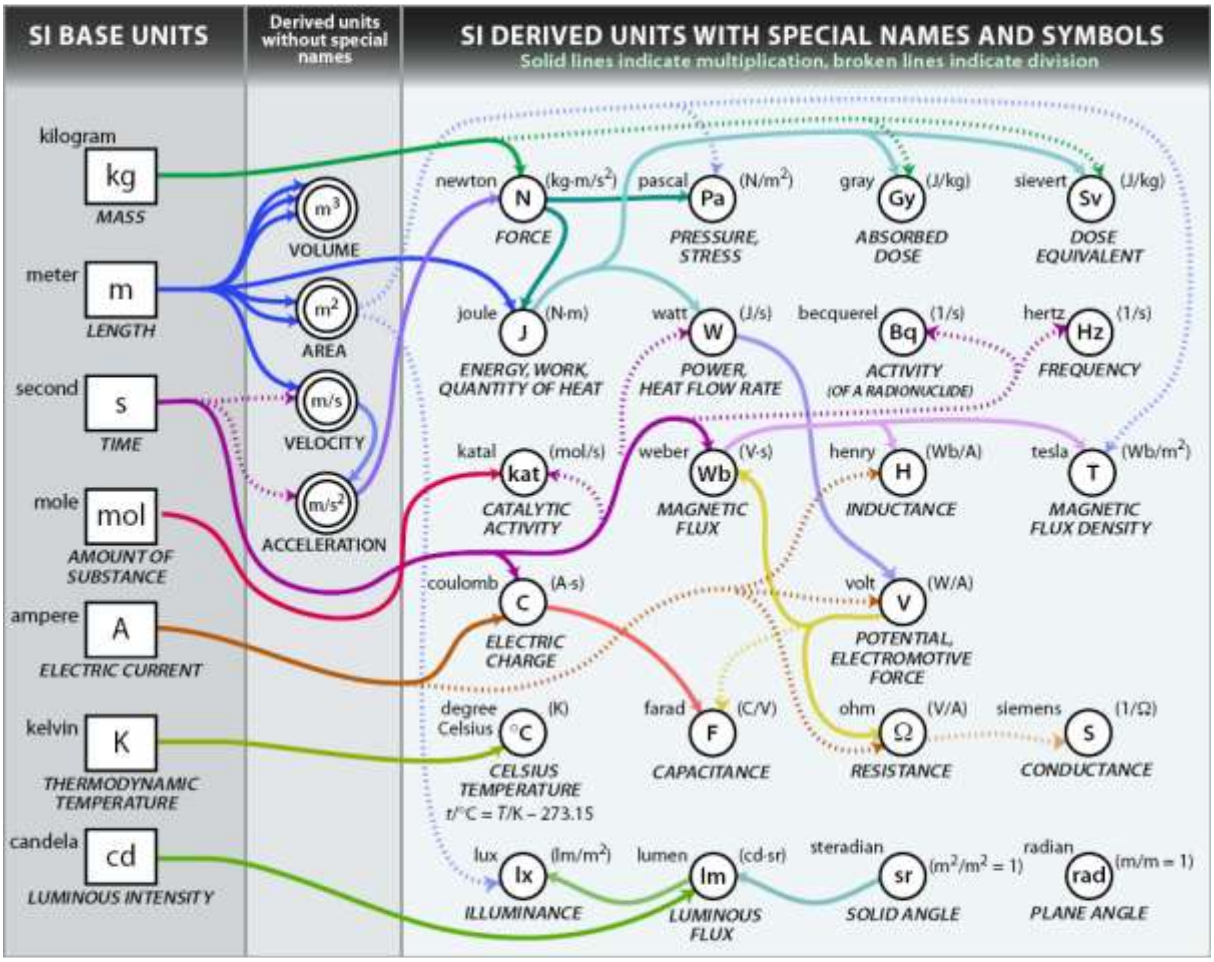


Figure. A.1. Relationships between SI derived and base units

SI, the quantities plane angle and solid angle are defined in such a way that their dimension is one – they are so-called dimensionless quantities. This means that the coherent SI derived unit for each of these quantities is the number one, symbol 1. That is, because plane angle is expressed as the ratio of two lengths, and solid angle as the ratio of an area and the square of a length, the SI derived unit for plane angle is m/m = 1, and the SI derived unit for solid angle is m2/m2 = 1. To aid understanding, the special name radian with symbol rad is given to the number 1 for use in expressing values of plane angle; and the special name steradian with symbol sr is given to the number 1 for use in expressing values of solid angle. However, one has the option of using or not using these names and symbols in expressions for other SI derived units, as is convenient.

The unit "degree Celsius,'' which is equal to the unit "kelvin," is used to express Celsius temperature *t*. In this case, "degree Celsius'' is a special name used in place of "kelvin.'' This equality is indicated in the diagram by the symbol K in parenthesis toward the upper right of the °C circle. The equation below "CELSIUS TEMPERATURE'' relates Celsius temperature *t* to thermodynamic temperature *T*. An interval or difference of Celsius temperature can, however, be expressed in kelvins as well as in degrees Celsius.

Standards

International standardization is an absolute must in today's world. World standards have been established as:

- The meter is the length equal to 1 650 763.73 wavelengths of radians in vacuum corresponding to the unperturbed transition between levels $2P_{10}$ and $5d_5$ of the atom of krypton 86, the orange-red line.
- The kilogram is the mass of a particular cylinder (of diameter 39 mm and height 39 mm) of platinum-iridium alloy, called the International prototype kilogram, which is preserved in a vault at Sevres, France, by the International Bureau of Weights and Measures.
- The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the fundamental state of the atom of cesium 133.
- The ampere is a constant current that, if maintained in two straight, parallel conductors of infinite length, of negligible circular cross sections, and placed 1 meter apart in a vacuum, will produce between these conductors a force equal to $2x10^{-7}$ Newton per meter of length.
- The Kelvin is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water. (Note that the triple point of water is 0.01 degree Celsius.)
- The candela is the luminous intensity, in the direction of the normal, of a black body surface 1/600 000 square meter in area, at the temperature of solidification of platinum under a pressure of 101 325 Newton per square meter.
- The mole is the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12.

The two auxiliary units of SI are defined as follows

- The radian is a unit of plane angular measurement equal to the angle at the center of a circle subtended by an arc equal in length to the radius. (The dimension of the radian is zero since it is a ratio of the quantities of the same dimensions.)
- The steradian is a unit of measure of solid angles that is expressed as solid angle subtended at the center of the sphere by a portion of the surface whose area is equal to the square of the radius of the sphere. (The dimension of the steradian is also zero, since it is a ratio of the quantities of the same dimension.)

B – OPERATIONAL AMPLIFIERS

Characteristics and basic amplifiers configurations using op-amps

Amplifiers are devices that increase the voltage, current and power of an input signal. The input may

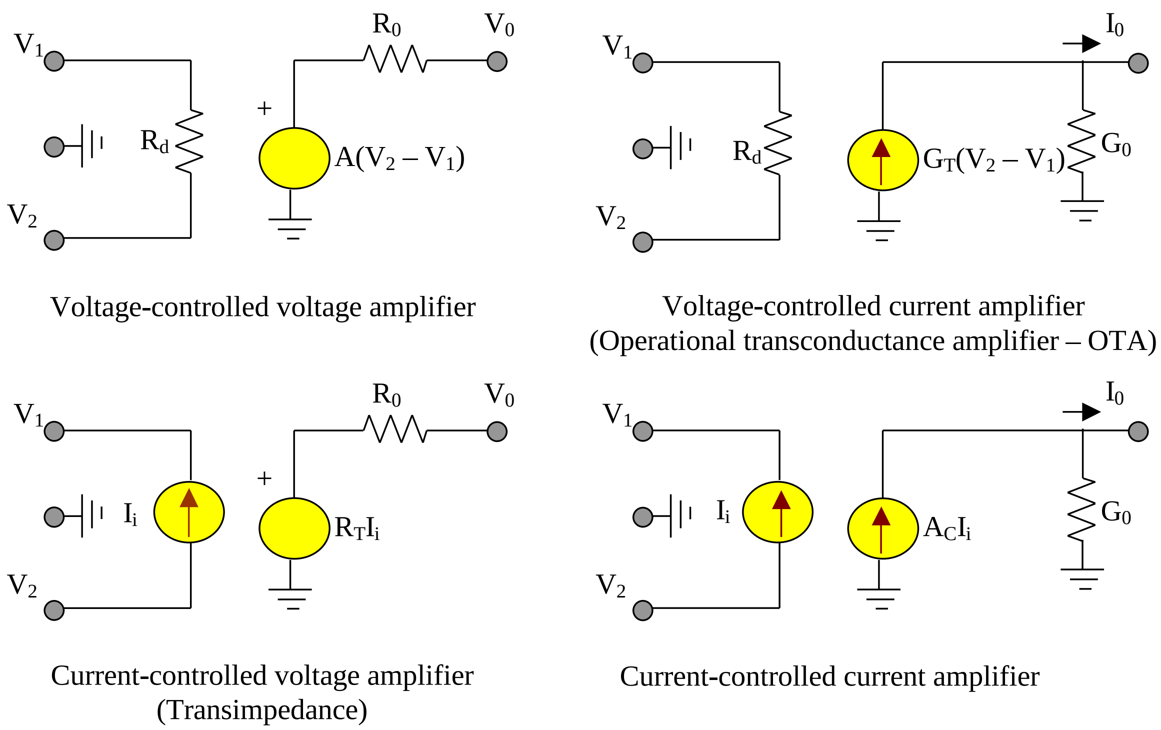


Figure B.1 Types of amplifiers

be in form of voltage or current. Accordingly, the amplifiers can be classified into four basic groups as illustrated in Figure B.1.

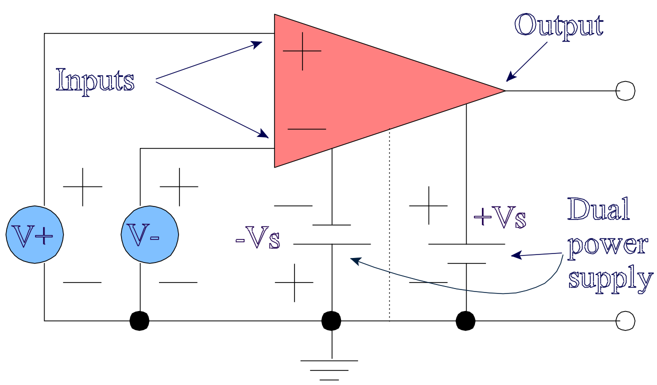


Figure B.2 Op-amp as a circuit element

Operational amplifiers (op-amps) are very versatile devices that are used in various signal amplification and processing applications. Figure B.2 illustrates utilization of an ordinary op-amp as a circuit element. It has two input terminals and one output terminal. It requires a dual symmetrical power supply for the operation. One of the inputs is in phase with the output and it is called the non-inverting input. The other input 180° out of phase with the output and it is called the inverting input.

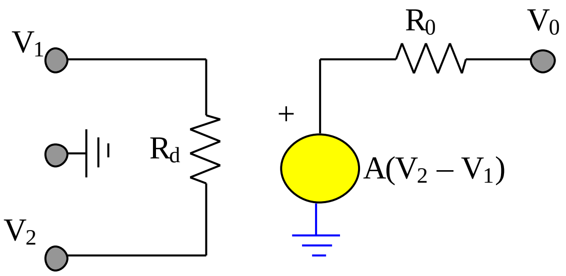


Figure B.3 Diagram of an ideal op-amp

Ordinary op-amp is a voltage-controlled voltage amplifier as illustrated in Figure B.3. When considered ideal, it has the following properties:

$A = \infty$ (gain is infinity)

$V_0 = 0$ when $V_1 = V_2$ (no offset voltage)

$R_d = \infty$ (input impedance is infinity)

$R_0 = 0$ (output impedance is zero)

Bandwidth = ∞ (no frequency response limitations) and no phase shift

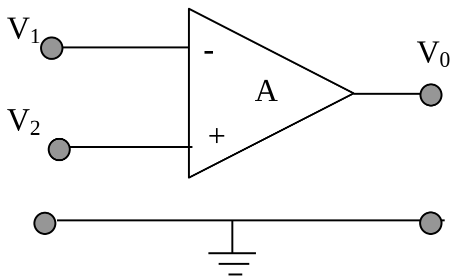


Figure B.4 An ideal op-amp

Of course all op-amps have limitations that let them deviate from ideal characteristics. However, the assumption of ideal characteristics simplifies the modeling and associated calculations, and it is valid for most applications. A symbolic diagram of the op-amp is shown in Figure B.4. There are two golden rules that are driven from the ideal characteristics:

When the op-amp output is in its linear range, the two input terminals are at the same voltage.

No current flows into either input terminals of the op-amp.

Inverting amplifiers

The simplest amplifier configuration is the inverting amplifier shown in Figure B.5. No current flows through the input terminals of the op-amp and the inverting terminal appears at ground potential (virtual ground) since the non-inverting terminal is physically connected to ground. The current from the input is

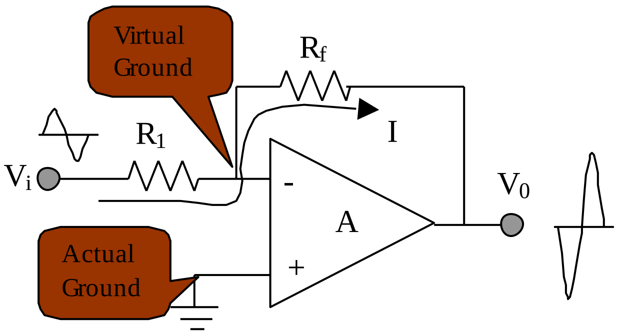


Figure B.5 The inverting amplifier

$I = V_i/R_1$

And it is rooted through the feedback resistor R_f yielding $V_0 = - V_iR_f/R_1$

Following websites contain very useful information about op-amps and their application.

http://www.electronics-tutorials.ws/opamp/opamp_1.html

http://users.ece.gatech.edu/mleach/ece4435/tutorial.pdf

http://holbert.faculty.asu.edu/ece201/opamp.html

C – VISUAL DISPLAYS

C.1 INTRODUCTION

C.1.1 Sources of Light

Light is a form of energy that requires another form of energy for generation. There are two common ways for this to occur: incandescence and luminescence. Incandescence is light from heat energy as in the case of the tungsten filament of an ordinary incandescent light bulb as in Figure C.1. Luminescence is all forms of visible radiant energy due to causes other than temperature. It is also called "cold light", which can take place at normal and lower temperatures as in the case of the fluorescent lamp in Figure C.1.

An incandescent lamp

A fluorescent lamp

Figure C.1 Incandecent and fluorescent lamps

In luminescence, some energy source kicks an electron in the outer orbit of an atom out of its "ground" (lowest-energy) state into an "excited" (higher-energy) state. Then the electron falls back into the ground state by delivering the excess energy in the form of light.There are a number of different types of luminescence, including: electroluminescence, chemiluminescence, cathodoluminescence, triboluminescence, and photoluminescence. Most "glow in the dark" toys take advantage of photoluminescence: light that is produced after exposing a photoluminescent material to intense light. Chemiluminescence is the name given to light that is produced as a result of chemical reactions, such as those that occur in the body of a firefly. Cathodoluminescence is the light given off by a material being bombarded by electrons (as in the phosphors on the faceplate of a cathode ray tube). Electroluminescence is the production of visible light by a substance exposed to an electric field without thermal energy generation.[3]

Fluorescence and photoluminescence are luminescence where the energy is supplied by electromagnetic radiation (rays such as light); photoluminescence is generally taken to mean luminescent from any electromagnetic radiation, while fluorescence is often used only for luminescence caused by ultraviolet light photones, although it may be used for other photoluminescences also.

---

[3] P.D. Rock, A. Naman, P.H. Holloway, Sey Shing Sun, and R.T. Tuenge, "Materials Used in Electroluminescent Displays," http://www.distec.com/Electro.htm, accessed on July 13, 1999.

The fluorescent lamp is a sealed glass tube that contains a small bit of mercury and an inert gas (typically argon, kept under very low pressure). Inside of the glass is coated with a phosphor powder. The tube has two electrodes, one at each end, and as we turn the lamp on, the current flows through the electrical circuit to the electrodes that produce electron clouds. There is a considerable voltage across the electrodes that causes the electrons to migrate through the gas from one end of the tube to the other. This energy changes some of the mercury in the tube from a liquid to a gas that generates a plasma (a gas made up of free-flowing ions and electrons). As electrons and charged atoms move through the tube, some of them collide with the gaseous mercury atoms. These collisions excite the atoms, bumping electrons up to higher energy levels. When the electrons return to their original energy level, they release light photons mostly in the ultraviolet wavelength range.

Phosphors are substances that give off light when they are exposed to light. When a photon hits a phosphor atom, one of the phosphor's electrons jumps to a higher energy level and the atom heats up. When the electron falls back to its normal level, it releases energy in the form of another photon. This photon has less energy than the original photon, because some energy was lost as heat. An incandescent lamp sheds off a lot of the energy as heat. Yet, a typical fluorescent lamp is four to six times more efficient than an incandescent one. In a fluorescent lamp, the phosphor gives off white light we can see. Manufacturers can vary the color of the light by using different combinations of phosphors.

Electroluminescent (EL) devices include light emitting diodes (LEDs), as well as EL displays (ELDs) which are matrix-addressed devices that can be used to display text, graphics, and other computer images. There is a new generation of leds, called the organic LED (OLED), in which an organic compound replaces the phosphor. Such systems can be used in television screens and computer displays. EL is also used in lamps and backlights.

C.1.2 Visual Displays

Visual displays can be divided into two gross categories as the cathode ray tube (CRT) and flat panel (screen) displays. Most video monitors today use the traditional CRT, which works on the same scientific principle as a television set. The vacuum tube produces an image when an electron beam strikes the phosphorescent surface inside the monitor. Traditional CRT technology is analog technology. Flat panel displays (FPD) are the technologies of the future and a variety of technologies are currently competing. Among these are advanced liquid crystal (LCD), plasma discharge (PDP), and field emission (FED) displays. Flat panel displays use digital technology. Figure C.2 shows examples of a CRT and a flat panel displays.

A CRT display

A flat panel display

Figure C.2 Examples of CRT and flat panel displays

The CRT technology has been used in oscilloscopes for scientific applications, in television sets for entertainment, and in monitors as a desktop PC peripheral. It's based on universally understood principles and employs commonly available materials. The result is cheap-to-make monitors capable of excellent performance, producing stable images in true color at high display resolutions. The technology has following salient advantages:

- phosphors have been developed over a long period of time, to the point where they offer excellent color saturation at the very small particle size required by high-resolution displays
- the fact that phosphors emit light in all directions means that viewing angles of close to 180 degrees are possible
- since an electron current can be focused to a small spot, CRTs can deliver peak luminance as high as 1000 cd/m^2 (or 1000 nits)
- CRTs use a simple and mature technology and can therefore be manufactured inexpensively in many industrialized countries
- whilst the gap is getting smaller all the time, they remain significantly cheaper than alternative display technologies.

Yet, immaterial of how good they are, CRT displays have significant limitations as:

- they're too big, heavy and bulky
- they're power hungry, suck too much electricity - typically 150W for a 17-inch monitor
- their high-voltage electric field, high- and low frequency magnetic fields and x-ray radiation have proven to be harmful to humans in the past
- the scanning technology they employ makes flickering unavoidable, causing eye strain and fatigue
- their susceptibility to electro-magnetic fields makes them vulnerable in military environments
- their surface is often either spherical or cylindrical, with the result that straight lines do not appear straight at the edges. Also, there may be color variations across the screen.

C.2 CATHODE RAY TUBE (CRT)

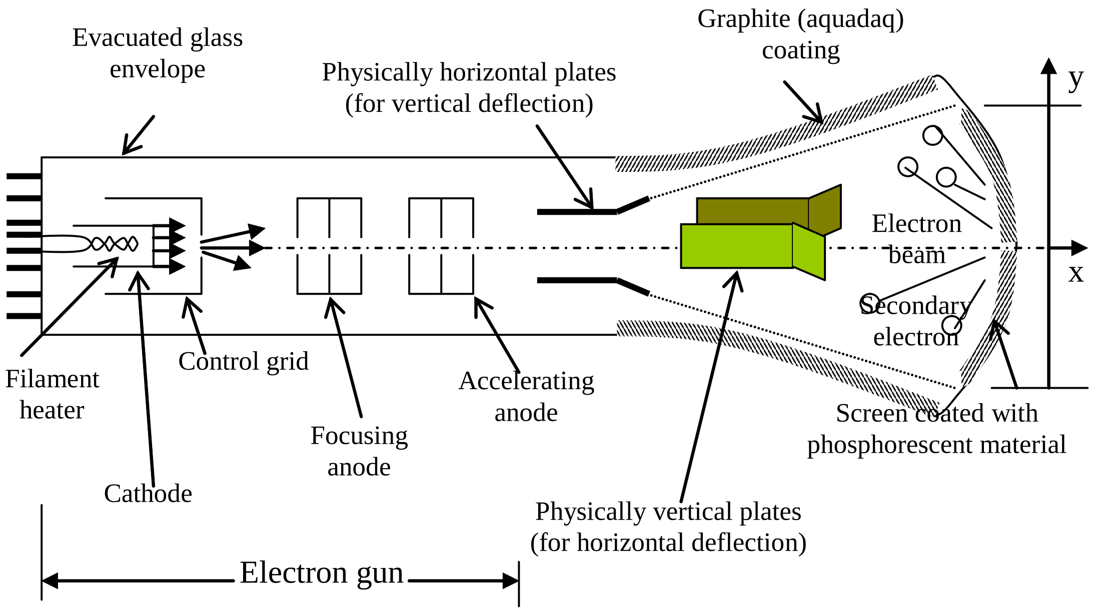


Figure C.3 Schematic diagram of a cathode ray tube (CRT)

C.2.1 The schematic of the CRT

The schematic diagram of a cathode ray tube is shown in Figure C.3. The CRT consists of four basic parts:

1. An evacuated glass envelope
2. An electron gun for producing a beam of electron (a heater, a cathode, a control grid, a focusing anode, and an accelerating anode)
3. Electrostatic (or electromagnetic) structure for deflecting the electron beam (usually rectangular set of horizontal and vertical plates)
4. A phosphorescent screen for converting the kinetic energy of the electron beam into light energy.

C.2.2 The Electron Gun

The electron gun produces and accelerates electrons. It contains controls to adjust the brightness and sharpness of the display. It is composed of the heater, cathode, control grid, and accelerating and focusing anodes. All electrical connection to the electron gun are made through the pins at the back of the tube.

The heater: It is a filament that provides heat energy to the cathode. Generally it is fed by 6.3 volts AC voltage.

The cathode: A nickel cylinder covered by a layer of barium and strontium oxides to obtain high electron emission.

The control grid: Cylindrical and partially surround the cathode. Biased negative with respect to the cathode, and controls the intensity (density) of the electron beam. Hence, it provides the intensity control.

Typically, the control grid voltage of (with respect to cathode) -20 volt provides a high intensity beam, -50 volt a slightly intense beam and -100 volt may cut-off the beam (screen blanking). The grid voltage is adjusted through the *intensity (brightness)* control on the front panel of the oscilloscope. The Z-input that is provided at the back of some laboratory oscilloscopes can also be used to control the brightness using external signals.

The anodes: Focusing and accelerating anodes are cylindrical with diaphragm mounted inside, the diaphragm having a hole at its center. They form an electrostatic lens that brings the beam to a focal point on the screen, and makes this electrons attain high speed.

C.2.2.1 Focusing

The focusing anode is generally placed between two accelerating anodes and kept at a lower potential. Hence, an electrostatic force is applied to the electrons, so that the electron beam becomes parallel first and then focused to a point on the screen. The voltage difference between the focusing anode and the accelerating anodes can be set externally using the *focus* control at the front panel to adjust the size of the spot as illustrated in Figure C.4.

C.2.2.2 Acceleration of Electrons

Two plates separated by a distance L and have voltages V_b and V_a produces an electric field strength or electric stress $\varepsilon = dV/dl = (V_b - V_a)/L$. The electric field strength is also defined as the force per unit of (+) charge; $\varepsilon = F/q$. Let $V_{ba} = V_b - V_a$, then $V_{ba}/L = F/q$

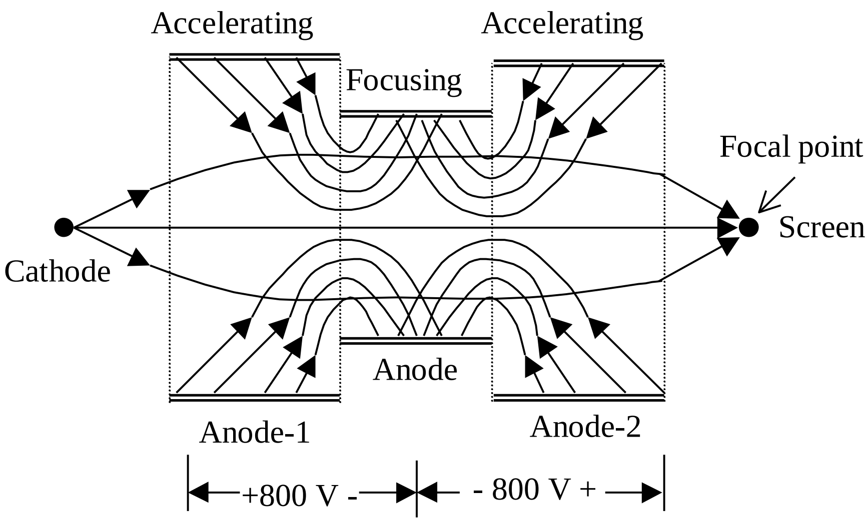


Figure C.4 The focusing of electron beam

Work done (energy) = force x distance = FL = qV_{ba} .... (C.1)

The potential energy (PE) applied to the electron as it enters into this field = $q(V_b - V_a) = -e(V_b - V_a)$. Therefore

$PE = eV_{ab}$ .... (C.2)

The voltage difference V_a is applied between the accelerating anodes and the cathode. This is normally done by holding the cathode at a high negative voltage (typically -2000 V) and anodes at around 0 V. The electrostatic force exerted on electrons causes them to accelerate and gain kinetic energy. The kinetic energy

$KE = \frac{1}{2}mv_e^2$ .... (C.3)

Where m is the electron mass = $9.1x10^{-31}$kg; e is the electron charge = $1.6x10^{-19}$coulomb; and v_e is the electron velocity in m/sec.

Energy gained by the electron = $e(V_{acc.anode} - V_{cathode}) = eV_a$ and $KE = PE$, yielding

$\frac{1}{2}mv_e^2 = e(V_{acc.anode} - V_{cathode}) = eV_a$ .... (C.4)

$v_e = \sqrt{\frac{2eV_a}{m}}$ m/sec.. (C.5)

Example C.1

Let the cathode voltage $V_{cathode}$= -2000 volt, accelerating anode voltage $V_{acc.anode}$ = 0 volt

Then, $v_e = \sqrt{\frac{2x1.6x10^{-19}x2000}{9.1x10^{-31}}} = \sqrt{\frac{6.4}{9.1}x10^{15}} = 2.65x10^7 m/\sec = 26500 km/\sec$

The velocity is less than one tenth of the speed of the light. Hence, relativistic correction is not necessary.

C.2.3 Deflection of the Beam

Two types of deflection mechanisms are used as the electrostatic and electromagnetic deflections. In the electrostatic deflection, the tube has a long neck and narrow display area. It can be used to display high frequency signals efficiently. Tubes using the electromagnetic deflection have shorter and thinner neck and larger display areas. They are limited to low frequency applications as in medical displays, television and computer monitors.

C.2.3.1 Electrostatic Deflection

Two parallel plates separated by a distance *d* and a deflection voltage V_d is applied in between as shown in Figure C.5. The length of the plate is l_d. Electric field

$$E = \frac{-V_d}{d} \quad \text{(C.6)}$$

is developed where y indicates the direction of the electric field along the y-axis. If an electron appears between the plates, then a force is exerted on it and causes it to accelerate along the y-axis.

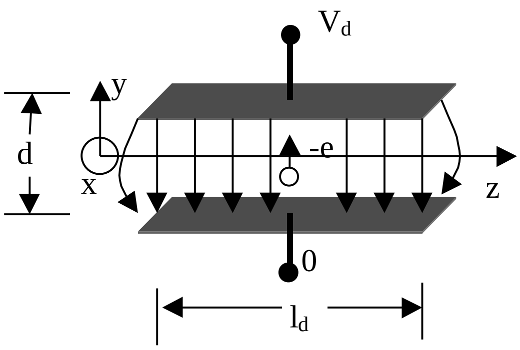


Figure C.5 Electrostatic deflection

$$F = -eE = e\frac{V_d}{d} = m\frac{d^2y}{dt^2} \quad \text{(C.7)}$$

yields, the acceleration of the electron along the y-axis as

$$\frac{d^2y}{dt^2} = \frac{eV_d}{md} \quad \text{(C.8)}$$

The electron enters the deflection plates in z-direction with a velocity of v_e as specified previously. Hence,

$$v_{ez} = \sqrt{\frac{2eV_a}{m}} \quad \text{(C.9)}$$

and the velocity along the y-axis is zero at this moment. The velocity along the z-axis remains constant since there is no force effecting the electron in this direction. However, the electrostatic force along the y-axis causes the electrons to accelerate in y-direction. Distances covered in a given time *t* along both directionsis expressed by a pair of parametric equation:

$$z = v_{ez}t = \sqrt{\frac{2eV_a}{m}}t \quad \text{(C.10)}$$

and

$$y = \frac{1}{2}a_y t^2 = \frac{eV_d}{2md}t^2 \quad \text{(C.11)}$$

Taking *t* from the first one and replacing it in the second one yields:

$$t = \frac{z}{\sqrt{\frac{2eV_a}{m}}} \quad \text{...(C.12)}$$

$$y = \frac{eV_d}{2md}\frac{z^2}{\frac{2eV_a}{m}} \Rightarrow y = \frac{V_d}{4dV_a}z^2 \quad \text{...(C.13)}$$

Hence, the electron travels through a parabolic path in y-z plane while it is inside the deflection plates. As it leaves the plate however, the electrostatic force is not influential. Therefore, the trajectory outside the plates becomes a straight path as indicated in Figure C.6. The straight line drawn is tangent to the parabola at $z = l_d$. The slope of the line

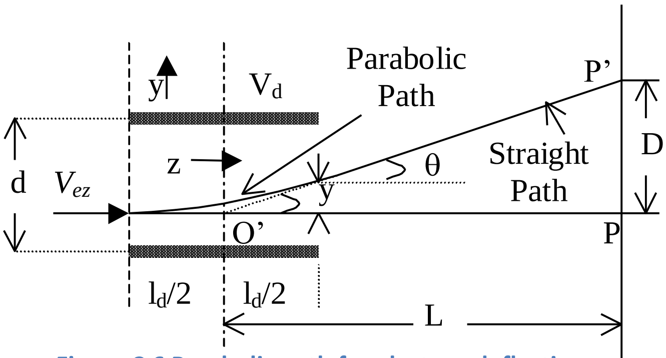


Figure C.6 Parabolic path for electron deflection

$$\tan\theta = \frac{dy}{dz}_{|z=l_d} = \frac{V_d l_d}{2dV_a} \quad \text{(C.14)}$$

This tangent intersects with the z-axis at point O’. The vertical deflection at the point of tangent is

$$y_{|z=l_d} = \frac{V_d l_d^2}{4dV_a} \quad \text{..........................}$$

This is limited to *d/2* since any value of y >d/2 would cause the electron to hit the deflection plate.

The position of the apparent center can be found

$$z - O' = \frac{y}{\tan\theta} = \frac{V_d l_d^2}{4dV_a}\frac{2dV_a}{V_d l_d} = \frac{l_d}{2} \quad \text{...(C.16)}$$

Thus, the apparent origin is at the center (O) of the deflection plates that is L (m) from the screen.

The beam would hit the screen at point P if $V_d = 0$. With $V_d \neq 0$, there will be a deflection on the screen given by

$$D = L\tan\theta = L\frac{V_d l_d}{2dV_a} = \frac{Ll_d}{2dV_a}V_d = SV_d \text{ (m)} \quad \text{...(C.17)}$$

Where *S* is the *deflection sensitivity*;

$$S = \frac{D}{V_d} = \frac{Ll_d}{2dV_a} \text{ (m/V)} \quad \text{...(C.18)}$$

The *deflection factor G* of the CRT is the reciprocal of the sensitivity *S* and is expressed as

$$G = \frac{1}{S} = \frac{2dV_a}{Ll_d} \text{ (V/m)} \quad \text{...(C.19)}$$

High accelerating voltages produce an electron beam with more kinetic energy and eventually a brighter image on the screen. However, this beam is more difficult to deflect and requires higher deflection voltages for a given excursion on the screen. Typical values of deflection factors range from 10 V/cm to 100 V/cm. In CRTs designed for high frequency operation, a post-deflection mechanism is also used after the deflection plate assembly.

Example C.2: Deflection on the Screen

For a cathode ray tube the accelerating voltage is 1,500 V, the length of deflection plates is 2 cm and separation between plates is 1 cm.

1. How much voltage is required across two deflection plates to deflect an electron beam 1° as it leaves the plates?
2. How much the deflection on the screen if it is 12 cm away from the center of the plates?
3. What is the velocity of the electron as it enters the deflection plates?
4. What is the deflection sensitivity? What is the deflection factor?

SOLUTION

For V_a=1500 V, l_d=2 cm and d=1 cm;

a. $\tan\theta = \frac{l_d}{2d}\frac{V_d}{V_a}$ yields $V_d = \frac{2dV_a \tan\theta}{l_d}$ that can be calculated as V_d=26.2 V

b. $\tan\theta = \frac{D}{L}$ yields D=2.1 mm

c. $v = \sqrt{\frac{2eV_a}{m}}$ yields v=23x10^6 m/sec

d. Deflection sensitivity S=D/V_d = 0.08 cm/V and the deflection factor G=1/S = 12.5 V/cm

Example C.3: Length of the CRT

What is the minimum distance , *L*, that will allow full deflection of 4 cm at the oscilloscope screen with a deflection factor of 100 V/cm and with an accelerating voltage of 2000 V?

SOLUTION

The beam can be deflected before it leaves the plates by *d/2* at maximum, otherwise it will hit the plates. Hence, the deflection angle is $\theta = \tan^{-1}(\frac{d}{l})$ at maximum. From the triangular geometry it can be shown that, as we have the maximum possible deflection

$$\frac{D}{d/2} = \frac{L}{l_d/2} \Rightarrow \frac{L}{D} = \frac{l_d}{d}$$

L can also be found from the equation for the deflection factor as $L = \frac{2dV_a}{Gl_d}$

Substituting the previous equation into the later one yields $L^2 = \frac{2DV_a}{G} = \frac{2x4x10^{-2}x2x10^3}{10^4}$ and L = 0.126 m.

Thus, the distance from the center of the deflection plates to the oscilloscope screen is 12.6 cm. If the acceleration potential is increased to 8000 V, then the distance becomes 25.2 cm. Lower deflection factors are desirable since it allows utilization of lower-voltage deflection amplifiers voltages that are easy do design electronically. Lowering the deflection factor would necessitate even longer CRTs. Hence, the CRT becomes the most expansive part of the oscilloscope.

C.2.3.2 Electromagnetic Deflection

Electron beam can be deflected from its path if it is subjected to a magnetic field as well. In this case, the force acting on the electron is perpendicular to both the direction of electron flow and the magnetic field itself.

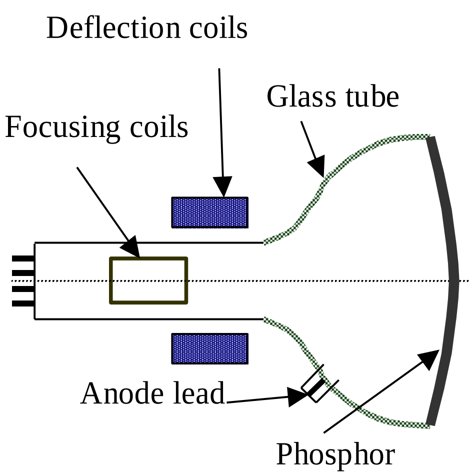


Figure C.7 Magnetic deflection

Two sets of coils are placed perpendicular to each other over the neck of the CRT outside the glass envelope as shown in Figure C.7. The current in these coils provide the two magnetic fields in X and Y directions. As the electron comes in Z direction, it is deflected in Y and X directions respectively. The mechanism of coils is called the *deflection yoke*.

The neck of the CRT is considerably shorter and thinner than the case of electrostatic deflection. There is also no geometric limitation on the deflection angle resulting in larger display area. There are two basic limitations in application of the electromagnetic deflection. Firstly, the inductance and distributed capacitance of the coil require higher voltages to be applied for a given current as the frequency of the deflection current increases. Practical tubes are limited to frequencies up to 20 - 25 kHz. The minimum deflection frequency in the cheapest laboratory oscilloscope is 20 MHz. Eventually, almost all laboratory oscilloscopes utilize electrostatic deflection mechanisms. The second limitation comes from the increased screen size. The trajectory of the spot covers varying lengths as it travels along the screen. This requires a more complicated focusing circuitry.

C.2.4 The Screen

The screen of the oscilloscope is coated with a deposit of phosphor salt and it is semitransparent. Incoming electrons strike on the phosphorous side and transfer their kinetic energies to the coating material. Light is produced as a result and radiated outside through the other (semitransparent) face of this phosphorescent screen. There are important terms related to the screen:

Fluorescence: property of some crystalline materials to emit light when stimulated by radiation.

Phosphorescence: property of material to continue light emission even after the source of excitation is cut-off.

Persistence: the length of duration of phosphorescence.

Luminance: the intensity of light emitted from the CRT screen. The luminance is affected by:

a. Number of bombarding electrons/sec. (the beam current).
b. Spot size.
c. Energy of bombarding electrons (accelerating potential)
d. Sweep speed (how long the electron keeps bombarding a given spot before progressing to the neighboring one).
e. Characteristic of the phosphor itself.

There are several materials available for the phosphorous coating. The efficiency, spectrum of light and persistence vary. The screen for laboratory oscilloscopes is generally a deposit of zinc silicate, which produces a green light output. It has short persistence to ensure that a changing pattern will not appear confused and overlapped. The phosphors used for medical displays have long persistence and gives a yellow-green light.

A graphite (aquadaq) coating covers the inside surface of the glass envelope. It is connected to a high voltage supply and attracts the secondary electrons emitted by the screen. Hence, it is connected to the cathode via the power supply (not to the screen) and completes the electric circuit.

The outside surface of the screen is covered with a transparent plastic or glass that contains calibrated horizontal and vertical lines (graticules) to facilitate the use of the oscilloscope and ease the measurements. In some oscilloscopes, there are ordinary light bulbs fixed between the CRT and outside sheet to illuminate the graticules. This is useful in taking photographs from the screen since it allows the rules to appear on the photograph. Also it helps in reducing the parallax errors in measurements.

C.2.5 CRT Controls

There are three controls on the CRT as:

1. **Intensity**: it adjusts the beam current by varying the potential between the cathode and the control grid.

2. **Focus**: it controls the focal length of the electrostatic lens by varying the voltage difference between the accelerating and focusing anodes.
3. **Astigmatism**: it controls the roundness of the spot by adjusting the potential between the deflection plates and first accelerating anode. This control is not available in some laboratory oscilloscopes.

Some oscilloscopes also have a control called the trace rotation. It is used to obtain a straight line horizontally (with zero slope) when a DC voltage is applied. It controls the current applied to a wire placed inside the CRT and provides a magnetic field.

C.3 IMPORTANT OSCILLOSCOPE CIRCUITS

C.3.1 Signal Input and Vertical Amplifier Assembly

The oscilloscope measures the magnitude of an input signal according to the deflection of the electron it causes on the CRT screen. Schematic diagram of signal input and vertical deflection amplifier assembly is shown in Figure C.8. It contains a connector and switch assembly, an attenuator

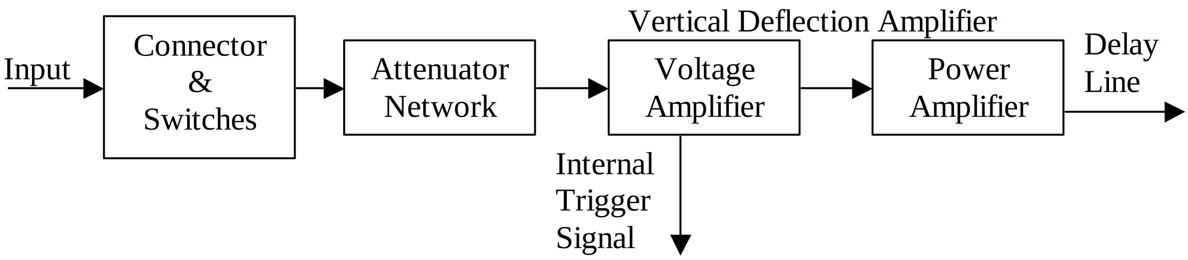


Figure C.8 Schematic diagram of signal input and vertical amplifier assembly

assembly and an amplifier assembly. The amplifier assembly contains two distinct sections as the voltage amplifier to increase the amplitude of the input signal and a power amplifier that provides necessary current to derive the deflection plates. The output of the amplifier assembly is applied to the vertical deflection plates through the delay line. The amplifier stage also presents signal to the triggering circuit.

C.3.1.1 Input connection

The oscilloscope is basically a voltage-measuring instrument. All inputs are applied through a special connector called the BNC (***B**ayonet **N**eill **C**oncelman)* connector as illustrated in Figure C.9. The outside of the connector is attached to the chassis of the oscilloscope. Therefore, one end is

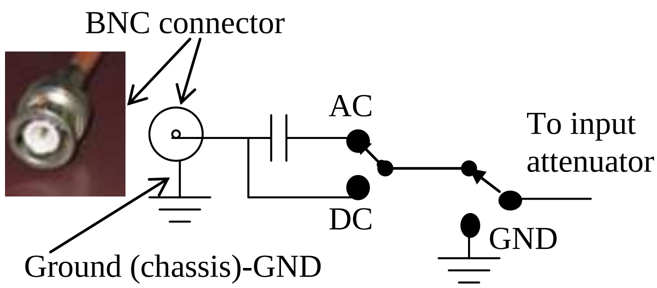


Figure C.9 Input connection to the scope automatically connected to the system ground through the chassis.

The connection of the oscilloscope to electronic circuits is made through a special cable that has an inner conductor wire and outer wire mesh that is called shield. The two conductors are isolated from each other by a plastic coating on the inner one. One end of the cable is fixed to the male of the BNC connector. The shield goes to the outside and the inner conductor is connected to the central terminal of the BNC. The other end of the cable accommodates a crocodile clip fixed to the shield via a wire and a hook attachment at the tip connected to the internal conductor. The whole assembly is called *the probe*. Hence, **the voltage measurement is always made between the point where the tip is connected and the ground**.

Practically all input connections have a provision to select AC or DC coupling via a switch on the front panel of the oscilloscope. In AC coupling, a capacitor is used to block the DC component of the AC signal and let only the AC component applied to the input. This allows measurement of AC voltages in the presence of high DC voltages. In DC coupled mode the connection is made directly hence, whatever comes from the input, AC or DC is applied to the attenuator. A second switch is involved in the signal flow path to interrupt the signal connection and apply “0” volt (ground - GND) to the input to establish the base-line (reference) for measurements.

C.3.1.2 Input Attenuator

The input connector feeds an input attenuator, after which follows the vertical amplifier. The oscilloscope is versatile equipment that is used in a wide amplitude and frequency ranges. Thus, it accepts inputs as low as a few millivolts per centimeter of deflection up to few tens of volts per centimeter. The frequency range goes from a few hertz up to tens of megahertz. The vertical amplifier is designed to have a wide bandwidth (frequency response). The gain of the amplifier is made high enough to drive the vertical deflection assembly to obtain the required deflection on the

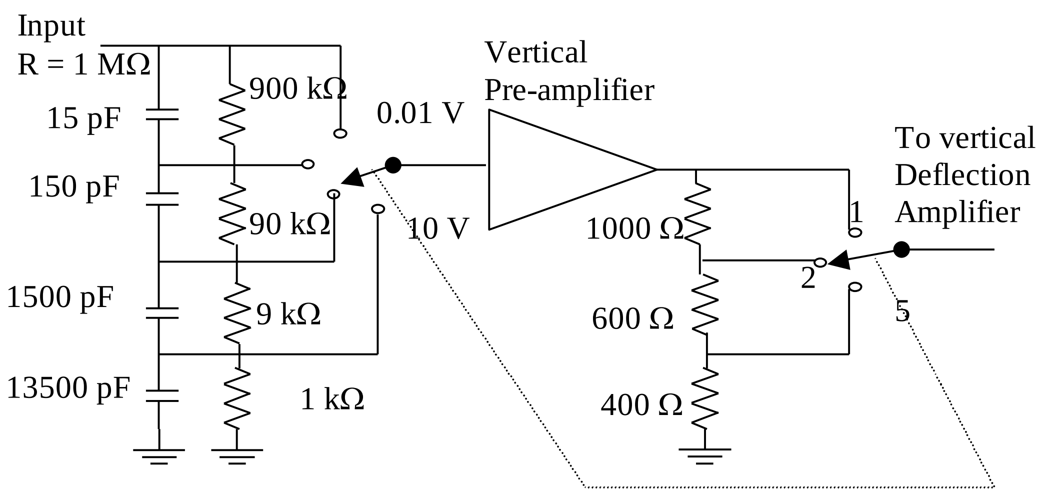


Figure C.10 The two-stage compensated input attenuator screen even in the case of the smallest amplitude at the input. An attenuator assembly is attached to the front end between the amplifier and the input connector to reduce the amplitude of the incoming signals to the desired level. Figure C.10 shows a simplified diagram of a two-stage compensated attenuator. It is composed of two attenuator networks controlled by two switches. The first one selects the decade and the second one selects the correct 1-2-5-sequence attenuation within the decade. It covers at least three decades from 10 mV/cm to 10 V/cm.

The input impedance of the oscilloscope is set to 1 MΩ with the help of the attenuator network shown. The input capacitance is 13.5 nF in the example. Both the input impedance and the attenuation must stay constant over the frequency range for which the oscilloscope was designed. For high frequency operations, this is difficult condition to maintain. Therefore, more sophisticated attenuator networks are designed and also compensating probes are used.

C.3.1.3 High Impedance Probes

The input impedance of an oscilloscope is 1 MΩ shunted with a 10 to 30 pF capacitance. When the probe is connected, the capacitance of the probe assembly is added to the input capacitance. The input to the oscilloscope behaves as a low-pass filter and high frequency signals receive additional attenuation. High impedance probe has 9 MΩ at the tip shunted with a capacitor as illustrated in Figure C.11. In the base of the probe at the oscilloscope connector, there is an adjustable capacitor. A rather flat frequency response within the frequency range of the oscilloscope is achieved by adjusting the capacitor. Overall input to the oscilloscope is reduced 10-to-1 and such a probe is also called X10 probe. The amplitude readings must be multiplied by 10 unless it is done by the oscilloscope itself through special sensing of the presence of the attenuating (X10) probe.

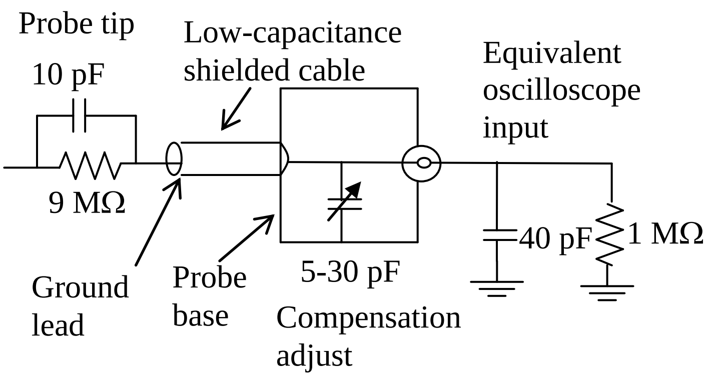


Figure C.11 The high-impedance probe (X10)

C.3.1.4 Vertical Deflection Amplifier

The vertical deflection amplifier is a DC amplifier with a bandwidth that covers the maximum frequency that can be displayed by the oscilloscope. There are two controls on the amplifier. The first one is the gain that can be changed through a potentiometer, generally fixed inside of the attenuator switches. It must be kept at calibrated (CAL) position while measuring the amplitude of the signal. The second one is the vertical position of the display that is adjusted by the position (POS) potentiometer. Adding a DC voltage that can be varied to the input signal does this. The input must

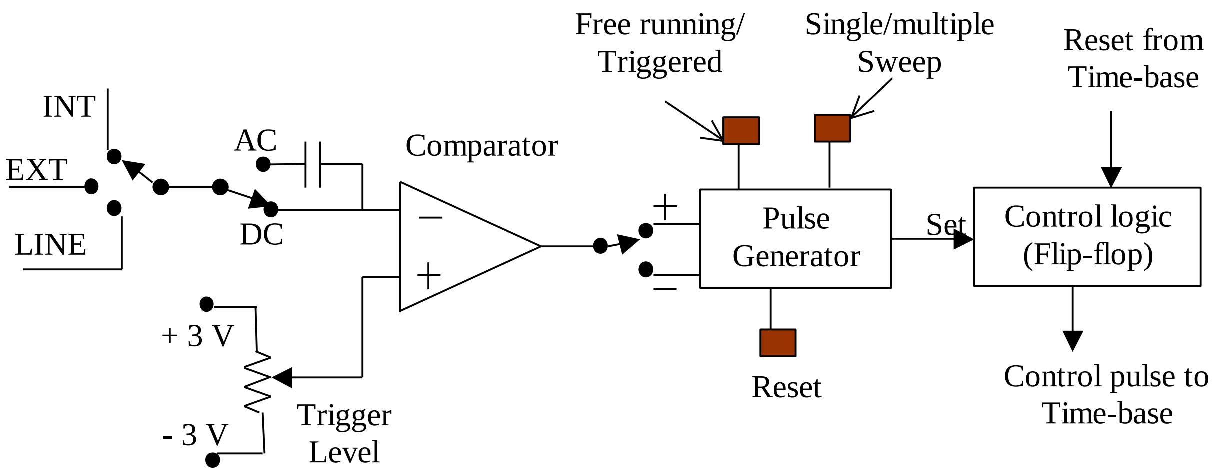


Figure C.12 Schematic diagram of the trigger circuits

be grounded and level must be set to a proper location before any DC voltage measurement.

C.3.2 The Trigger and Time-Base Circuits

C.3.2.1 The Trigger Circuit and Control of The Time-Base Circuit

The trigger circuit is responsible for producing pulses that initiate the ramp signal generated by the time-base circuitry. Figure C.12 shows a symbolic diagram of the organization of the trigger circuits. It accepts input from the vertical deflection amplifier internally (INT), from the step-down transformer of the power supply (LINE) or from an external source. The input is applied to the comparator either directly (DC) or via a capacitor (AC) to remove the DC component of the signal.

The comparator compares the incoming signal with a DC voltage whose value is set by the trigger level potentiometer. For periodic input signals, the output of the comparator is a pulse whose duration depends upon how long the input remains above trigger level as illustrated in Figure C.13. There will be two coincidences between the input signal and the trigger level as indicated in the figure, provided that the trigger level is set within the amplitude range of the signal.

The pulse generator is a multivibrator that produces an output pulse to set the flip-flop. It has several controls that are selected by push-button switches. The pulse can be synchronized with the positive (rising) edge or negative (falling) edge of the input selectable via the slope switch. It can run in free or triggered mode selected by another switch. In a free running mode, it produces an output pulse to allow the time-base to generate the sweep signal even if no input to the oscilloscope is applied or the trigger level is set beyond the signal range so that no coincidence occurs between the inputs to the comparator. The operation may also synchronize with the incoming pulses from the comparator.

There is also a provision to have a single pulse that causes a single sweep pulse to be generated by the time base. This is useful in displaying signals that occur only once and traces can be stored in the storage facility of the oscilloscope or pictures can be taken using photographic cameras. This is strictly a triggered operation and the set pulse is synchronized with the output of the comparator. The reset button initializes the multivibrator for a new operation.

C.3.2.2 The Time-Base Circuit

The time-base circuit is responsible in generating the ramp signal that derives the horizontal deflection plates. It is basically composed of a constant current, a capacitor and an electronic switch in parallel as illustrated in Figure C.14. Upon reception of a proper trigger pulse, the control logic

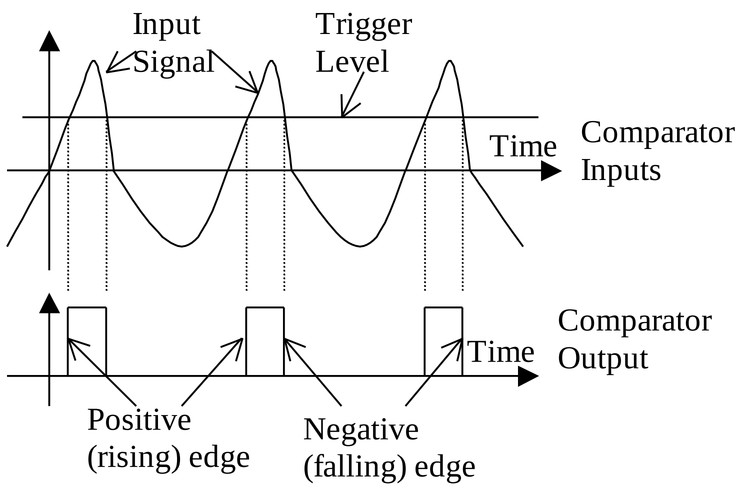


Figure C.13 Inputs and output of the comparator

opens the switch and allows the capacitor to charge. The voltage across the capacitor gradually increases as

$$V_C = \frac{I}{C}t \qquad \text{(C.20)}$$

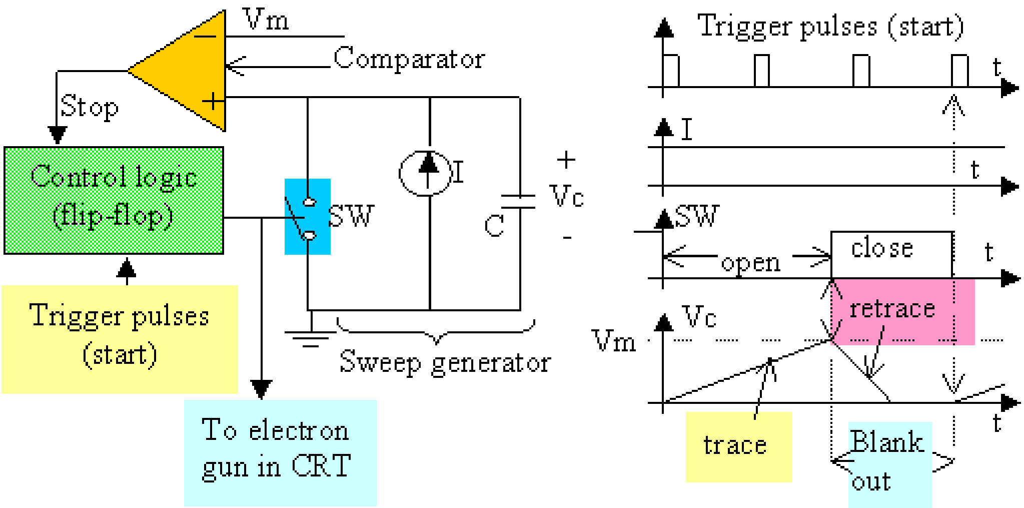


Figure C.14 Schematic and timing diagrams of the sweep generator

where I is the current provided by the source and C is the value of the capacitor in farad. Given I and C constant, V_C increases with a fixed slope and it causes the electron to sweep the screen horizontally at a constant rate. This is called *the trace*.

The electron beam reaches to the right hand side of the screen as the V_C assumes its maximum value. This point is detected via a comparator and the switch is turned "on" by the control logic. The capacitor then discharges through the switch. The switch is kept "on" until reception of a new trigger pulse.

During the discharge of the capacitor, the electron beam returns to the right hand side of the screen. Hence, this is called the retrace. All trigger pulses that appear between the trace and retrace intervals are ignored. The time from the termination of retrace and beginning of the next trace is called *the hold-off time*. The discharge switch is kept "on" during the retrace and hold-off times. During this interval, the screen is blanked out by switching off the electron gun in the cathode ray tube.

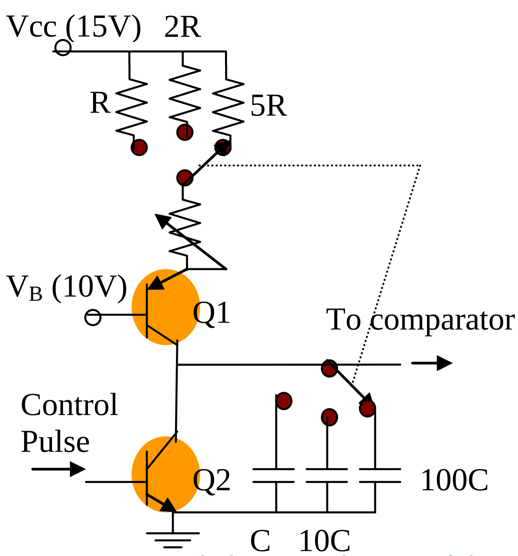


Figure C.15 Symbolic circuit diagram of the sweep generator

The symbolic circuit diagram of the sweep generator is shown in Figure C.15. It has two transistors Q1 and Q2. Q1 works as a constant current source. The current is set according to 1-2-5 sequence by selecting the proper bias resistor. Capacitors are arranged to select the decade. Q2 works as the switch. It saturates with the high level of the control pulse from the trigger circuit. This is equivalent to a close switch (“on” position). With the control pulse having “low” status, Q2 is cut-off indicating an “open” switch (“off” position).

C.4 CATHODE RAY TUBE (CRT) BASED PICTURE DISPLAYS

C.4.1 Principles of Operation

Main elements of a CRT display is illustrated in Figure C.16. Most CRT displays used in computer monitors and television screen use electromagnetic deflection principle discussed in C.2.3.2. The

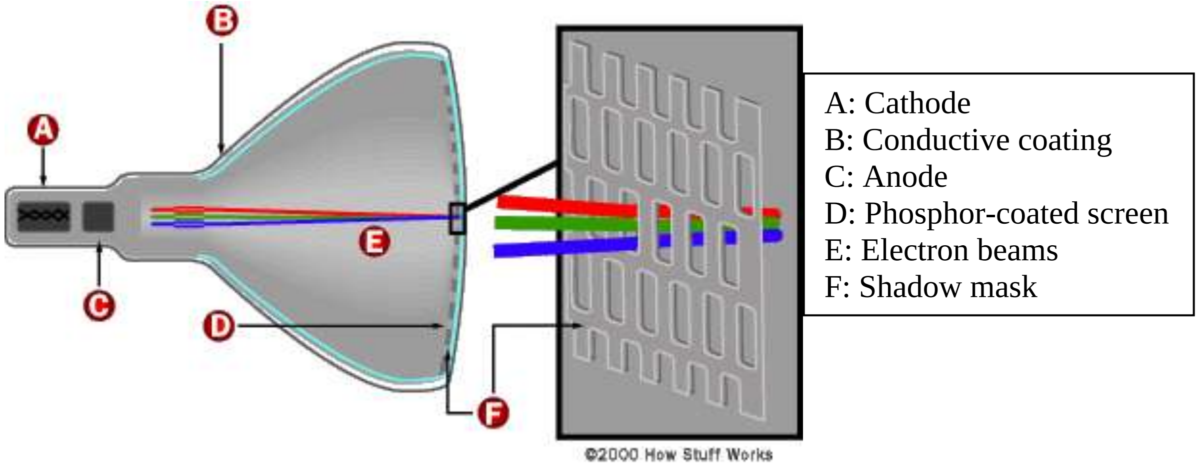


Figure C.16 Main elements of a CRT display (taken from 2000 How Stuff Works; http://electronics.howstuffworks.com/tv1.htm)

screen is not uniformly coated with the phosphorous material; rather it contains a matrix of thousands of tiny phosphor dots. Each dot consists of three blobs of coloured phosphor: one red, one green, one blue. These groups of three phosphors make up what is known as a single pixel. There are three separate electron guns, one for each phosphor color. Images are produced when electrons, fired from the electron guns, converge to strike their respective phosphor blobs. Combinations of different intensities of red green and blue phosphors generate the illusion of millions of colors.

C.4.2 Advantages of CRT Monitors

The CRT technology is a classical and well established one. It is still advantages as compared to other emerging technologies in terms of price, color representation, responsiveness to fast changes and ruggedness as briefed below.

- **Less expensive** - Although LCD monitor prices have decreased, comparable CRT displays still cost less.
- **Better color representation** - CRT displays have historically represented colors and different gradations of color more accurately than LCD displays. However, LCD displays are gaining ground in this area, especially with higher-end models that include color-calibration technology.

- **More responsive** - Historically, CRT monitors have had fewer problems with ghosting and blurring because they redrew the screen image faster than LCD monitors. Again, LCD manufacturers are improving on this with displays that have faster response times than they did in the past.
- **Multiple resolutions** - If you need to change your display's resolution for different applications, you are better off with a CRT monitor because LCD monitors don't handle multiple resolutions as well.
- **More rugged** - Although they are bigger and heavier than LCD displays, CRT displays are also less fragile and harder to damage.

C.5 LIQUID CRYSTAL DISPLAYS[4]

C.5.1 Principles of Operation

There are three common states of matter: solid, liquid or gaseous. **Solids** act the way they do because their molecules always maintain their orientation and stay in the same position with respect to one another. The molecules in **liquids** are just the opposite: They can change their orientation and move anywhere in the liquid. But there are some substances that can exist in an odd state that is sort of like a liquid and sort of like a solid. When they are in this state, their molecules tend to maintain their orientation, like the molecules in a solid, but also move around to different positions, like the molecules in a liquid. This means that liquid crystals are neither a solid nor a liquid. That's how they ended up with their seemingly contradictory name.

One feature of liquid crystals is that they're affected by **electric current**. A particular sort of nematic liquid crystal, called **twisted nematics** (TN), is naturally twisted. Applying an electric current to these liquid crystals will untwist them to varying degrees, depending on the current's voltage. LCDs use these liquid crystals because they react predictably to electric current in such a way as to control light passage.

C.5.1.1The Basics of LCD

Liquid crystal display technology works by blocking light. Specifically, an LCD is made of two pieces of polarized glass (also called substrate) that contain a liquid crystal material between them. A backlight produces light that passes through the first substrate. At the same time, electrical currents cause the liquid crystal molecules to align to allow varying levels of light to pass through to the second substrate and generate the colors and images that we see.

---

[4] Extracted from Jeff Tyson, How LCDs Work, http://electronics.howstuffworks.com (visited in June 2007)

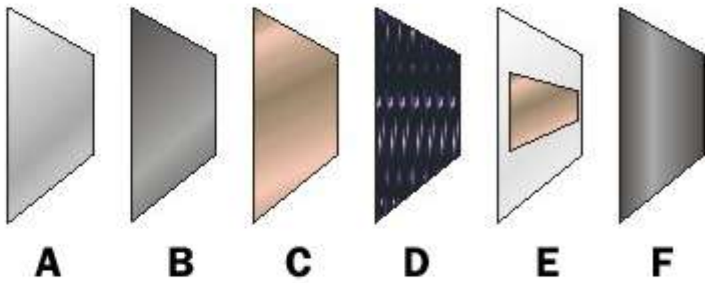


Figure C.17 Layer of the liquid crystal display

The LCD needed to do this job is very basic and it has six layers as illustrated in Figure C.17.

- It has a mirror (**A**) in back, which makes it reflective.
- Then, we add a piece of glass (**B**) with a polarizing film on the bottom side,
- And a common electrode plane (**C**) made of indium-tin oxide on top. A common electrode plane covers the entire area of the LCD.
- Above that is the layer of liquid crystal substance (**D**).
- Next comes another piece of glass (**E**) with an electrode in the shape of the rectangle on the bottom and,
- On top, another polarizing film (**F**), at a right angle to the first one.

The electrode is hooked up to a power source like a battery. When there is no current, light entering through the front of the LCD will simply hit the mirror and bounce right back out. But when the battery supplies current to the electrodes, the liquid crystals between the common-plane electrode and the electrode shaped like a rectangle untwist and block the light in that region from passing through. That makes the LCD show the rectangle as a black area.

C.5.1.2 Reflective LCD (liquid crystal display)

Figure C.18 An LCD type calculator display

An LCD type calculator display is shown in Figure C.18. The LCD requires an **external light source**. Liquid crystal materials emit no light of their own. Small and inexpensive LCDs are often **reflective**, which means to display anything they must reflect light from external light sources. Look at an LCD calculator: The numbers appear where small electrodes charge the liquid crystals and make the layers untwist so that light is not transmitting through the polarized film.

C.5.1.3 Backlit LCD

Most computer displays are lit with built-in **fluorescent tubes** above, beside and sometimes behind the LCD (Figure C.19). A white diffusion panel behind the LCD redirects and scatters the light evenly to ensure a uniform display. On its way through filters, liquid crystal layers and electrode layers, a lot of this light is lost - often more than half! A Cold Cathode Fluorescent Lamp (CCFL) is used.

Figure C.19 A backlit type LCD monitor with a cold cathode fluorescent lamp

C.5.2 Display Types

C.5.2.1 Passive and Active Matrix

Passive matrix LCDs use a simple grid to supply the charge to a particular pixel on the display. It starts with two glass layers called **substrates**. One substrate is given columns and the other is given rows made from a transparent conductive material. This is usually **indium-tin oxide**. The rows or columns are connected to **integrated circuits** that control when a charge is sent down a particular column or row. The liquid crystal material is sandwiched between the two glass substrates, and a polarizing film is added to the outer side of each substrate. To turn on a pixel, the integrated circuit sends a charge down the correct column of one substrate and a ground activated on the correct row of the other. The row and column **intersect** at the designated pixel, and that delivers the voltage to untwist the liquid crystals at that pixel.

Active-matrix LCDs depend on **thin film transistors** (TFT). Basically, TFTs are tiny switching transistors and capacitors. They are arranged in a matrix on a glass substrate. To address a particular pixel, the proper row is switched on, and then a charge is sent down the correct column. Since all of the other rows that the column intersects are turned off, only the capacitor at the designated pixel receives a charge. The capacitor is able to hold the charge until the next refresh cycle. And if we carefully control the amount of voltage supplied to a crystal, we can make it untwist only enough to allow some light through. By doing this in very exact, very small increments, LCDs can create a **gray scale**. Most displays today offer 256 levels of brightness per pixel.

C.5.2.2 The Color

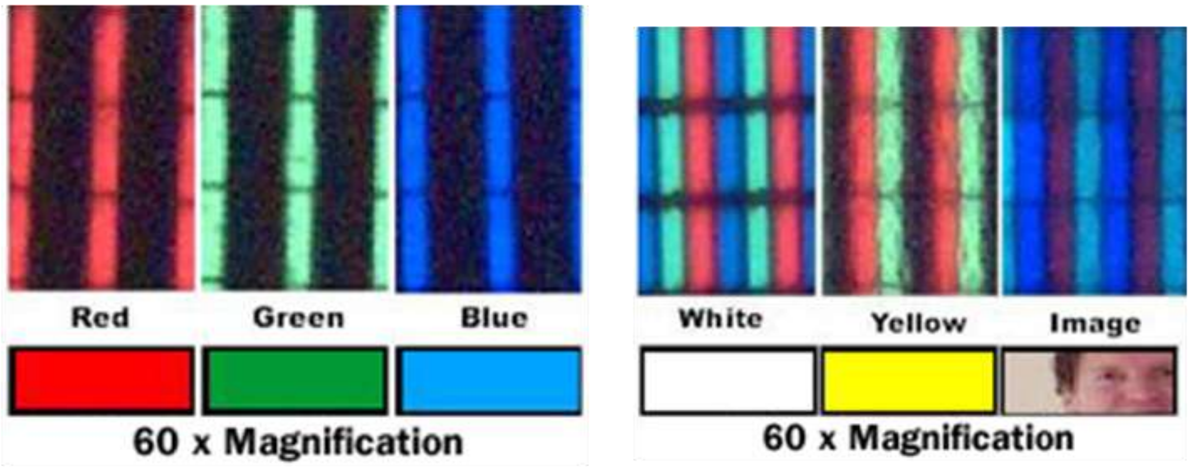


Figure C.20 Subpixels and color filters in LCD

An LCD that can show colors must have **three subpixels** with red, green and blue color filters to create each color pixel as illustrated in Figure C.20. Through the careful control and variation of the voltage applied, the intensity of each subpixel can range over **256 shades**. Combining the subpixels produces a possible palette of **16.8 million colors** (256 shades of red x 256 shades of green x 256 shades of blue). These color displays take an enormous number of transistors. For example, a typical laptop computer supports resolutions up to 1,024x768. If we multiply 1,024 columns by 768 rows by 3 subpixels, we get 2,359,296 transistors etched onto the glass! If there is a problem with any of these transistors, it creates a "bad pixel" on the display. Most active matrix displays have a few bad pixels scattered across the screen.

C.5.2.3 Popular Screen Sizes

Popular screen sizes are 15, 17, 19 and 21 inches. Notebook screen sizes are smaller, typically ranging from 12 to 17 inches. As technologies improve in both desktop and notebook displays, even larger screen sizes are becoming available. For professional applications, such as medical imaging or public information displays, some LCD monitors are 40 inches or larger! Obviously, the size of the display directly affects resolution. The same pixel resolution is sharper on a smaller monitor and fuzzier on a larger monitor because the same number of pixels is spread out over a larger number of inches. An image on a 21-inch monitor with an 800x600 resolution will not appear nearly as sharp as it would on a 15-inch display at 800x600.

C. 5.3 Advantages of LCD Monitors

The LCDs are used as alternative to CRT screens in monitors and text display applications due to their power meagerness, lightness in weight and adaptability into specific applications as briefed below. Yet, they have lagged behind plasma displays in size because they are harder to make. An LCD's polarised light is highly directional, making it harder to view from the side than a cathode-ray tube (CRT) or plasma display. And the speed at which picture frames are refreshed is slower than a plasma display, causing blurring in some fast action scenes.

- **Require less power** - Power consumption varies greatly with different technologies. CRT displays are somewhat power-hungry, at about 100 watts for a typical 19-inch display. The average is about 45 watts for a 19-inch LCD display. LCDs also produce less heat.
- **Smaller and weigh less** - An LCD monitor is significantly thinner and lighter than a CRT monitor, typically weighing less than half as much. In addition, you can mount an LCD on an arm or a wall, which also takes up less desktop space.
- **More adjustable** - LCD displays are much more adjustable than CRT displays. With LCDs, you can adjust the tilt, height, swivel, and orientation from horizontal to vertical mode. As noted previously, you can also mount them on the wall or on an arm.
- **Less eye strain** - Because LCD displays turn each pixel off individually, they do not produce a flicker like CRT displays do. In addition, LCD displays do a better job of displaying text compared with CRT displays.

C.6 PAINTING THE PICTURE

C.6.1 The Raster Scan

The screen is coated with phosphor and the electron beam "paints" an image onto the screen by moving the electron beam across the phosphor a line at a time. To "paint" the entire screen, electronic circuits inside the monitor use the magnetic coils shown in Figure C.7 to move the electron beam in a "**raster scan**" pattern across and down the screen. The beam paints one line across the screen from left to right. It then quickly flies back to the left side, moves down slightly and paints another horizontal line, and so on down the screen, like the one shown in Figure C.21. In this figure, the continuous lines represent lines that the electron beam is "painting" on the screen from left to right, while the dashed lines represent the beam flying back to the left. When the beam reaches the right side of the bottom line, it has to move back to the upper left corner of the screen, as represented by the thick line in the figure. When the beam is "painting," it is

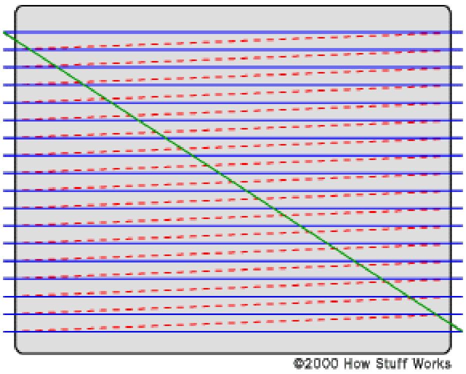


Figure C.21 Painting the screen by a raster scan on, and when it is flying back, it is off so that it does not leave a trail on the screen. The term **horizontal retrace** is used to refer to the beam moving back to the left at the end of each line, while the term **vertical retrace** refers to its movement from bottom to top. As the beam paints each line from left to right, the intensity of the beam is changed to create different shades of the colors across the screen. Because the lines are spaced very closely together, your brain integrates them into a single image.

C.6.2 Refresh Rate

In monitors based on CRT technology, the refresh rate is the number of times that the image on the display is drawn each second. If your CRT monitor has a refresh rate of 72 Hertz (Hz), then it cycles through all the pixels from top to bottom 72 times a second. Refresh rates are very important because they control flicker, and you want the refresh rate as high as possible. Too few cycles per second and you will notice a **flickering**, which can lead to headaches and eye strain.

Because your monitor's refresh rate depends on the number of rows it has to scan, it limits the maximum possible resolution. Most monitors support multiple refresh rates. Keep in mind that there is a tradeoff between flicker and resolution, and then pick what works best for you. This is especially important with larger monitors where flicker is more noticeable. Recommendations for refresh rate and resolution include 1280x1024 at 85 Hertz or 1600x1200 at 75 Hertz.

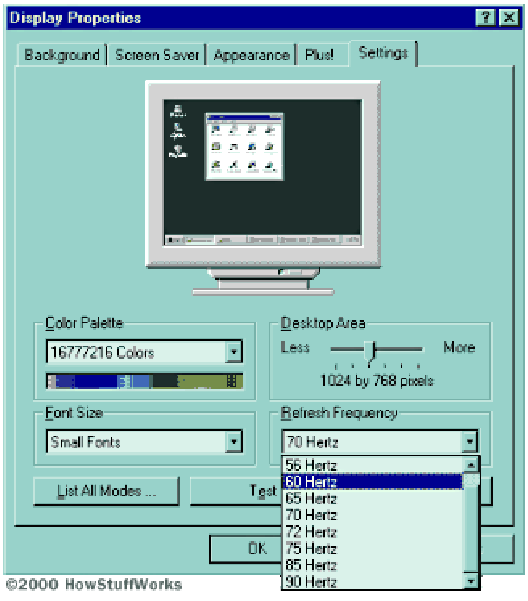


Figure C.22 Selection of the refresh rate

A CRT uses electron beams to create images on a phosphor screen, it supports the resolution that matches its physical dot (pixel) size as well as several lesser resolutions. For example, a display with a physical grid of 1280 rows by 1024 columns can obviously support a maximum resolution of 1280x1024 pixels (Figure C.22). It also supports lower resolutions such as 1024x768, 800x600, and 640x480. An LCD monitor works well only at its native resolution.

C.6.3 Aspect Ratio and Viewable Area

Two measures describe the size of your display: the **aspect ratio** and the **screen size**. Historically, computer displays, like most televisions, have had an aspect ratio of 4:3. This means that the ratio of the width of the display screen to the height is 4 to 3. For widescreen LCD monitors, the aspect ratio is 16:9 (or sometimes 16:10 or 15:9). Widescreen LCD displays are useful for viewing DVD movies in widescreen format, playing games and displaying multiple windows side by side. High definition television (HDTV) also uses a widescreen aspect ratio.

Screen sizes are normally measured in inches from one corner to the corner diagonally across from it. This diagonal measuring system actually came about because the early television manufacturers wanted to make the screen size of their TVs sound more impressive. Interestingly, the way in which the screen size is measured for CRT and LCD monitors is different. For CRT monitors, screen size is measured diagonally from outside edges of the display casing. In other words, the exterior casing is included in the measurement as illustrated in Figure C.23. For LCD monitors, screen

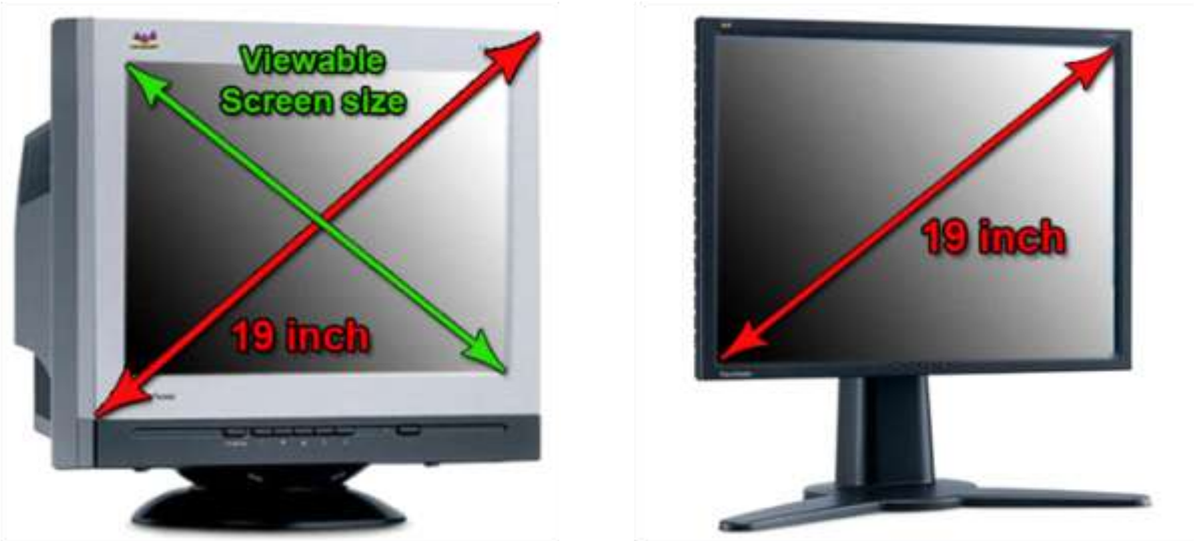


Figure C. 23 Aspect ratios and screen sizes for CRT and LCD monitors

size is measured diagonally from the inside of the beveled edge. The measurement does not include the casing as indicated in the image in Figure C.23. Because of the differences in how CRT and LCD monitors are measured, a 17-inch LCD display is comparable to a 19-inch CRT display. For a more accurate representation of a CRT's size, find out its viewable screen size. This is the measurement of a CRT display without its outside casing.

C.7 EMERGING DISPLAY TECHNOLOGIES

There are emerging technologies as well for the screen in addition to the classical CRT and LCD displays. Important ones among them are the plasma displays, Organic Light-Emitting Diode (OLED) and Surface-Conduction Electron Emitter Displays (SED). Each type will be briefed below and details will be left to the reader who may refer to the references for further information.

C.7.1 Plasma Panel Displays

Under normal conditions, the individual gas atoms include equal numbers of protons and electrons that makes the net charge of the atom zero. If we introduce many free electrons into the gas by establishing an electrical voltage across it, the situation changes very quickly. The free electrons collide with the atoms, knocking loose other electrons. With a missing electron, an atom loses its balance. It has a net positive charge, making it an ion. Plasma is the central element in a fluorescent light. It is generated in a gas made up of free-flowing ions and electrons.

In a plasma with an electrical current running through it, negatively charged particles are rushing toward the positively charged area of the plasma, and positively charged particles are rushing toward the negatively charged area. In this mad rush, particles are constantly bumping into each other. These collisions excite the gas atoms in the plasma, causing them to release photons of energy. Xenon and neon atoms, the atoms used in plasma screens, release light photons when they are excited. Mostly, these atoms release ultraviolet light photons, which are invisible to the human eye. But ultraviolet photons have higher energy than the visible light photons and they can be used to excite visible light photons.

Figure C.24 A plasma display (courtesy of Sony)

A plasma panel display is made up of millions of phosphor-coated gas-filled pixel cells. Each pixel is made up of three fluorescent lights: a red light, a green light and a blue light. Just like a CRT television, the plasma display varies the intensities of the different lights to produce a full range of colors. When excited by a voltage, the gas emits UV light that makes the cells red, green or blue phosphor coating emit visible light.

By varying the pulses of current flowing through the different cells, the control system can increase or decrease the intensity of each subpixel color to create hundreds of different combinations of red, green and blue. In this way, the control system can produce colors across the entire spectrum.

Plasma displays have wide screens, comparable to the largest CRT sets, but they are only about 6 inches (15 cm) thick as illustrated in Figure C.24. Having each pixel lit individually makes the image very bright and looks good from almost every angle. The biggest drawback of this technology has been the price. However, falling prices and advances in technology mean that the plasma display may soon replace the old CRT sets. Proponents say that the plasma technology produces more natural colors and a softer picture than the stark brightness of a uniformly backlit LCD making viewing easier for tired eyes. However, PDP screens have a shorter lifetime than an LCD and consume more power.

C.7.2 Organic Light-Emitting Diode (OLED)

An organic light-emitting diode (OLED) is any light-emitting diode (LED) whose emissive electroluminescent layer comprises a film of organic compounds. The layer usually contains a polymer substance that allows suitable organic compounds to be deposited. They are deposited in rows and columns onto a flat carrier by a simple "printing" process. The resulting matrix of pixels can emit light of different colors.

OLEDs consist of stacks of organic layers (thickness about 100 nm), which are inserted between a cathode and an anode as illustrated in Figure C.25. Usually, the substrate is glass coated with a transparent conductive oxide being the anode, followed by the organic stack, consisting of hole transport and electron transport materials, followed by the inorganic cathode. Key advantages of the organic luminescence are the chemical variability of the organic light-emitting diodes, allowing virtually any color including white, and the thin film system, allowing large-area and low-cost deposition, and the possibility to use thin and even flexible substrates to realize a novel class of lighting and display solutions not possible for other technologies.

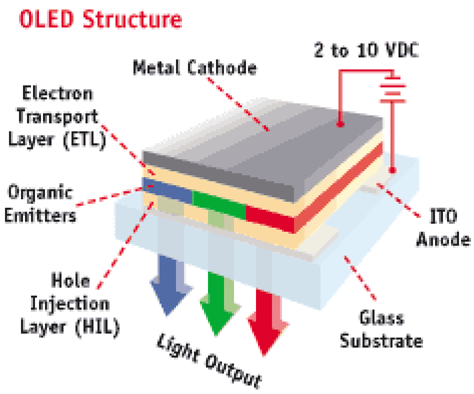


Figure C.25 Structure of an OLED

OLEDs are thin-film LED (Light-Emitting Diode) displays that don't require a backlight to function. Thus they draw far less power and, when powered from a battery, can operate longer on the same charge. The material emits light when stimulated by an electrical current, which is known as **electroluminescence** as mentioned in the introduction.They consist of red, green and blue elements, which combine to create the desired colors. Such systems can be used in television screens, computer displays, portable system screens, advertising, information and indication. OLEDs can also be used in light sources for general space illumination, and large-area light-emitting elements. OLEDs typically emit less light per area than inorganic solid-state based LEDs which are usually designed for use as point-light sources.

Advantages of OLEDs include lower power requirements, a less-expensive manufacturing process, improvements in contrast and color, and the ability to bend. OLED-based display devices also can be more effectively manufactured than LCDs and plasma displays. Yet, electroluminence has not reached the wide audiences, it is generally used on only special applications. Degradation of OLED materials has limited the use of these materials. Displays based on electroluminence traditionally have problems in getting the full color spectrum (problems especially on blue color generation), and have thus been useful only on applications that need only few colors.

C.7.3 Surface-Conduction Electron Emitter (SED) and Field Emission (FED) Displays

The Surface-Conduction Electron Emitter (SED) and Field Emission (FED) Displays are new technologies originated jointly by Canon and Toshiba as flat panel electronic visual displays. They are based on the cathodoluminescence principle and they can be recognized as millions of miniature CRTs filling up the screen. Hence, similar to a CRT, an SED and FED display utilizes electrons and a phosphor-coated screen to create images. The difference is that instead of a deep tube with an electron gun, these displays use tiny electron emitters and a flat-panel display.

SEDs use nanoscopic-scale electron emitters to energize colored phosphors and produce an image. In a general sense, a SED consists of a matrix of tiny cathode ray tubes, each "tube" forming a single sub-pixel on the screen, grouped in threes to form red-green-blue (RGB) pixels. The difference is that instead of a deep tube with an electron gun, an SED uses tiny electron emitters and a flat-panel display as illustrated in Figure C.26.

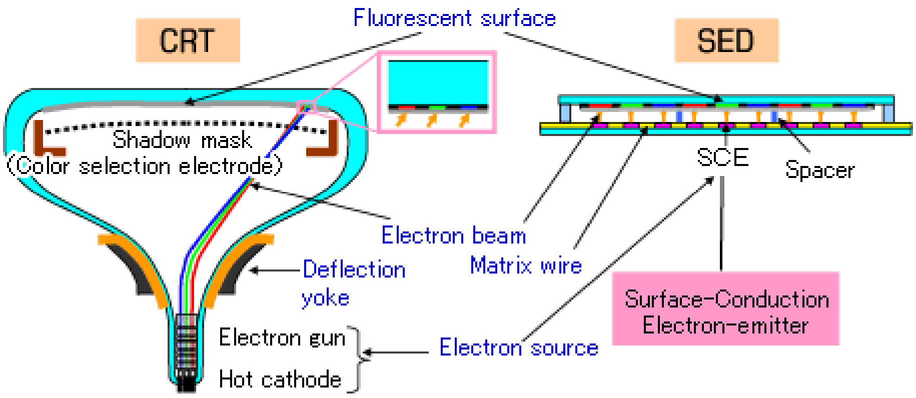


Figure C.26 Structural comparison between CRT and SED displays

After considerable time and effort in the early and mid-2000s, SED efforts started winding down in 2009 as LCD became the dominant technology. However, in August 2010, Canon announced they were shutting down their joint effort to develop SEDs commercially, signaling the end of development efforts. SEDs are closely related to another developing display technology, the field emission display, or FED, differing primarily in the details of the electron emitters. Sony, the main backer of FED, has similarly backed off from their development efforts. In a general sense, a FED consists of a matrix of cathode ray tubes, each tube producing a single sub-pixel, grouped in threes to form red-green-blue (RGB) pixels as illustrated in Figure C.27.

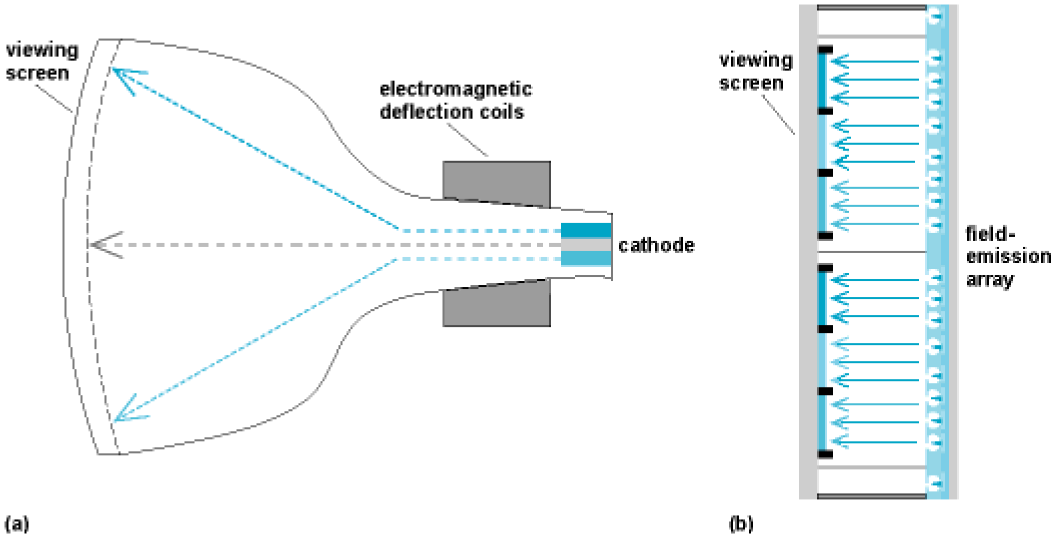


Figure C.27 Comparison between CRT (a) and FED (b) displays

SEDs and FEDs combine the advantages of CRTs, namely their high contrast ratios, wide viewing angles and very fast response times, with the packaging advantages of LCD and other flat panel displays. They also use much less power than an LCD television of the same size (for an FED, it is about half of an LCD system).

The technologies described above have all native resolution defined by the pixel count of the display. All flat panel displays (FPDs) operate best at their native resolution. To display a non-native resolution, the panel manufacturers interpolate the incoming signal and usually have different resolutions they support. However, not all supported resolution always give a good picture, but they give some kind of picture.

C.8 TOUCH SCREEN MONITORS

C.8.1 Touch Screens

Displays with touch-screen technology let you input information or navigate applications by touching the surface of the display as illustrated in Figure C.28. A touchscreen is any monitor, based either on LCD (Liquid Crystal Display) or CRT (Cathode Ray Tube) technology, that accepts direct onscreen input. The ability for direct onscreen input is facilitated by an external (for example pen) or an internal device (touch overlay and controller) that relays the X-Y coordinates of point touched to the computer. Touchscreen technology gives us the power to make our computer react without using a mouse or keyboard. We just press what we see on the screen. Touchscreens are also ideal for unattended public applications in high traffic environments. They are extremely user-friendly and durable.

Figure C.28 A touch screen display

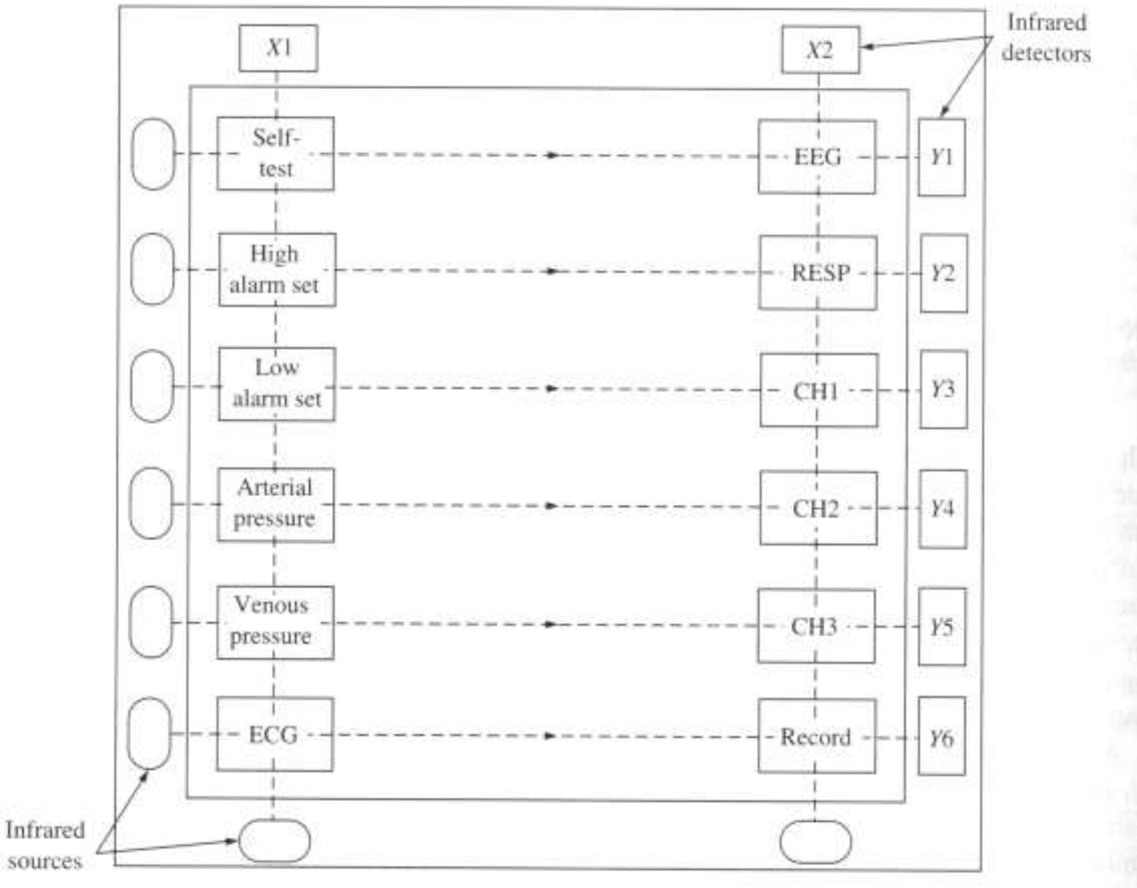
C.8.2 Touch Screen Technologies

Touchscreen monitors make use of a range of technologies to detect touch, including capacitive-sensing, sound and light sensors, and pressure on the screen surface. Capacitive touchscreens sense electrical signals to determine the presence and location of our finger as it makes contact with the surface of the touchscreen. Strengths of capacitive technology include a fast response time, durability and a tolerance for surface contamination. Quantum Tunneling Composite (QTC) is a new class of electrically conductive material that has been developed to advance the capability of switching and sensing systems. QTC is a pressure switching and sensing material technology and it will be briefly explained later in relation to mechanical pressure sensors.

Resistive screens use a flexible membrane with a coating of transparent metal oxide and a grid of spacers to locate the touchpoint. Resistive LCD touchscreen monitors rely on a touch overlay, which is composed of a flexible top layer and a rigid bottom layer separated by insulating dots, attached to a touchscreen controller. The inside surface of each of the two layers is coated with a transparent metal oxide coating (ITO) that facilitates a gradient across each layer when voltage is applied. Pressing the flexible top sheet creates electrical contact between the resistive layers, producing a switch closing in the circuit. The control electronics alternate voltage between the layers and pass the resulting X and Y touch coordinates to the touchscreen controller. Resistive touchscreen technology exists in 4-wire, 5-wire, or 8-wire forms. 4-wire resistive technology is restricted to small flatpanels (less than 10 inches). Because of its versatility and cost-effectiveness, resistive touchscreen technology is the touch technology of choice for many markets and applications (like retail point-of-sale (POS), medical monitoring devices, industrial process control, handhelds). The downside of resistive technology is that metal oxide coating and spacers may reduce the picture quality and brightness.

Infrared screens generate a grid of light across the face of the screen and check for interruptions to that grid as illustrated in Figure C.29. Surface acoustic wave (SAW) touchscreens send sound waves across our screen surface to look for interruptions caused by touch. Guided acoustic wave runs on principles similar to SAW but sends waves through the screen substrate rather than over the surface. The following are key factors in assessing touchscreen performance:

- ....
  Response time (usually between 8ms and 20ms, more than 25ms may create problems for



  users),
- .... to
  uch contact requirement (measured in milliseconds) and

- .... to
  uch resolution (points per inch).

Figure C.29 An infrared-based touchscreen display

C.8.3 Wireless Monitors

Similar in looks to a tablet PC, wireless monitors use technology such as 802.11b/g to connect to your computer without a cable. Most include buttons and controls for mousing and web surfing, and some also include keyboards. The displays are battery-powered and relatively lightweight. Most also include touch-screen capabilities (Figure C.30).

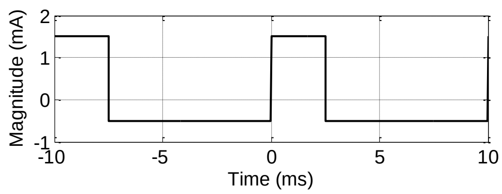
Figure C.30 A wireless patient monitor

D – PRETEST

Knowledge checkout in fundamental topics

Mark the correct choice in the following questions (time allowed 20 minutes):

1. A shorted resistor always has
   a. Infinite current through it
   b. Infinite voltage across it
   c. Zero voltage across it
   d. Zero current through it
2. A four-band resistor with color code <u>orange-white-brown-red</u> is
   a. 390 Ω ± 5%
   b. 39 Ω ± 2%
   c. 39 Ω ± 5%
   d. 390 Ω ± 2%
3. Thevenin's theorem replaces a complicated circuit facing a load by an
   a. Ideal voltage source and parallel resistor
   b. Ideal current source and parallel resistor
   c. Ideal voltage source and series resistor
   d. Ideal current source and series resistor
4. The voltage and current into a network are measured to be 10Vcos(377t) and 1mAcos(377t+60°) respectively. The input impedance of the network is
   a. 10kΩ + j10kΩ
   b. 10kΩ - j10kΩ
   c. 5.0kΩ + j8.6kΩ
   d. 5.0kΩ - j8.6kΩ



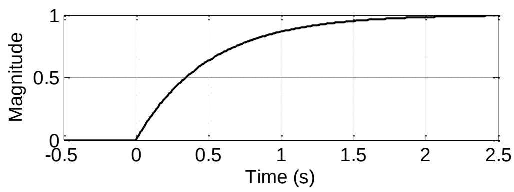
5. The waveform represents a current in 5 kΩ resistor. The power dissipated by the resistor is (time in millisecond and current in milliampere)
   a. 2.25 mW b. 3.7 mW c. 0.01 W d. 0.45 W

6. The fundamental frequency of the signal in the above question is
   a. 0.1 Hz b. 400 Hz c. 100 Hz d. Undefined

7. Which one of the following sinusoidal signals has a period of 1 msec?
   a. 2sin(1000πt) b. cos(2000πt-45°) c. 3sin(2πt+53°) d. 1000sin(1000πt-53°)

8. The voltage and current into a network are measured to be 10Vcos(100πt) and 1mAcos(100πt +60°) respectively. Input impedance of the network is
   a. 10kΩ + j10kΩ b. 10kΩ - j10kΩ c. 5.0kΩ + j8.6kΩ d. 5.0kΩ - j8.6kΩ

9. The figure can be represented by
   a. x(t) = (1 – e-2t)u(t)
   b. x(t) = 1 - e-tu(t)
   c. x(t) = (1 - e-t/2)u(t)
   d. $x(t) = e^{-t}u(t)$



10. A series RC is designed with R = 1 kΩ and C = 1μF. The impedance seen by the source at f = 159 Hz (1000 rad/sec) is
    a. 1kΩ - j1.44kΩ b. 1kΩ - j1kΩ c. 1.44kΩ + j1kΩ d. 1kΩ + j1kΩ

E – EXIT SURVEY

King Abdulaziz University
Faculty of Engineering
Department of Electrical and Computer Engineering
EE 306 – ELECTRICAL ENGINEERING TECHNOLOGIES
EXIT SURVEY

May 2011

Please mark the appropriate boxes in the following table for your current GPA and grade you expect from the course.

GPA Range		☐ <2	☐ 2-2.5	☐ 2.5-3	☐ 3-3.5	☐ 3.5-4	☐ 4-4.5	☐ 4.5-4.75	☐ 4.75-5
Expected Grade	☐ F	☐ D	☐ D+	☐ C	☐ C+	☐ B	☐ B+	☐ A	☐ A+

Please fill in the tables concerning the skills, abilities and attributes that you have acquired, teaching methods and assessment tools used and quality of teaching while studying EE 306 as well as your perception of contribution of the course to your career.

1. Assessment of Abilities, Skills and Attributes Acquired at EE 306.

Please rate how well you have been prepared in each of the following skills, abilities or attributes as stated in the Course Learning Outcomes (CLOs).

Skills, abilities, and attributes	*Level of preparation*			
At finishing of the course EE 306, I am able to: (3 = High, 2 = Average, 1 = Low, 0 = Not Applicable)	3	2	1	0
1. Recognize the commonly used electrical engineering components and choose the proper ones for specific applications				
2. Compare and contrast the electrical energy sources,				
3. Determine the energy requirement of an application				
4. Select protection schemes and devices for safe operations of electrically operated devices				
5. Recognize basic operations and limitations of devices and facilities that use electrical energy				
6. Describe the instrument functions and define terms related to electrical measurements				
7. Illustrate the error sources in measurements and apply statistical analysis of errors				
8. Identify the critical issues for sensor choice, placement, and circuit implementation				
9. Appreciate the applications and limitations of various electronic/electrical measuring instruments				
10. Describe the instrument functions and define terms related to electrical measurements				
11. Determine the energy requirement of an application				

2. Assessment of Educational Methods and Assessment Tools

Please indicate your satisfaction with each the following methods and tools used in the course.

Educational and Assessment Methods Used	*Level of satisfaction*			
(3 = High, 2 = Average, 1 = Low, 0 = Not Applicable)	3	2	1	0
Educational Methods				
1. Classroom lectures				
2. Lab Demonstrations				
3. Lab work				
4. Lab project(s)				
Assessment Methods				
1. Quizzes				
2. Homework				
3. Major exams				
4. Lab project report				
5. Short lab reports				

3. Assessment of Quality of Teaching and Teaching Tools

Please indicate your satisfaction with each the following methods and tools used in the course.

Quality of Teaching and Teaching Tools	*Level of satisfaction*			
(3 = High, 2 = Average, 1 = Low, 0 = Not Applicable)	3	2	1	0
1. Instructor, his way of lecturing				
2. Instructor, his attitudes and interests in teaching				
3. Textbook and lecture notes				
4. General lab facilities				
5. Experiments to illustrate the principles				
6. Lab project				
7. Lab engineer, his attitude and interests				

4. Assessment of Contribution of the Course to Your Electrical Engineering Career

Please indicate your satisfaction with each the following professional components developed in the course.

Professional contribution	*Level of satisfaction*			
After completing the course, I can (3 = High, 2 = Average, 1 = Low, 0 = Not Applicable)	3	2	1	0
1. Design and conduct experiments				
2. Collect experimental data and use statistical analysis				
3. Accept responsibilities as a team member, share information and provide assistance to others				
4. Cooperate with others in obtaining knowledge of technical skills, issues and approaches relevant to disciplines outside of electrical engineering				
5. Determine statistical measures such as accuracy, precision, resolution etc for a measuring equipment and uses them				
6. Correctly infers tolerances of electronic components in design of medical devices				
7. Recognize the statistical variability of electrical components and systems and value the population statistics and calculate important measures such as the mean and standard deviation from a normal distribution of data.				
8. obtain mathematical models, translate academic theory into engineering applications and accept limitations of mathematical models of physical reality.				
9. Develops correct models for electrical engineering problems using electrical circuit analogies, explains their behaviors and solves model equations and relate solutions to real system behaviors				
10. Demonstrate innovative synthesis of solution and initiate new alternatives by combining knowledge and information				
11. relate theoretical concepts to practical problem solving, predict and defend problem outcomes				

5. Important notes:

1. ABET accreditation identifies to the general public, students, school counselors, educational institutions, professional societies, employers, governmental agencies, and state boards of examiners, programs that meet minimum criteria.
2. Assessment activities will not affect your grades or any other factor related to your academic standing.
3. Much of the data collected will be anonymous and all will be kept confidential. The data will only be reported to faculty in aggregate form.

6. General Comments

Please feel free to express yourself. Thanks for your cooperation.

Departmental ABET committee

PS: Please send the filled form to atrigui@yahoo.com

F – RUBRICS FOR STUDENT OUTCOMES SUPPORTED BY EE 306

Assessment Rubric for Outcome "b"

The Graduate of the Electrical and Computer Engineering at King Abdulaziz University is expected to demonstrate an ability to design and conduct experiments, analyze and interpret data.

Indicator	Best (5)	Acceptable (3)	Poor (1)
Lab safety	**Observes** good laboratory safety procedures including electrical safety, hygiene and environmental protection	Unsafe lab procedures observed occasionally in electrical safety matters, hygiene and environmental protection	Practices unsafe, risky behaviors in lab frequently
Define objectives	**Establishes** well the need for the experiment and clearly defines the objectives	The need for the experiment is poorly stated or not mentioned at all; but the objectives are clearly defined	No mention for the experiment is available and objectives are poorly defined
Selection of variables to measure	**Identifies** important variables and chooses relevant responses to measure	Chooses relevant responses to measure, yet fails to identify all important parameters and variables that affect the measurement	Can't identify the important variables to measure without a significance clue from outside
Data gathering	**Formulates** an experimental plan of data gathering to attain a stated objective (develop correlation, test a model, ascertain performance of equipment, etc.)	Develops a simplistic experimental plan of data gathering, does not recognize entire scope of study (e.g. not all parameters affecting the results are investigated)	No systematic plan of data gathering; experimental data collection is disorganized, even random, and incomplete
Tool selection	Can **select** appropriate equipment and instruments to perform the experiment	Needs some guidance in selecting appropriate equipment and instrumentation	Cannot select the appropriate equipment and instrumentation required to run experiment(s)
Tool use	Is able to **operate** instrumentation and process equipment	Is tentative in operation of instruments and process equipment	Does not operate instrumentation and process equipment, or does so incorrectly or requires frequent supervision
Experimental procedures	**Develops** and **implements** logical experimental procedures	Experimental procedures most often are followed, but occasional oversight leads to loss of experimental efficiency and/or loss of data	Does not follow an experimental procedure
Documentation	Carefully **documents** data collected	Data collected are not all documented, units are missing, or some measurements are not recorded	Data are poorly documented
Analysis and theory of operation	Analyzes and **interprets** data carefully	Applies appropriate theory to data when prompted to do so, but misinterprets physical significance of theory or variable involved; makes errors in unit conversions	Makes a little attempt to relate data to theory
Measurement errors	Is **aware** of measurement errors and is able to account for them statistically	Is aware of measurement errors but does not account for them statistically or does	Is unaware of measurement errors

Indicator	Best (5)	Acceptable (3)	Poor (1)
		so at a minimal level	
Additional (multiple) sources (if possible)	**Seeks** information for experiment(s) from multiple sources	Seeks information for experiment(s) from a few sources - mainly from the textbook or the instructor	Seeks no extra information for experiments other than what is provided by instructor
Conclusion	**Evaluates** the experimental procedures including statistical analysis, **compares** the achievements against experimental objectives and **suggests** ways to improve the experiment	Attentively compares achievements against objectives	Compares achievements against objectives superficially

Assessment Rubric for Outcome "d"

The Graduate of the Electrical and Computer Engineering at King Abdulaziz University is expected to demonstrate an ability to function on multi-disciplinary teams.

Indicator	Best (5)	Acceptable (3)	Poor (1)
Contribution	Is **prepared** for the group meeting with clearly formulated ideas and **contributes** a fair share to the project workload. **Shares** information with others and provides assistance to others	Prepares somewhat for group meetings, but ideas are not clearly formulated. Contributes less than fair share. Sometimes keeps information to himself; not very willing to share	Routinely fails to prepare for meetings and does not contribute to group work at all or submits own work as the group's
Responsibility	**Demonstrates** the ability to assume a designated role in the group and routinely present at team meetings or work sessions	Takes charge when not in the position to lead; absent occasionally, but does not inconvenience group, sometimes depends on others to complete the work	Does not willingly assume team roles and hides in the background; only participates if strongly encouraged and is absent from team meetings or work sessions >50% of the time.
Valuing	Is courteous group member, **values** alternative perspectives and **encourages** participation among all team members. **Shares** credit for success with others and accountability for team results. Remains non-judgmental when disagreeing with others/seeks conflict resolution; does not "point fingers" or blame	Is not always considerate or courteous towards team members. Persuades others to adopt only his ideas or grudgingly accepts the ideas of others. Makes subtle references to other's poor performance or sometimes does not identify contributions of other team members and criticizes ideas of other team members or blames	Is discourteous to other group members and does not consider the ideas of others. Claims work of group as own or frequently blames others and is openly critical of the performance of others

Indicator	Best (5)	Acceptable (3)	Poor (1)
	others when things go wrong	others for errors	
Cooperation with other disciplines	**Cooperates** with others (outside of the discipline) and has knowledge of technical skills, issues and approaches relevant to disciplines outside of electrical engineering	Occasionally works as a loner or interacts to a minor extent with extra-disciplinary team members. Has some knowledge of other disciplines, but gets lost in discussions with extra-disciplinary team members	Does work on his own; does not value team work. Has no knowledge of disciplines outside of electrical engineering

Assessment Rubric for Outcome "f"

The Graduate of the Electrical and Computer Engineering at King Abdulaziz University is expected to demonstrate an understanding of professional and ethical responsibility.

Indicator	Best (5)	Acceptable (3)	Poor (1)
Participation in ethical discussions	**Participates** in class discussions and exercises on ethics and professionalism	Does not take the discussion of ethics seriously but is willing to accept its existence	Does not participate in or contribute to discussions of ethics but has some awareness for the need for professional ethics
Behavior	**Demonstrates** ethical behavior among peers and faculty	Does not model ethical behavior among peers and faculty	Student has been caught cheating or plagiarizing the work of others occasionally
Responsibility	**Takes** personal responsibility for his actions	Doesn't recognize the need to take personal responsibility for his actions	Blames others for own issues and problems
Respect to others	Is punctual, professional, and collegial; **attends** classes regularly	Sometimes exhibits unprofessional behavior; is sometimes absent from class without reason	Is frequently absent from class and is generally not collegial to fellow students, staff, and faculty
Objectivity	Evaluates and **judges** a situation in practice or as a case study, using facts and a professional code of ethics	Evaluates and judges a situation in practice or as a case study using personal understanding of the situation, possibly applying a personal value system	Evaluates and judges a situation in practice or as a case study using a biased perspective without objectivity
Personal versus professional ethics	Uses personal value system to support actions, but **understands** the role of professional ethical	Uses personal value system to support actions, but confuses personal ethics with professional	Uses personal value system to support actions to the exclusion of all other ethical standards

Indicator	Best (5)	Acceptable (3)	Poor (1)
	standards for corporate decisions	ethics	

Assessment Rubric for Outcome "k"

The Graduate of the Electrical and Computer Engineering at King Abdulaziz University is expected to demonstrate an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

Indicator	Best (5)	Acceptable (3)	Poor (1)
Modern software tools	Is able to learn and **implement** process simulation software and uses computer-based and other resources effectively in assignments/projects	Is able to implement process simulation software with little help and attempts to use computer-based and other resources in assignments/ projects	Is not able to learn and implement process simulation software even with considerable help and doesn't use computer-based and other resources effectively in assignments/ projects
Skill maintenance	Is able to **interpret** and understand information from a variety of resources	Is able to understand information from a variety of resources but can't properly interpret them	Has difficulty in understanding information from a variety of resources and eventually he can't draw healthy conclusions
Outside resources	Understands the organization and use of the library and **seeks** information on problems from multiple resources	Doesn't understand the organization, but can use the library and seeks information on problems from a few sources	Doesn't understand the organization and use of the library and seeks information on problems only from textbooks

Assessment Rubric for Outcome "l"

The Graduate of the Electrical and Computer Engineering Program at King Abdulaziz University is expected to demonstrate ***knowledge of probability and statistics, including applications in instrumentations, systems and measurements related to his specialization.***

Indicator	Best (5)	Acceptable (3)	Poor (1)
Statistical measures	Determines statistical measures such as accuracy, precision, resolution etc for a measuring equipment and uses them.	Mention statistical measures such as accuracy, precision, resolution etc for a measuring equipment and uses them.	No mention of statistical measures such as accuracy, precision, resolution etc for a measuring equipment but indications of some use of them.
Tolerances	Correctly infers tolerances of electronic components in	Infers tolerances of electronic components in	No use of tolerances of electronic components in

Indicator	Best (5)	Acceptable (3)	Poor (1)
	design of electronic devices.	design of electronic devices with errors in calculations.	design of electronic devices although there is some mentioning of errors.
Data analyzes	Correctly analyzes data sets using statistical concepts • Calculate important measures such as the mean and standard deviation from a normal distribution of data. • Recognize the statistical variability of biological systems and value the population statistics (for BME).	Minor errors in statistical analysis of data • Calculate important measures such as the mean and standard deviation from a normal distribution of data with errors. • Aware of the statistical variability of biological systems but doesn't value the population statistics (for BME).	Demonstrates some awareness of statistics to analysis of data without any application examples.

INDEX