The McGraw-Hill Companies

Electronic Instrumentation

Third Edition

About the Author

H S Kalsi obtained a Diploma in Electronics and Radio Engineering (DERE) from St Xavier's Technical Institute, Mumbai, and a Diploma in Technical Teaching from TTTI, Bhopal. Presently, he is Lecturer (Selection Grade) in the Department of Electronics at St Xavier's Technical Institute, Mumbai. His area of special interest is Electronic Instrumentation. He is also a member of ISTE.

The McGraw·Hill Companies

Electronic Instrumentation

Third Edition

H S Kalsi

Lecturer (Selection Grade)
St Xavier's Technical Institute
Mumbai

Tata McGraw Hill Education Private Limited
NEW DELHI

McGraw-Hill Offices

New Delhi New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

Tata McGraw-Hill

Published by Tata McGraw Hill Education Private Limited,
7 West Patel Nagar, New Delhi 110 008

Electronic Instrumentation, 3e

ISBN (13): 978-0-07-070206-6
ISBN (10): 0-07-070206-3

Managing Director: *Ajay Shukla*

Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Manager—Sponsoring SEM & Tech Ed: *Shalini Jha*
Assoc. Sponsoring Editor: *Suman Sen*
Development Editor: *Manish Choudhary*
Executive—Editorial Services: *Sohini Mukharjee*
Sr Production Manager: *P L Pandita*

Dy Marketing Manager—SEM & Tech Ed: *Biju Ganesan*

General Manager—Production: *Rajender P Ghansela*
Asst. General Manager—Production: *B L Dogra*

Typeset at Text-o-Graphics, B1/56 Arawali Apartment, Sector 34, Noida 201301 and
printed at Gopal Jee Enterprises, 190/5, Main Road, Maujpur, Delhi 110 053

Cover Printer: SDR Printers

DZXLCRAZRQCLY

The McGraw-Hill Companies

In the everloving and fond memory

of

My Elder Brother

Contents

Preface

The tremendous and overwhelming response received by the second edition of this book inspired me to bring out the third edition. This new edition has been revised and updated based on the suggestions received from the students and teachers using this book.

The book is written in a simple and lucid manner with systematically arranged chapters which enable the reader to get thorough knowledge, starting from the basic concepts to the sophisticated advancements of all types of measuring instruments and measurement techniques.

With the advancement of technology in integrated circuits, instruments are becoming increasingly compact and accurate. In view of this, sophisticated types of instruments covering digital and microprocessor-based instruments are dealt with in detail in a systematic manner for easy understanding. The basic concepts, working operation, capabilities and limitations of the instruments discussed in the book will also guide the users in selecting the right instrument for different applications.

New to this Edition

- ❖ Inclusion of new topics on Telemetry, Electric and Voltage Standards and Rotational Variable Differential Transducers (RVDT)
- ❖ Expanded coverage of Bridges which now includes Maxwell–Wien Bridge, Anderson Bridge, Carey–Foster Bridge, De Sauty Bridge and Owen Bridge
- ❖ Thoroughly revised pedagogy including
    - ☞ 300 Review questions
    - ☞ 200 Objective type questions
    - ☞ 125 Solved examples and practice questions, with easy steps introduced for solved examples

Chapter Organisation

Chapter 1 covers the basic characteristics and the errors associated with instruments. Different types of indicating and display devices are dealt with in detail in **Chapter 2**. This chapter discusses different types of printers and printer heads used with computers.

The basic analog-type ammeters for both dc and RF frequencies and different types of voltmeters, ohmmeters and multimeters are discussed in **Chapters 3** and **4**.

Digital instruments ranging from a simple digital voltmeter to a microprocessor-based instrument and their measurement techniques are presented in a comprehensible manner for easy understanding in **Chapters 5** and **6**. **Chapter 7** on oscilloscopes has been dealt with in depth to familiarize the students with the working of all types of Cathode Ray Oscilloscopes (CROs) and their measurement techniques. **Chapter 8** pertains to signal generation. **Chapter 9** analyses the frequency component of a generated wave, and its distortion.

In industry, it is required to transmit signals or the changes in parameters from the measurement site location to the control room. Hence in **Chapter 10**, telemetry systems have been covered to get a brief insight of the various transmission methods used in industry.

Most instruments used in process control plants measure various parameters such as resistance, inductance, capacitance, dissipation factor, temperature, etc. To obtain accurate measurement of the changes in parameters, bridges are used. Hence, **Chapter 11** covers most of the types of bridges used for measurement of different parameters, for example, Wheatstone's bridge, Maxwell's Bridge, Hay's Bridge, Schering Bridge, etc. Instruments and the instrumentation systems also use bridges as the input stage.

Chapters 12, 13 and **14** cover the essential components of industrial instruments used for measurements and their usage.

Different types of analog and digital filters are given in **Chapter 15.** A mathematical approach to explaining digital filters has been adopted to provide the students a clear insight into their working. **Chapter 16** is on the measurement of microwave frequencies. A detailed discussion on the data acquisition system along with the latest data logger is covered in **Chapter 17.** Instruments from remote places transmit signals over long distances to a master control room where they are displayed. This transmission of signals has been explained in detail in **Chapter 18.**

Frequency standards and measurement of power at RF and microwave frequencies are dealt with in **Chapters 19** and **20** respectively. **Chapter 21** discusses control systems, electronic control systems in particular. This chapter covers the basic control systems, electronic control systems, electronic controllers, PLC and advanced control systems such as DCS used in process control plants.

Web Supplements

The Web supplements can be accessed at *http://www.mhhe.com/kalsi/ei3*, which contains the following:

For Instructors

Solution Manual, PowerPoint Lecture Slides

For Students

Additional Review Questions and Web links for useful reference materials.

Acknowledgements

First of all, I express my deepest thanks and gratitude to my younger brother who gave me all his support, without which it would have been difficult to complete this project. Secondly, I thank all the reviewers for their important suggestions which helped me a lot while revising this book. Their names are given below.

Nilesh Chaurasia	Sri Vaishnav Institute of Technology and Science Indore, Madhya Pradesh
Praveen Tiwari	IIMT Engineering College Meerut, Uttar Pradesh
S B L Seksena	National Institute of Technology (NIT) Jamshedpur, Jharkhand
Sunit Kumar Sen	University College of Science and Technology Kolkata, West Bengal
Samir Ekbote	Datta Meghe College of Engineering Navi Mumbai, Maharashtra
V G Sarode	Xavier Institute of Engineering Mumbai, Maharashtra
Sreeprabha P K	Indian Institute of Space Science and Technology (IIST), Thiruvanathpuram, Kerala
P Thirumurthy	Kumaraguru College of Technology Coimbatore, Tamil Nadu
K E Srinivas Murthy	Sri Venkateswara Institute of Technology Anantpur, Andhra Pradesh

Signet Electronics Ltd, who made available the photographs of instruments, deserves a special note of appreciation.

I am deeply indebted to my wife and other family members for their never-ending encouragement, moral support and patience during the preparation of this book.

I appreciate the efforts of all members of the Tata McGraw-Hill family, especially Shalini Jha, Suman Sen, Manish Choudhary, Sohini Mukherjee and P L Pandita, whose help and cooperation shaped the book in all its stages.

Last, but not the least, I would like to thank my friends and colleagues who helped me in the writing of this book.

I hope that this edition of the book will prove useful to all readers, students as well as teachers. All suggestions for further improvement of the book are welcome and will be gratefully acknowledged. You can send your feedback to my email id: *hs_kalsi@yahoo.com*

H S Kalsi

Publisher's Note

Tata McGraw Hill invites comments, views and suggestions from readers, all of which can be sent to *tmh.ecefeedback@gmail.com*. Piracy-related issues may also be reported.

List of Abbreviations

µA	Micro Amperes
ADC	Analog to Digital Converter
AF	Audio Frequency
ALU	Arithmetic Logic Unit
AM	Amplitude Modulation
ASCII	American Standard Code for Information Interchange
B.W	Bandwidth
BCD	Binary Coded Decimal
BFO	Beat Frequency Oscillator
BJT	Bipolar Junction Transistor
CCIR	Committee Counsultatif International Radio telecommunique
CDRX	Critically Damping External Resistance
CL	Control Language
CMRR	Common Mode Rejection Ratio
cps	Character per second
CRO	Cathode Ray Oscilloscope
CRT	Cathode Ray Tube
CSC	Computer Supervisory Control
CVSD	Continuous Variable Slope DM
DAC	Digital to Analog Converter
DAS	Data Acquisition System
dB	deci-Bel
DCE	Data Circuit terminating Equipment
DCS	Distributed Control System
DDA	Decade Divider Assemblies
DDC	Direct Digital Control
DFT	Discrete Fourier Transform
DM	Delta Modulation
DMM	Digital Multimeter
DPDT	Double Pole Double Throw
DPM	Digital Panel Meter

DPST	Double Pole Single Throw
DSO	Digital Storage Oscilloscope
DTE	Data Terminal Equipment
DVM	Digital Voltmeter
DWR	Digital Waveform Recorder
EDM	Electrodynometer
EHT	Extra High Tension
EL	Electro- Luminescent
EM	Electro Magnetic
EPID	ElectroPhoretic Image Display
EPROM	Erasable Programmable Read Only Memory
EXT	External
F/F	Flip-Flop
FDM	Frequency Division Multiplexing
FET	Field Effect Transistor
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FM	Frequency Modulation
FSK	Frequency Shift Keying
GaAsP	Gallium Arsenide Phosphide
GaP	Gallium Phosphide
GHz	Giga Hertz
HDP	Horizontal Deflection Plates
HF	High Frequency
HV	High Voltage
Ifsd	Full scale deflection Current
IIR	Infinite Impulse Response
INT	Internal
Ku	Kilo unit
LCD	Liquid Crystal Display
LED	Light Emitting Diode
lpm	line per minute
LSB	Least Significant Bit
LV	Low voltage
LVD	Liquid Vapour Display
LVDT	Linear Variable Differential Transformer
mA/mV	milli-Amperes/milli-Volts
MCR	Master Control Reset
MHz	Mega Hertz
MOS	Metallic Oxide Semiconductor
MSB	Most Significant Bit
NIDM	Non-Impact Dot Matrix
NLC	Nematic Liquid Crystal
NO/NC	Normal Open/Norma Close

NRZ	Non Return to Zero
NTC	Negative Temperature Co-efficient
PCB	Printed Circuit Board
PCM	Pulse Code Modulation
PDM	Pulse Duration Modulation
PGA	Programmable Gain Amplifier
PLC	Programmable Logic Controller
PMMC	Permanent Magnet Moving Coil
p-p	peak to peak
PPM	Pulse Position Modulation
PSK	Phase Shift Keying
PTC	Positive Temperature Co-efficient
PWM	Pulse Width Modulation
RAM	Random Access Memory
RF	Radio Frequency
RMS	Root Mean Square
ROM	Read Only Memory
RTD	Resistance Temperature Detector
RVDT	Rotary Variable Differential Transformer
RZ	Return to Zero
S/H	Sample and Hold
SAR	Successive Approximation Register
SMPTE	Society of Motion Pictures and Television Engineers
TC	Thermocouple
TDM	Time Division Multiplexing
TRP	Total Radiation Pyrometer
TTL	Transistor Transistor logic
TV	Television
TVM	Transistor Voltmeter
UART	Universal Asynchronous Receiver Transmitter
UDC	Up-Down Counter
UHF	Ultra High Frequency
UJT	Unijunction Transistor
VDP	Vertical Deflection Plates
VHF	Very High Frequency
VLF	Very Low Frequency
VTVM	Vacuum Tube Voltmeter

List of Important Formulae

1.	Absolute Error	$E = Y_n - X_n$ (Where E = Absolute Error Y_n = Expected Value X_n = Measured Value)
2.	Accuracy	$A = 1 - \left\lvert \dfrac{Y_n - X_n}{X_n} \right\rvert$
3.	Deflecting torque	$\tau_d = B \times A \times N \times I$
4.	Shunt resistance for Ammeter	$R_{sh} = \dfrac{I_m R_m}{I - I_m}$
5.	Multiplier for dc Voltmeter	$R_s = \dfrac{V}{I_m} - R_m$
6.	Multiplier Resistor for ac Range	$R_s = \dfrac{0.45 \times E_{\text{rms}}}{I_{dc}} - R_m$
7.	R_1 and R_2 for an Ohmmeter	$R_1 = R_h - \dfrac{If_{sd} R_m R_h}{V}$ $R_2 = \dfrac{If_{sd} R_m R_h}{V - If_{sd} R_h}$
8.	Sensitivity of Digital Meters	$S = (f_s)_{\text{min}} \times R$ (Where $(f_s)_{\text{min}}$ = lowest full scale on meter $R = 1/10^n$ n = number of full digits)
9.	Resistance Value for Wheatstone's Bridge	$R_x = \dfrac{R_2 R_3}{R_1}$

10.	Maxwell's Bridge	$R_x = \dfrac{R_2 R_3}{R_1}$ and $L_x = C_1 R_2 R_3$ $Q = w\, C_1 R_1$
11.	Hay's Bridge	$L_x = \dfrac{R_2 R_3 C_1}{1 + w^2 C_1^2 R_1^2}$ $R_x = \dfrac{w C_1 R_2 R_3}{1 + w^2 C_1^2 R_1^2}$
12.	Schering Bridge	$R_x = \dfrac{R_2 C_1}{C_3}$ $C_x = \dfrac{R_1 C_3}{R_2}$
13.	Wien's Bridge	$R_2/R_4 = 2$ and $f = \dfrac{1}{2\pi RC}$
14.	Resistance	$R = \dfrac{\sigma x l}{A}$
15.	Gage Factor	$GF(K) = \dfrac{\Delta R/R}{\Delta l/l}$
16.	Resistance of Conductor	$R_t = R_{\text{ref}}(1 + \alpha\,\Delta t)$
17.	For a Dual Slope DVM	$e_i = \dfrac{n_2\, xer}{n_1}$
18.	Input Capacitance of a CRO Probe	$C_1 = \dfrac{R_{in}(C_{in} + C_2)}{R_1}$
19.	Distributed Capacitance	$C_s = \dfrac{C_1 - 4C_2}{3}$
20.	Closed-loop Voltage gain for Non-Inverting Amplifier	$A_F = \left(1 + \dfrac{RF}{R_1}\right)$

21.	Closed-loop Voltage gain for Inverting Amplifier	$A_F = \left(-\frac{RF}{R_1}\right)$
22.	Output Voltage of an Instrumentation Amplifier	$A_F = \left(1+\frac{2R_2}{R_1}\right)(e_2 - e_1)$

Qualities of Measurements

Chapter 1

INTRODUCTION 1.1

Instrumentation is a technology of measurement which serves not only science but all branches of engineering, medicine, and almost every human endeavour. The knowledge of any parameter largely depends on the measurement. The indepth knowledge of any parameter can be easily understood by the use of measurement, and further modifications can also be obtained.

Measuring is basically used to monitor a process or operation, or as well as the controlling process. For example, thermometers, barometers, anemometers are used to indicate the environmental conditions. Similarly, water, gas and electric meters are used to keep track of the quantity of the commodity used, and also special monitoring equipment are used in hospitals.

Whatever may be the nature of application, intelligent selection and use of measuring equipment depends on a broad knowledge of what is available and how the performance of the equipment renders itself for the job to be performed.

But there are some basic measurement techniques and devices that are useful and will continue to be widely used also. There is always a need for improvement and development of new equipment to solve measurement problems.

The major problem encountered with any measuring instrument is the error. Therefore, it is obviously necessary to select the appropriate measuring instrument and measurement method which minimises error. To avoid errors in any experimental work, careful planning, execution and evaluation of the experiment are essential.

The basic concern of any measurement is that the measuring instrument should not effect the quantity being measured; in practice, this non-interference principle is never strictly obeyed. Null measurements with the use of feedback in an instrument minimise these interference effects.

PERFORMANCE CHARACTERISTICS 1.2

A knowledge of the performance characteristics of an instrument is essential for selecting the most suitable instrument for specific measuring jobs. It consists of two basic characteristics—static and dynamic.

STATIC CHARACTERISTICS 1.3

The static characteristics of an instrument are, in general, considered for instruments which are used to measure an unvarying process condition. All the static performance characteristics are obtained by one form or another of a process called calibration. There are a number of related definitions (or characteristics), which are described below, such as accuracy, precision, repeatability, resolution, errors, sensitivity, etc.

1. ***Instrument*** A device or mechanism used to determine the present value of the quantity under measurement.
2. ***Measurement*** The process of determining the amount, degree, or capacity by comparison (direct or indirect) with the accepted standards of the system units being used.
3. ***Accuracy*** The degree of exactness (closeness) of a measurement compared to the expected (desired) value.
4. ***Resolution*** The smallest change in a measured variable to which an instrument will respond.
5. ***Precision*** A measure of the consistency or repeatability of measurements, i.e. successive reading do not differ. (Precision is the consistency of the instrument output for a given value of input).
6. ***Expected value*** The design value, i.e. the most probable value that calculations indicate one should expect to measure.
7. ***Error*** The deviation of the true value from the desired value.
8. ***Sensitivity*** The ratio of the change in output (response) of the instrument to a change of input or measured variable.

ERROR IN MEASUREMENT 1.4

Measurement is the process of comparing an unknown quantity with an accepted standard quantity. It involves connecting a measuring instrument into the system under consideration and observing the resulting response on the instrument. The measurement thus obtained is a quantitative measure of the so-called "true value" (since it is very difficult to define the true value, the term "expected value" is used). Any measurement is affected by many variables, therefore the results rarely reflect the expected value. For example, connecting a measuring instrument into the circuit under consideration always disturbs (changes) the circuit, causing the measurement to differ from the expected value.

Some factors that affect the measurements are related to the measuring instruments themselves. Other factors are related to the person using the instrument. The degree to which a measurement nears the expected value is expressed in terms of the error of measurement.

Error may be expressed either as absolute or as percentage of error.

Absolute error may be defined as the difference between the expected value of the variable and the measured value of the variable, or

$$e = Y_n - X_n$$

where e = absolute error
Y_n = expected value
X_n = measured value

Therefore $$\% \text{ Error} = \frac{\text{Absolute value}}{\text{Expected value}} \times 100 = \frac{e}{Y_n} \times 100$$

Therefore $$\% \text{ Error} = \left(\frac{Y_n - X_n}{Y_n}\right) \times 100$$

It is more frequently expressed as a accuracy rather than error.

Therefore $$A = 1 - \left|\frac{Y_n - X_n}{Y_n}\right|$$

where A is the relative accuracy.
Accuracy is expressed as % accuracy

$$a = 100\% - \% \text{ error}$$
$$a = A \times 100 \%$$

where a is the % accuracy.

Example 1.1 (a) *The expected value of the voltage across a resistor is 80 V. However, the measurement gives a value of 79 V. Calculate (i) absolute error, (ii) % error, (iii) relative accuracy, and (iv) % of accuracy.*

Solution

(i) Absolute error $e = Y_n - X_n = 80 - 79 = 1$ V

(ii) $\% \text{ Error} = \dfrac{Y_n - X_n}{Y_n} \times 100 = \dfrac{80 - 79}{80} \times 100 = 1.25\%$

(iii) Relative Accuracy

$$A = 1 - \left|\frac{Y_n - X_n}{Y_n}\right| = 1 - \left|\frac{80 - 79}{80}\right|$$

$\therefore$ $A = 1 - 1/80 = 79/80 = 0.9875$

(iv) % of Accuracy $a = 100 \times A = 100 \times 0.9875 = 98.75\%$
or $a = 100\% - \%$ of error $= 100\% - 1.25\% = 98.75\%$

Example 1.1 (b) *The expected value of the current through a resistor is 20 mA. However the measurement yields a current value of 18 mA. Calculate (i) absolute error (ii) % error (iii) relative accuracy (iv) % accuracy*

Solution

Step 1: Absolute error

$$e = Y_n - X_n$$

where e = error, Y_n = expected value, X_n = measured value

Given $Y_n = 20$ mA and $X_n = 18$ mA
Therefore $e = Y_n - X_n = 20 \text{ mA} - 18 \text{ mA} = 2 \text{ mA}$

Step 2: % error

$$\% \text{ error} = \frac{Y_n - X_n}{Y_n} \times 100 = \frac{20 \text{ mA} - 18 \text{ mA}}{20 \text{ mA}} \times 100 = \frac{2 \text{ mA}}{20 \text{ mA}} \times 100 = 10\%$$

Step 3: Relative accuracy

$$A = 1 - \left| \frac{Y_n - X_n}{Y_n} \right| = 1 - \left| \frac{20 \text{ mA} - 18 \text{ mA}}{20 \text{ mA}} \right| = 1 - \frac{2}{20} = 1 - 0.1 = 0.90$$

Step 4: % accuracy

$$a = 100\% - \%\text{error} = 100\% - 10\% = 90\%$$

and
$$a = A \times 100\% = 0.90 \times 100\% = 90\%$$

If a measurement is accurate, it must also be precise, i.e. Accuracy means precision. However, a precision measurement may not be accurate. (The precision of a measurement is a quantitative or numerical indication of the closeness with which a repeated set of measurement of the same variable agree with the average set of measurements.) Precision can also be expressed mathematically as

$$P = 1 - \left| \frac{X_n - \bar{X}_n}{\bar{X}_n} \right|$$

where X_n = value of the nth measurement
$\bar{X}_n$ = average set of measurement

Example 1.2 *Table 1.1 gives the set of 10 measurement that were recorded in the laboratory. Calculate the precision of the 6th measurement.*

Table 1.1

Measurement number	*Measurement value X_n*
1	98
2	101
3	102
4	97
5	101
6	100
7	103
8	98
9	106
10	99

Solution The average value for the set of measurements is given by

$$\overline{X}_n = \frac{\text{Sum of the 10 measurement values}}{10}$$

$$= \frac{1005}{10} = 100.5$$

$$\text{Precision} = 1 - \left| \frac{X_n - \overline{X}_n}{\overline{X}_n} \right|$$

For the 6th reading

$$\text{Precision} = 1 - \left| \frac{100 - 100.5}{100.5} \right| = 1 - \frac{0.5}{100.5} = \frac{100}{100.5} = 0.995$$

The accuracy and precision of measurements depend not only on the quality of the measuring instrument but also on the person using it. However, whatever the quality of the instrument and the case exercised by the user, there is always some error present in the measurement of physical quantities.

TYPES OF STATIC ERROR 1.5

The static error of a measuring instrument is the numerical difference between the true value of a quantity and its value as obtained by measurement, i.e. repeated measurement of the same quantity gives different indications. Static errors are categorised as gross errors or human errors, systematic errors, and random errors.

1.5.1 Gross Errors

These errors are mainly due to human mistakes in reading or in using instruments or errors in recording observations. Errors may also occur due to incorrect adjustment of instruments and computational mistakes. These errors cannot be treated mathematically.

The complete elimination of gross errors is not possible, but one can minimise them. Some errors are easily detected while others may be elusive.

One of the basic gross errors that occurs frequently is the improper use of an instrument. The error can be minimized by taking proper care in reading and recording the measurement parameter.

In general, indicating instruments change ambient conditions to some extent when connected into a complete circuit. (Refer Examples 1.3(a) and (b)).

(One should therefore not be completely dependent on one reading only; at least three separate readings should be taken, preferably under conditions in which instruments are switched off and on.)

1.5.2 Systematic Errors

These errors occur due to shortcomings of the instrument, such as defective or worn parts, or ageing or effects of the environment on the instrument.

These errors are sometimes referred to as bias, and they influence all measurements of a quantity alike. A constant uniform deviation of the operation of an instrument is known as a systematic error. There are basically three types of systematic errors—(i) Instrumental, (ii) Environmental, and (iii) Observational.

(i) Instrumental Errors

Instrumental errors are inherent in measuring instruments, because of their mechanical structure. For example, in the D'Arsonval movement, friction in the bearings of various moving components, irregular spring tensions, stretching of the spring, or reduction in tension due to improper handling or overloading of the instrument.

Instrumental errors can be avoided by

(a) selecting a suitable instrument for the particular measurement applications. (Refer Examples 1.3 (a) and (b)).
(b) applying correction factors after determining the amount of instrumental error.
(c) calibrating the instrument against a standard.

(ii) Environmental Errors

Environmental errors are due to conditions external to the measuring device, including conditions in the area surrounding the instrument, such as the effects of change in temperature, humidity, barometric pressure or of magnetic or electrostatic fields.

These errors can also be avoided by (i) air conditioning, (ii) hermetically sealing certain components in the instruments, and (iii) using magnetic shields.

(iii) Observational Errors

Observational errors are errors introduced by the observer. The most common error is the parallax error introduced in reading a meter scale, and the error of estimation when obtaining a reading from a meter scale.

These errors are caused by the habits of individual observers. For example, an observer may always introduce an error by consistently holding his head too far to the left while reading a needle and scale reading.

In general, systematic errors can also be subdivided into static and dynamic errors. Static errors are caused by limitations of the measuring device or the physical laws governing its behaviour. Dynamic errors are caused by the instrument not responding fast enough to follow the changes in a measured variable.

Example 1.3 (a) *A voltmeter having a sensitivity of 1 kΩ/V is connected across an unknown resistance in series with a milliammeter reading 80 V on 150 V scale. When the milliammeter reads 10 mA, calculate the (i) Apparent resistance of the unknown resistance, (ii) Actual resistance of the unknown resistance, and (iii) Error due to the loading effect of the voltmeter.*

Solution

(i) The total circuit resistance $R_T = \dfrac{V_T}{I_T} = \dfrac{80}{10 \text{ mA}} = 8 \text{ k}\Omega$

(Neglecting the resistance of the milliammeter.)

(ii) The voltmeter resistance equals $R_v = 1000\ \Omega/\text{V} \times 150 = 150 \text{ k}\Omega$

$\therefore$ actual value of unknown resistance $R_x = \dfrac{R_T \times R_v}{R_v - R_T} = \dfrac{8\text{k} \times 150\text{k}}{150\text{k} - 8\text{k}}$

$$= \frac{1200\text{k}^2}{142\text{k}} = 8.45\ \Omega$$

(iii) % error $= \dfrac{\text{Actual value} - \text{Apparent value}}{\text{Actual value}} = \dfrac{8.45\text{k} - 8\text{k}}{8.45\text{k}} \times 100$

$= 0.053 \times 100 = 5.3\%$

Example 1.3 (b) *Referring to Ex. 1.3 (a), if the milliammeter reads 600 mA and the voltmeter reads 30 V on a 150 V scale, calculate the following: (i) Apparent, resistance of the unknown resistance. (ii) Actual resistance of the unknown resistance. (iii) Error due to loading effect of the voltmeter.*

Comment on the loading effect due to the voltmeter for both Examples 1.3 (a) and (b). (Voltmeter sensitivity given 1000 Ω/V.)

Solution

1. The total circuit resistance is given by

$$R_T = \frac{V_T}{I_T} = \frac{30}{0.6} = 50\ \Omega$$

2. The voltmeter resistance R_v equals

$$R_v = 1000\ \Omega/\text{V} \times 150 = 150 \text{ k}\Omega$$

Neglecting the resistance of the milliammeter, the value of unknown resistance = 50 Ω

$$R_x = \frac{R_T \times R_v}{R_v - R_T} = \frac{50 \times 150\text{ k}}{150\text{ k} - 50} = \frac{7500\text{ k}}{149.5\text{ k}} = 50.167\ \Omega$$

$$\%\ \text{Error} = \frac{50.167 - 50}{50.167} \times 100 = \frac{0.167}{50.167} \times 100 = 0.33\%$$

In Example 1.3 (a), a well calibrated voltmeter may give a misleading resistance when connected across two points in a high resistance circuit. The same voltmeter, when connected in a low resistance circuit (Example 1.3 (b)) may give a more dependable reading. This show that voltmeters have a loading effect in the circuit during measurement.

1.5.3 Random Errors

These are errors that remain after gross and systematic errors have been substantially reduced or at least accounted for. Random errors are generally an accumulation of a large number of small effects and may be of real concern only in measurements requiring a high degree of accuracy. Such errors can be analyzed statistically.

These errors are due to unknown causes, not determinable in the ordinary process of making measurements. Such errors are normally small and follow the laws of probability. Random errors can thus be treated mathematically.

For example, suppose a voltage is being monitored by a voltmeter which is read at 15 minutes intervals. Although the instrument operates under ideal environmental conditions and is accurately calibrated before measurement, it still gives readings that vary slightly over the period of observation. This variation cannot be corrected by any method of calibration or any other known method of control.

SOURCES OF ERROR 1.6

The sources of error, other than the inability of a piece of hardware to provide a true measurement, are as follows:

1. Insufficient knowledge of process parameters and design conditions
2. Poor design
3. Change in process parameters, irregularities, upsets, etc.
4. Poor maintenance
5. Errors caused by person operating the instrument or equipment
6. Certain design limitations

DYNAMIC CHARACTERISTICS 1.7

Instruments rarely respond instantaneously to changes in the measured variables. Instead, they exhibit slowness or sluggishness due to such things as mass, thermal capacitance, fluid capacitance or electric capacitance. In addition to this, pure delay in time is often encountered where the instrument waits for some reaction to take place. Such industrial instruments are nearly always used for measuring quantities that fluctuate with time. Therefore, the dynamic and transient behaviour of the instrument is as important as the static behaviour.

The dynamic behaviour of an instrument is determined by subjecting its primary element (sensing element) to some unknown and predetermined variations in the measured quantity. The three most common variations in the measured quantity are as follows:

1. *Step* change, in which the primary element is subjected to an instantaneous and finite change in measured variable.
2. *Linear* change, in which the primary element is following a measured variable, changing linearly with time.

3. *Sinusoidal* change, in which the primary element follows a measured variable, the magnitude of which changes in accordance with a sinusoidal function of constant amplitude.

The dynamic characteristics of an instrument are (i) speed of response, (ii) fidelity, (iii) lag, and (iv) dynamic error.

(i) *Speed of Response* It is the rapidity with which an instrument responds to changes in the measured quantity.

(ii) *Fidelity* It is the degree to which an instrument indicates the changes in the measured variable without dynamic error (faithful reproduction).

(iii) *Lag* It is the retardation or delay in the response of an instrument to changes in the measured variable.

(iv) *Dynamic Error* It is the difference between the true value of a quantity changing with time and the value indicated by the instrument, if no static error is assumed.

When measurement problems are concerned with rapidly varying quantities, the dynamic relations between the instruments input and output are generally defined by the use of differential equations.

1.7.1 Dynamic Response of Zero Order Instruments

We would like an equation that describes the performance of the zero order instrument exactly. The relations between any input and output can, by using suitable simplifying assumptions, be written as

$$a_n \frac{d^n x_o}{dt^n} + a_{n-1} \frac{d^{n-1} x_o}{dt^{n-1}} + \cdots + a_1 \frac{dx_o}{dt} + a_0 x_o$$

$$= b_m \frac{d^m x_i}{dt^m} + \cdots + b_{m-1} \frac{d^{m-1} x_i}{dt^{m-1}} + \cdots + b_1 \frac{dx_i}{dt} + b_0 \simeq x_i \quad (1.1)$$

where x_o = output quantity
x_i = input quantity
t = time

a's and b's are combinations of systems physical parameters, assumed constant.

When all the a's and b's, other than a_0 and b_0 are assumed to be zero, the differential equation degenerates into the simple equation given as

$$a_0 x_o = b_0 x_i \quad (1.2)$$

Any instrument that closely obeys Eq. (1.2) over its intended range of operating conditions is defined as a zero-order instrument. The static sensitivity (or steady state gain) of a zero-order instrument may be defined as follows

$$x_o = \frac{b_0}{a_0} x_i = K.x_i$$

where $K = b_0/a_0$ = static sensitivity

Since the equation $x_o = K\,x_i$ is an algebraic equation, it is clear that no matter how x_i might vary with time, the instrument output (reading) follows it perfectly with no distortion or time lag of any sort. Thus, a zero-order instrument represents ideal or perfect dynamic performance. A practical example of a zero order instrument is the displacement measuring potentiometer.

1.7.2 Dynamic Response of a First Order Instrument

If in Eq. (1.1) all a's and b's other than a_1, a_0, b_0 are taken as zero, we get

$$a_1 \frac{dx_o}{dt} + a_0 x_o = b_0 x_i$$

Any instrument that follows this equation is called a first order instrument. By dividing by a_0, the equation can be written as

$$\frac{a_1}{a_0}\frac{dx_o}{dt} + x_o = \frac{b_0}{a_0} x_i$$

or
$$(\tau \cdot D + 1) \cdot x_o = K\,x_i$$

where $\tau = a_1/a_0$ = time constant

$K = b_0/a_0$ = static sensitivity

The time constant τ always has the dimensions of time while the static sensitivity K has the dimensions of output/input. The operational transfer function of any first order instrument is

$$\frac{x_o}{x_i} = \frac{K}{\tau D + 1}$$

A very common example of a first-order instrument is a mercury-in-glass thermometer.

1.7.3 Dynamic Response of Second Order Instrument

A second order instrument is defined as one that follows the equation

$$a_2 \frac{d^2 x_o}{dt^2} + a_1 \frac{dx_o}{dt} + a_0 x_o = b_0 x_i$$

The above equations can be reduced as

$$\left(\frac{D^2}{\omega_n^2} + \frac{2\xi D}{\omega_n} + 1\right) \cdot x_o = K x_i$$

where $\omega_n = \sqrt{\dfrac{a_0}{a_2}}$ = undamped natural frequency in radians/time

$2\xi = a_1/\sqrt{a_0\,a_2}$ = damping ratio

$K = b_0/a_0$ = static sensitivity

Any instrument following this equation is a second order instrument. A practical example of this type is the spring balance. Linear devices range from mass spring arrangements, transducers, amplifiers and filters to indicators and recorders.

Most devices have first or second order responses, i.e. the equations of motion describing the devices are either first or second order linear differentials. For example, a search coil and mercury-in-glass thermometer have a first order response. Filters used at the output of a phase sensitive detector and amplifiers used in feedback measuring systems essentially have response due to a single time constant. First order systems involve only one kind of energy, e.g. thermal energy in the case of a thermometer, while a characteristic feature of second order system is an exchange between two types of energy, e.g. electrostatic and electromagnetic energy in electrical LC circuits, moving coil indicators and electromechanical recorders.

STATISTICAL ANALYSIS 1.8

The statistical analysis of measurement data is important because it allows an analytical determination of the uncertainty of the final test result. To make statistical analysis meaningful, a large number of measurements is usually required. Systematic errors should be small compared to random errors, because statistical analysis of data cannot remove a fixed bias contained in all measurements.

1.8.1 Arithmetic Mean

The most probable value of a measured variable is the arithmetic mean of the number of readings taken. The best approximation is possible when the number of readings of the same quantity is very large. The arithmetic mean of n measurements at a specific count of the variable x is given by the expression

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{n=1}^{n} x_n}{n}$$

where $\bar{x}$ = Arithmetic mean

x_n = nth reading taken

n = total number of readings

1.8.2 Deviation from the Mean

This 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,

The deviations from the mean can be expressed as

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

The deviation may be positive or negative. The algebraic sum of all the deviations must be zero.

Example 1.4 *For the following given data, calculate*
(i) Arithmetic mean; (ii) Deviation of each value; (iii) Algebraic sum of the deviations
Given

$$x_1 = 49.7;\ x_2 = 50.1;\ x_3 = 50.2;\ x_4 = 49.6;\ x_5 = 49.7$$

Solution

(i) The arithmetic mean is calculated as follows

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5}$$

$$= \frac{49.7 + 50.1 + 50.2 + 49.6 + 49.7}{5} = 49.86$$

(ii) The deviations from each value are given by

$$d_1 = x_1 - \bar{x} = 49.7 - 49.86 = -0.16$$
$$d_2 = x_2 - \bar{x} = 50.1 - 49.86 = +0.24$$
$$d_3 = x_3 - \bar{x} = 50.2 - 49.86 = +0.34$$
$$d_4 = x_4 - \bar{x} = 49.6 - 49.86 = -0.26$$
$$d_5 = x_5 - \bar{x} = 49.7 - 49.86 = -0.16$$

(iii) The algebraic sum of the deviation is

$$d_{\text{total}} = -0.16 + 0.24 + 0.34 - 0.26 - 0.16$$
$$= +0.58 - 0.58 = 0$$

1.8.3 Average Deviations

The average deviation is an indication of the precision of the instrument used in measurement. Average deviation is defined as the sum of the absolute values of the deviation divided by the number of readings. The absolute value of the deviation is the value without respect to the sign.

Average deviation may be expressed as

$$D_{\text{av}} = \frac{|d_1| + |d_2| + |d_3| + \cdots + |d_n|}{n}$$

or

$$D_{\text{av}} = \frac{\sum |d_n|}{n}$$

where D_{av} = average deviation

$|d_1|, |d_2|, \ldots, |d_n|$ = Absolute value of deviations

and n = total number of readings

Highly precise instruments yield a low average deviation between readings.

Example 1.5 *Calculate the average deviation for the data given in Example 1.4.*

Solution The average deviation is calculated as follows

$$D_{av} = \frac{|d_1| + |d_2| + |d_3| + \cdots + |d_n|}{n}$$

$$= \frac{|-0.16| + |0.24| + |0.34| + |-0.26| + |-0.16|}{5}$$

$$= \frac{1.16}{5} = 0.232$$

Therefore, the average deviation = 0.232.

1.8.4 Standard Deviation

The standard deviation of an infinite number of data is the Square root of the sum of all the individual deviations squared, divided by the number of readings. It may be expressed as

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

where σ = standard deviation

The standard deviation is also known as root mean square deviation, and is the most important factor in the statistical analysis of measurement data. Reduction in this quantity effectively means improvement in measurement.

For small readings ($n < 30$), the denominator is frequently expressed as $(n - 1)$ to obtain a more accurate value for the standard deviation.

Example 1.6 *Calculate the standard deviation for the data given in Example 1.4.*

Solution

$$\text{Standard deviation } = \sqrt{\frac{d_1^2 + d_2^2 + d_3^2 + \cdots + d_n^2}{n-1}}$$

$$\sigma = \sqrt{\frac{(-0.16)^2 + (0.24)^2 + (0.34)^2 + (-0.26)^2 + (-0.16)^2}{5-1}}$$

$$\sigma = \sqrt{\frac{0.0256 + 0.0576 + 0.1156 + 0.0676 + 0.0256}{4}}$$

$$\sigma = \sqrt{\frac{0.292}{4}} = \sqrt{0.073} = 0.27$$

Therefore, the standard deviation is 0.27.

1.8.5 Limiting Errors

Most manufacturers of measuring instruments specify accuracy within a certain % of a full scale reading. For example, the manufacturer of a certain voltmeter may specify the instrument to be accurate within ± 2% with full scale deflection. This specification is called the limiting error. This means that a full scale deflection reading is guaranteed to be within the limits of 2% of a perfectly accurate reading; however, with a reading less than full scale, the limiting error increases.

Example 1.7 *A 600 V voltmeter is specified to be accurate within ± 2% at full scale. Calculate the limiting error when the instrument is used to measure a voltage of 250 V.*

Solution The magnitude of the limiting error is 0.02 × 600 = 12 V.

Therefore, the limiting error for 250 V is 12/250 × 100 = 4.8%

Example 1.8 (a) *A 500 mA voltmeter is specified to be accurate with ±2%. Calculate the limiting error when instrument is used to measure 300 mA.*

Solution Given accuracy of 0.02 = ±2%

Step 1: The magnitude of limiting error is = 500 mA × 0.02 = 10 mA

Step 2: Therefore the limiting error at 300 mA = $\dfrac{10\text{ mA}}{300\text{ mA}} \times 100\% = 3.33\%$

Example 1.8 (b) *A voltmeter reading 70 V on its 100 V range and an ammeter reading 80 mA on its 150 mA range are used to determine the power dissipated in a resistor. Both these instruments are guaranteed to be accurate within ±1.5% at full scale deflection. Determine the limiting error of the power.*

Solution The magnitude of the limiting error for the voltmeter is

$$0.015 \times 100 = 1.5\text{ V}$$

The limiting error at 70 V is

$$\frac{1.5}{70} \times 100 = 2.143\ \%$$

The magnitude of limiting error of the ammeter is

$$0.015 \times 150\text{ mA} = 2.25\text{ mA}$$

The limiting error at 80 mA is

$$\frac{2.25\text{ mA}}{80\text{ mA}} \times 100 = 2.813\ \%$$

Therefore, the limiting error for the power calculation is the sum of the individual limiting errors involved.

Therefore, limiting error = 2.143 % + 2.813 % = 4.956 %

STANDARD 1.9

A standard is a physical representation of a unit of measurement. A known accurate measure of physical quantity is termed as a standard. These standards are used to determine the values of other physical quantities by the comparison method.

In fact, a unit is realized by reference to a material standard or to natural phenomena, including physical and atomic constants. For example, the fundamental unit of length in the International system (SI) is the metre, defined as the distance between two fine lines engraved on gold plugs near the ends of a platinum-iridium alloy at 0°C and mechanically supported in a prescribed manner.

Similarly, different standards have been developed for other units of measurement (including fundamental units as well as derived mechanical and electrical units). All these standards are preserved at the International Bureau of Weight and Measures at Sèvres, Paris.

Also, depending on the functions and applications, different types of "standards of measurement" are classified in categories (i) international, (ii) primary, (iii) secondary, and (iv) working standards.

1.9.1 International Standards

International standards are defined by International agreement. They are periodically evaluated and checked by absolute measurements in terms of fundamental units of Physics. They represent certain units of measurement to the closest possible accuracy attainable by the science and technology of measurement. These International standards are not available to ordinary users for measurements and calibrations.

International Ohms It is defined as the resistance offered by a column of mercury having a mass of 14.4521 gms, uniform cross-sectional area and length of 106.300 cm, to the flow of constant current at the melting point of ice.

International Amperes It is an unvarying current, which when passed through a solution of silver nitrate in water (prepared in accordance with stipulated specifications) deposits silver at the rate of 0.00111800 gm/s.

Absolute Units International units were replaced in 1948 by absolute units. These units are more accurate than International units, and differ slightly from them. For example,

1 International ohm = 1.00049 Absolute ohm

1 International ampere = 0.99985 Absolute ampere

1.9.2 Primary Standards

The principle function of primary standards is the calibration and verification of secondary standards. Primary standards are maintained at the National Standards Laboratories in different countries.

The primary standards are not available for use outside the National Laboratory. These primary standards are absolute standards of high accuracy that can be used as ultimate reference standards.

1.9.3 Secondary Standards

Secondary standards are basic reference standards used by measurement and calibration laboratories in industries. These secondary standards are maintained by the particular industry to which they belong. Each industry has its own secondary standard. Each laboratory periodically sends its secondary standard to the National standards laboratory for calibration and comparison against the primary standard. After comparison and calibration, the National Standards Laboratory returns the Secondary standards to the particular industrial laboratory with a certification of measuring accuracy in terms of a primary standard.

1.9.4 Working Standards

Working standards are the principal tools of a measurement laboratory. These standards are used to check and calibrate laboratory instrument for accuracy and performance. For example, manufacturers of electronic components such as capacitors, resistors, etc. use a standard called a working standard for checking the component values being manufactured, e.g. a standard resistor for checking of resistance value manufactured.

ELECTRICAL STANDARDS 1.10

All electrical measurements are based on the fundamental quantities I, R and V. A systematic measurement depends upon the definitions of these quantities. These quantities are related to each other by the Ohm's law, $V = I.R$. It is therefore sufficient to define only two parameters to obtain the definitions of the third. Hence, in electrical measurements, it is possible to assign values of the remaining standard, by defining units of other two standards. Standards of emf and resistance are, therefore, usually maintained at the National Laboratory. The base values of other standards are defined from these two standards. The electrical standards are

(a) Absolute Ampere (b) Voltage Standard (c) Resistance Standard

1.10.1 Absolute Ampere

The International System of Units (SI) defines the Ampere, that is, the fundamental unit of electric current, as the constant current which if maintained in two straight parallel conductors of infinite length placed one metre apart in vacuum, will produce between these conductors a force equal to 2×10^7 newton per metre length. These measurements were not proper and were very crude. Hence, it was required to produce a more practical, accurate and reproducible standard for the National Laboratory.

Hence, by international agreement, the value of international ampere as discussed in the previous topic, was then based on the electrolytic deposition of silver from a silver nitrate solution. In this method, difficulties were encountered in determining the exact measurement of the deposited silver and slight differences existed between the measurements made independently by various National Standard laboratories.

The International Ampere was then replaced by the Absolute Ampere. This Absolute Ampere was determined by means of a current balance, which measures the force exerted between two current-carrying coils. This technique of force measurement was further improved to a value of ampere which is much superior to the early measurement (the relationship between force and the current which produces the force, can be calculated from the fundamental electromagnetic theory concepts).

The Absolute Ampere is now the fundamental unit of electric current in the SI system and is universally accepted by international agreements.

Voltage (V), current (I) and resistance (R) are related by Ohm's law $V = I.R$. If any of the two quantities is defined, the third can be easily known. In order to define the Ampere with high precision over long periods of time, the standard voltage cell and the standard resistor are used.

1.10.2 Voltage Standards

As described before, if two parameters of Ohm's law are known, the third can be easily derived. Standard voltage cell is used as one of the parameter.

The *standard voltage* called the *saturated standard cell* or *standard cell* was based on the principle of electrochemical cell for many years. But the standard cell had a drawback that it suffered from temperature dependence. This voltage was a function of a chemical reaction and was not directly related to any other physical constants. Hence, a new standard for volt was developed. This standard used a thin film junction, which is cooled to nearly absolute zero and irradiated with microwave energy, a voltage is developed across the junction. This voltage is related to the irradiating frequency by the relation $V = (h.f)/2e$ where h is the Planck's constant (6.63×10^{-34} J-s), 'e' is the charge of an electron (1.602×10^{-19}C) and 'f' is the frequency of the microwave irradiation. Since 'f', the irradiating frequency is the only variable in the equation, hence the standard volt is related to the standard frequency (or time).

The accuracy of the standard volt including all system inaccuracies approximately is one part in 10^9, when the microwave irradiating frequency is locked to an atomic clock or to a broadcast frequency standard.

Standard cells are used for transferring the volt from the standard. Based on the thin film junction to the secondary standards used for calibration, this device is called the normal or saturated Weston cell.

The saturated cell has mercury as the positive electrode and cadium amalgam (10% cadium). The electrolyte used is a solution of cadium sulphate. These electrodes with the electrolyte are placed in a H-shaped glass container as shown in Fig 1.1.

There are two types of Weston cells called the saturated cell and the unsaturated cell. In a saturated cell, the electrolyte used is saturated at all temperatures by the cadium sulphate crystals covering the electrodes.

In the unsaturated cell, the concentration of cadium sulphate is such that it produces saturation at 40°C. The unsaturated cell has a negligible temperature coefficient of voltage at normal room temperature.

The saturated cell has a voltage variation of approximately 40 μV/°C ,but is better reproducible and more stable than the unsaturated cell.

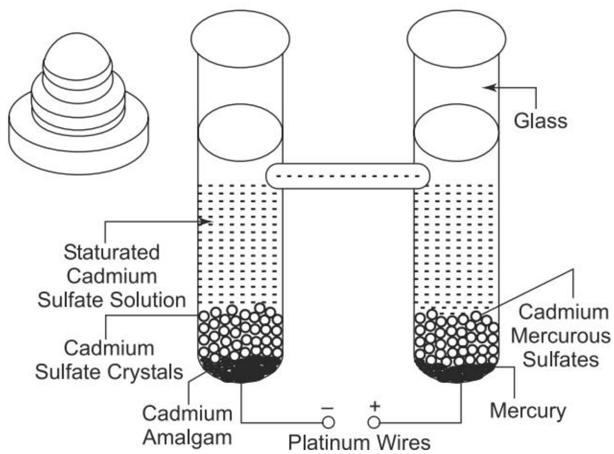


Fig 1.1 Voltage Standard

More rugged portable secondary and working standards are made of unsaturated Weston cells. These cells are very similar to the normal cell, but they do not require exact temperature control. The emf of an unsaturated cell lies in the range of 1.0180 V to 1.0200 V and the variation is less than 0.01%.

The internal resistance of Weston cells range from 500 to 800 ohms. The current drawn from these cells should therefore not exceed 100 μA.

Laboratory working standards have been developed based on the operation of Zener diodes as the voltage reference element, having accuracy of the same order as that of the standard cell. This instrument basically consists of Zener controlled voltage placed in a temperature controlled environment to improve it's long-term stability and having a precision output voltage. The temperature controlled oven is held within + 0.03°C over an ambient temperature range of 0 to 50°C giving an output stability in the order of 10 ppm/month. Zener controlled voltage sources are available in different ranges such as

(a) 0–1000 μV source with 1 μV resolution
(b) A 1.000 V reference for volt box potentiometric measurements
(c) A 1.018 V reference for saturated cell comparison

1.10.3 Resistance Standards

The absolute value of resistance is defined as ohms in the SI system of units. We know that the resistance R is given by $R = \rho \cdot l/A$ in terms of the length of wire (l), area of cross-section (A) of the wire and the resistivity of the wire (ρ) . Standard resistors are made of high resistivity conducting material with low temperature coefficient of resistance. Manganin, an alloy of copper, having a resistivity and whose temperature resistance relationship is almost constant, is used as the resistance wire. The construction of a resistance standard is as shown in Fig 1.2.

A coil of manganin wire as shown in Fig 1.2. is mounted on a double-walled sealed container to prevent the change in resistance due to humidity. The unit of resistance can be represented with precision values of a few parts in 10^7 over several years, with a set of four or five resistors of this type.

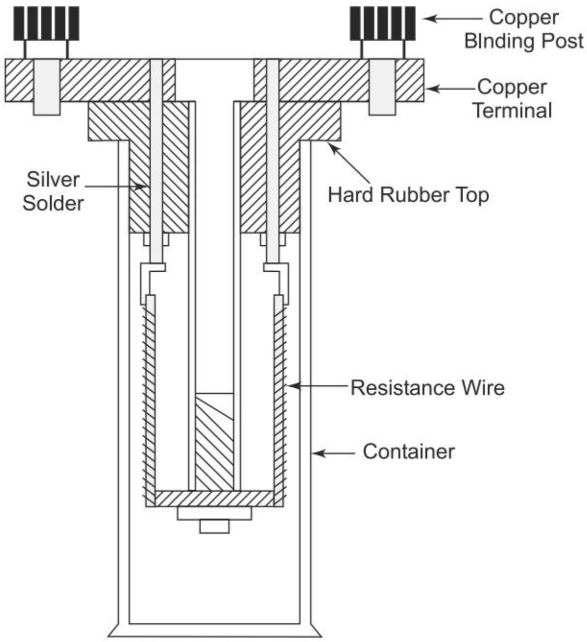


Fig 1.2 Resistance Standard

The secondary standard resistors are made of alloy of resistance wire such as manganin or Evan Ohm. The secondary standard or working standard are available in multiple of 10. These laboratory standards can also be referred to as transfer resistors. The resistance coil of the transfer resistance is supported between polyester film, in order to reduce stress on the wire and to improve it's stability. The coil is immersed in moisture-free oil and placed in a sealed container. The connections to the coils are silver soldered and terminals hooks are made of nickel-plated oxygen-free copper. These are checked for stability and temperature characteristics at it's rated power and operating.

Transfer resistors are used in industrial research and calibration laboratory. It is used in determining the value of the unknown resistance and ratio value. These resistors are also used as linear decade dividers resistors. These dividers are used in calibrating universal ratio sets and volt boxes.

ATOMIC FREQUENCY AND TIME STANDARDS 1.11

The measurement of time has two different aspects, civil and scientific. In most scientific work, it is desired to know how long an event lasts, or if dealing with an oscillator, it is desired to know its frequency of oscillation. Thus any time standard must be able to answer both the question "what time is it" and the two related questions "how long does it last" or "what is its frequency".

Any phenomena that repeats itself can be used as a measure of time, the measurement consisting of counting the repetitions. Of the many repetitive phenomena occurring in nature, the rotation of the earth on its axis which determines the length of the day, has been long used as a time standard. Time defined in terms of rotation of the earth is called *Universal time* (UT).

Time defined in terms of the earth's orbital motion is called *Ephemersis time* (ET). Both UT and ET are determined by astronomical observation. Since these astronomical observations extend over several weeks for UT and several years for ET, a good secondary terrestrial clock calibrated by astronomical observation is needed. A quartz crystal clock based on electrically sustained natural periodic vibrations of a quartz wafer serves as a secondary time standard. These clocks have a maximum error of 0.02 sec per year. One of the most common of time standards is the determination of frequency.

In the RF range, frequency comparisons to a quartz clock can be made electronically to a precision of atleast 1 part in 10^{10}.

To meet a better time standard, atomic clocks have been developed using periodic atomic vibrations as a standard. The transition between two energy levels, E_1 and E_2 of an atom is accompanied by the emission (or absorption) of radiation given by the following equation

$$v = \frac{E_2 - E_1}{h}$$

where v = frequency of emission and depends on the internal structure of an atom

h = Planck's constant = 6.636×10^{-34} J-sec.

Provided that the energy levels are not affected by the external conditions such as magnetic field etc.

Since frequency is the inverse of the time interval, time can be calibrated in terms of frequency.

The atomic clock is constructed on the above principle. The first atomic clock was based on the Cesium atom.

The International Committee of Weights and Measures defines the second in terms of the frequency of Cesium transitions, assigning a value of 9,192, 631,770 Hz to the hyperfine transitions of the Cesium atom unperturbed by external fields. If two Cesium clocks are operated at one precision and if there are no other sources of error, the clocks will differ by only 1s in 5000 years.

GRAPHICAL REPRESENTATION OF MEASUREMENTS AS A DISTRIBUTION 1.12

Suppose that a certain voltage is measured 51 times. The result which might be obtained are shown in Table 1.2.

Table 1.2

x Voltage (V)	Number of Occurrences (n)	x_n (v)	$d_n = x_n - \bar{x}$	$n\,\lvert d_n \rvert$	$(d_n)^2$	$n\,(d_n)^2$
1.01	1	1.01	– 0.04	0.04	16×10^{-4}	16×10^{-4}
1.02	3	3.06	– 0.03	0.09	9×10^{-4}	27×10^{-4}
1.03	6	6.18	– 0.02	0.12	4×10^{-4}	24×10^{-4}
1.04	8	8.32	– 0.01	0.08	1×10^{-4}	8×10^{-4}
1.05	10	10.50	0.00	0.00	0×10^{-4}	00×10^{-4}
1.06	7	7.42	+ 0.01	0.07	1×10^{-4}	7×10^{-4}
1.07	8	8.56	+ 0.02	0.16	4×10^{-4}	32×10^{-4}
1.08	4	4.32	+ 0.03	0.12	9×10^{-4}	36×10^{-4}
1.09	3	3.27	+ 0.04	0.12	16×10^{-4}	48×10^{-4}
1.10	0	0.00	+ 0.05	0.00	25×10^{-4}	00×10^{-4}
1.11	1	1.11	+ 0.06	0.06	36×10^{-4}	36×10^{-4}
	51 $= \Sigma_n$	53.75 $= \sum_{n=1}^{51} x_n$		0.86 $= \sum_{n=1}^{51} \lvert d_n \rvert$		234×10^{-4} $\sum_{n=1}^{51} (d_n)^2$

$$\text{Average } \bar{x} = \frac{\sum_{n=1}^{51} x_n}{n} = \frac{53.75}{51} = 1.054 \text{ V}$$

$$\text{Average deviation } D_{\text{av}} = \frac{\sum_{n=1}^{51} \lvert d_n \rvert}{n} = \frac{0.86}{51} = 0.0168 \text{ V}$$

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum_{n=1}^{51} (d_n)^2}{n}} = \sqrt{\frac{234 \times 10^{-4}}{51}} = \sqrt{4.588 \times 10^{-4}} \text{ V}$$

$$= 2.142 \times 10^{-2} \text{ V}$$

The first column shows the various measured values and the second column, the number of times each reading has occurred. For example, in the fourth row, the measured value is 1.04 V and the next column indicates that this reading is obtained 8 times.

The data given in Table 1.2 may be represented graphically as shown in Fig. 1.3.

We imagine the range of values of x to be divided into equal intervals dx, and plot the number of values of x lying in the interval versus the average value of x

within that interval. Hence the eight measurements of 1.04 V might be thought of lying in an 0.01 V interval centred upon 1.04 V, i.e. between 1.035 V to 1.045 V on the horizontal scale. Since with a small number (such as 51), these points do not lie on a smooth curve, it is conventional to represent such a plot by a histogram consisting of series of horizontal lines of length *dx* centred upon the individual points. The ends of adjacent horizontal lines being connected by vertical lines of appropriate length.

If another 51 measurements are taken and plotted we would, in general get a graph which does not coincide with the previous one. The graph plotted is called a Gauss error or Gaussian graph, shown in Fig. 1.3.

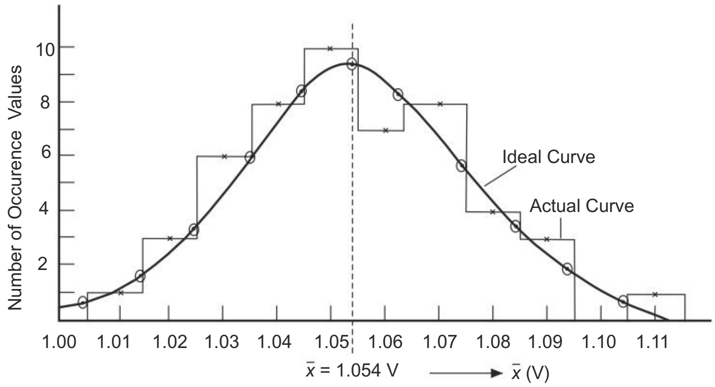


Fig. 1.3 Gaussian graph

Review Questions

1. What do you understand by static characteristics?
2. List different static characteristics.
3. Define the terms: instrument, accuracy, precision and errors.
4. Define the terms: resolution, sensitivity and expected value.
5. Discuss the difference between accuracy and precision of a measurement
6. List different types of errors.
7. Explain gross error in details. How can it be minimized?
8. Explain systematic error in detail. How can it be minimized?
9. Explain random error in detail.
10. A person using an ohmmeter reads the measured value as 470 Ω , when the actual value is 47 Ω. What kind of error does this represent?
11. What are the causes of environment errors?
12. How are instrumental errors different from gross errors? Explain.
13. Define absolute errors.
14. How is accuracy expressed?
15. What are the different types of errors that occur during measurement? Explain each.
16. What do you understand by dynamic characteristics of an instrument?
17. Define speed of response and fidelity.
18. Differentiate between lag and dynamic error.
19. What are limiting errors? What is the significance of limiting errors?
20. Define the following terms:

(i) Average value (ii) Arithmetic mean (iii) Deviation (iv) Standard deviation

21. What do you mean by a standard? What is the significance of standard?
22. What are international standards? List various international standards.
23. Define primary and secondary standards?
24. What are primary standards? Where are they used?
25. What are secondary standards? Where are they used?
26. What do you understand by a working standard?
27. State the difference between secondary and working standards.
28. Explain in brief atomic frequency and time standards.
29. How is time defined?
30. What do you understand by electrical standard?
31. List different types of electrical standards.

Multiple Choice Questions

1. The closeness of values indicated by an instrument to the actual value is defined as
   (a) repeatability (b) reliability
   (c) uncertainty (d) accuracy.
2. Precision is defined as
   (a) repeatability (b) reliability
   (c) uncertainty (d) accuracy
3. The ratio of change in output to the change in the input is called
   (a) precision (b) resolution
   (c) sensitivity (d) repeatability
4. The deviation of the measured value to the desired value is defined as
   (a) error (b) repeatability
   (c) hystersis (d) resolution
5. Improper setting of range of a multimeter leads to an error called
   (a) random error
   (b) limiting error
   (c) instrumental error
   (d) observational error
6. Errors that occur even when all the gross and systematic errors are taken care of are called
   (a) environmental errors
   (b) instrumental errors
   (c) limiting errors
   (d) random errors.
7. A means of reducing environmental errors is the regulation of ambient
   (a) noise (b) temperature
   (c) light (d) mains voltage
8. The ability of an instrument to respond to the weakest signal is defined as
   (a) sensitivity (b) repeatability
   (c) resolution (d) precision.
9. The difference between the expected value of the variable and the measured variable is termed
   (a) absolute error
   (b) random error
   (c) instrumental error
   (d) gross error
10. Accuracy is expressed as
    (a) relative accuracy
    (b) % accuracy
    (c) error
    (d) % error
11. Error is expressed as
    (a) absolute error (b) relative error
    (c) % error (d) % accuracy
12. Gross errors occurs due to
    (a) human error
    (b) instrumental error
    (c) environmental error
    (d) random error
13. Static errors are caused due to
    (a) measuring devices
    (b) human error
    (c) environmental error
    (d) observational error

14. Dynamic errors are caused by
    (a) instrument not responding fast
    (b) human error
    (c) environmental error
    (d) observational error
15. Limiting errors are
    (a) manufacturer's specifications of accuracy
    (b) manufacturer's specifications of instrumental error
    (c) environmental errors
    (d) random errors

Practice Problems

1. The current through a resistor is 3.0 A, but measurement gives a value of 2.9 A. Calculate the absolute error and % error of the measurement.
2. The current through a resistor is 2.5 A, but measurement yields a value of 2.45 A. Calculate the absolute error and % error of the measurement.
3. The value of a resistor is 4.7 k-ohms, while measurement yields a value of 4.63 K-ohms.
   Calculate (a) the relative accuracy, and (b) % accuracy.
4. The value of a resistor is 5.6 K-ohms, while measurement reads a value of 5.54 K-ohms. Calculate (a) the relative accuracy, and (b) % accuracy.
5. The output voltage of an amplifier was measured at eight different intervals using the same digital voltmeter with the following results: 20.00, 19.80, 19.85, 20.05, 20.10, 19.90, 20.25, 19.95.
   Which is the most precise measurement?
6. A 270 Ω ± 10% resistance is connected to a power supply source operating at 300 V dc.
   What range of current would flow if the resistor varied over the range of ± 10% of its expected value? What is the range of error in the current?
7. A voltmeter is accurate to 98% of its full scale reading.
   (i) If a voltmeter reads 200 V on 500 V range, what is the absolute error?
   (ii) What is the percentage error reading of Part (i)?
8. The expected value of voltage across a resistor is 100 V. However, the voltmeter reads a value of 99 V. Calculate (a) absolute error, (b) % error, (c) relative error, and (d) % accuracy.

Further Reading

1. Barry Jones, *Instrumentation Measurements and Feedback*.
2. Larry D. Jones and A. Foster Chin, *Electronic Instruments and Measurement*, John Wiley and Sons, 1987.
3. Yardley Beers, *Theory of Errors*, 1967.
4. Resnick and Halliday, *Physics*, Wiley Eastern, 1987.

Chapter 2

Indicators and Display Devices

INTRODUCTION 2.1

Analogue ammeters and voltmeters are classified together, since there is no basic difference in their operating principles. The action of all ammeters and voltmeters, except those of the electrostatic variety, depends upon a deflecting torque produced by an electric current. In an ammeter this torque is produced by the current to be measured, or by a definite fraction of it. In a voltmeter it is produced by a current that is proportional to the voltage to be measured. Hence both voltmeters and ammeters are essentially current measuring devices.

The essential requirements of a measuring instrument are (a) that its introduction into the circuit where measurements are to be made, should not alter the circuit conditions, and (b) the power consumed by it be small.

2.1.1 Types of Instrument

The following types of instrument are mainly used as ammeters and voltmeters.

1. PMMC	2. Moving Iron
3. Electrodynamometer	4. Hot wire
5. Thermocouple	6. Induction type
7. Electrostatic	8. Rectifier

Of these, the PMMC type can be used for dc measurements only, and the induction type for ac measurements only. The other types can be used for both.

The moving coil and moving iron types depend upon the magnitude effect of current. The latter is the most commonly used form of indicating instrument, as well as the cheapest. It can be used for both ac and dc measurements and is very accurate, if properly designed.

The PMMC instrument is the most accurate type for dc measurement. Instrument of this type are frequently constructed to have substandard accuracy.

The calibration of the electrodynamometer type of instrument is the same for ac and dc. The same situation prevails for thermal instruments. These are particularly suitable for ac measurements, since their deflection depends directly upon the heating effect of the ac, i.e. upon the rms value of the current. Their readings are therefore independent of the frequency.

Electrostatic instruments used as voltmeters have the advantage that their power consumption is exceedingly small. They can be made to cover a large range of voltage and can be constructed to have sub-standard accuracy.

The induction principle is most generally used for Watt-hour meters. This principle is not preferred for use in ammeters and voltmeters because of the comparatively high cost and inaccuracy of the instrument.

BASIC METER MOVEMENT 2.2

The action of the most commonly dc meter is based on the fundamental principle of the motor. The motor action is produced by the flow of a small current through a moving coil, which is positioned in the field of a permanent magnet. This basic moving coil system is often called the D'Arsonval galvanometer.

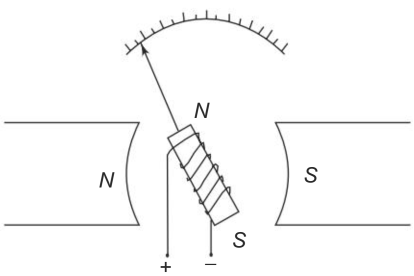


Fig. 2.1 D'Arsonval principle

The D'Arsonval movement shown in Fig. 2.1 employs a spring-loaded coil through which the measured current flows. The coil (rotor) is in a nearly homogeneous field of a permanent magnet and moves in a rotary fashion. The amount of rotation is proportional to the amount of current flowing through the coil. A pointer attached to the coil indicates the position of the coil on a scale calibrated in terms of current or voltage. It responds to dc current only, and has an almost linear calibration. The magnetic shunt that varies the field strength is used for calibration.

2.2.1 Permanent Magnetic Moving Coil Movement

In this instrument, we have a coil suspended in the magnetic field of a permanent magnet in the shape of a horse-shoe. The coil is suspended so that it can rotate freely in the magnetic field. When current flows in the coil, the developed (electromagnetic) torque causes the coil to rotate. The electromagnetic (EM) torque is counterbalanced by a mechanical torque of control springs attached to the movable coil. The balance of torques, and therefore the angular position of the movable coil is indicated by a pointer against a fixed reference called a scale. The equation for the developed torque, derived from the basic law for electromagnetic torque is

$$\tau = B \times A \times I \times N$$

where τ = torque, Newton-meter
B = flux density in the air gap, Wb/m^2
A = effective coil area (m^2)
N = number of turns of wire of the coil
I = current in the movable coil (amperes)

The equation shows that the developed torque is proportional to the flux density of the field in which the coil rotates, the current coil constants (area and number of turns). Since both flux density and coil constants are fixed for a given instrument, the developed torque is a direct indication of the current in the coil. The pointer deflection can therefore be used to measure current.

Example 2.1 (a) *A moving coil instrument has the following data.*
Number of turns = 100
Width of the coil = 20 mm
Depth of the coil = 30 mm
Flux density in the gap = 0.1 Wb/m²
Calculate the deflecting torque when carrying a current of 10 mA. Also calculate the deflection, if the control spring constant is 2×10^{-6} Nm/degree.

Solution The deflecting torque is given by

$$\begin{aligned}\tau_d &= B \times A \times N \times I\\ &= 0.1 \times 30 \times 10^{-3} \times 20 \times 10^{-3} \times 100 \times 10 \times 10^{-3}\\ &= 600 \times 1000 \times 0.1 \times 10^{-9}\\ &= 600 \times 1000 \times 10^{-10}\\ &= 60 \times 10^{-6}\ \text{Nm}\end{aligned}$$

The spring control provides a restoring torque, i.e. $\tau_c = K\theta$,
where K is the spring constant
As deflecting torque = restoring torque

$$\therefore \qquad \tau_c = 6 \times 10^{-5}\ \text{Nm} = K\theta, \qquad \therefore \theta = \frac{6 \times 10^{-5}}{2 \times 10^{-6}} = 3 \times 10 = 30°$$

Therefore, the deflection is 30°.

Example 2.1 (b) *A moving coil instrument has the following data*
No. of turns = 100
Width of the coil = 20 mm
Depth of the coil = 30 mm
Flux density in the gap = 0.1 Wb/m²
The deflection torque = 30×10^{-6} Nm
Calculate the current through the moving coil.

Solution The deflecting torque is given by

$$\tau_d = B \times A \times N \times I$$

Therefore $$30 \times 10^{-6} = 0.1 \times 30 \times 10^{-3} \times 20 \times 10^{-3} \times 100 \times I$$

$$I = \frac{30 \times 10^{-6}}{0.1 \times 30 \times 10^{-3} \times 20 \times 10^{-3} \times 100}$$

$$I = \frac{30 \times 10^{-6}}{0.1 \times 600 \times 10^{-6} \times 100} = 5\ \text{mA}$$

2.2.2 Practical PMMC Movement

The basic PMMC movement (also called a D'Arsonval movement) offers the largest magnet in a given space, in the form of a horse-shoe, and is used when a large flux is required in the air gap. The D'Arsonval movement is based on the principle of a moving electromagnetic coil pivoted in a uniform air gap between the poles of a large fixed permanent magnet. This principle is illustrated in Fig. 2.1 With the polarities as shown, there is a repelling force between like poles, which exerts a torque on the pivoted coil. The torque is proportional to the magnitude of current being measured. This D'Arsonval movement provides an instrument with very low power consumption and low current required for full scale deflection (fsd).

Figure 2.2 shows a permanent horse-shoe magnet with soft iron pole pieces attached to it. Between the pole pieces is a cylinder of soft iron which serves to provide a uniform magnetic field in the air gap between the pole pieces and the cylindrical core.

The coil is wound on a light metal frame and is mounted so that it can rotate freely in the air gap. The pointer attached to the coil moves over a graduated scale and indicates the angular deflection of the coil, which is proportional to the current flowing through it.

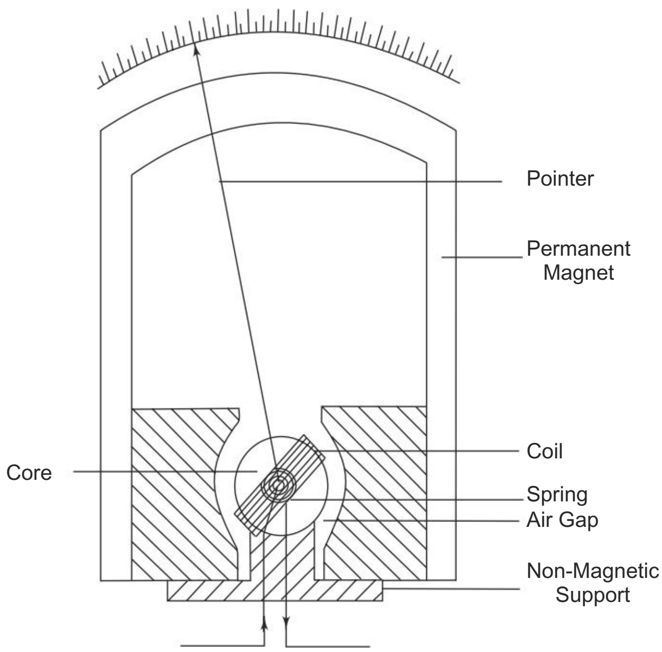


Fig. 2.2 Modern D'Arsonval movement

The Y-shaped member shown in Fig. 2.3 is the zero adjust control, and is connected to the fixed end of the front control spring. An eccentric pin through the instrument case engages the Y-shaped member so that the zero position of the pointer can be adjusted from outside. The calibrated force opposing the moving torque is provided by two phosphor-bronze conductive springs, normally equal in strength. (This also provides the necessary torque to bring the pointer back to its original position after the measurement is over.)

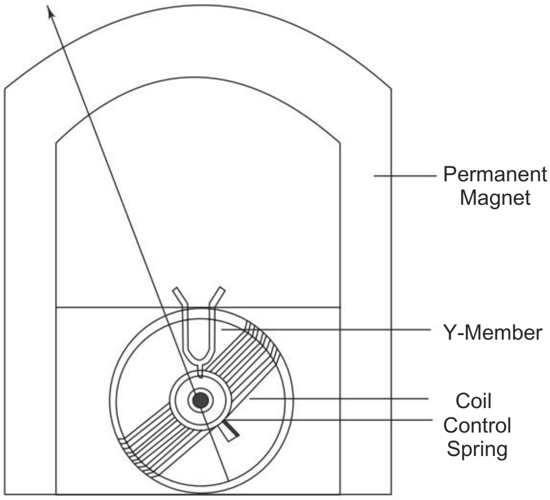


Fig. 2.3 Simplified diagram of a PMMC movement showing the Y-member

The accuracy of the instrument can be maintained by keeping spring performance constant. The entire moving system is statically balanced at all positions by three (counterweights) balance weights. The pointer, springs, and pivots are fixed to the coil assembly by means of pivot bases and the entire movable coil element is supported by jewel bearings.

PMMC instruments are constructed to produce as little viscious damping as possible and the required degree of damping is added.

In Fig. 2.4, Curve 2 is the underdamped case; the pointer attached to the movable coil oscillates back and forth several times before coming to rest. As in curve 1, the overdamped case, the pointer tends to approach the steady state position in a sluggish manner. In Curve 3, the critically damped case, the pointer moves up to its steady state position without oscillations. Critical damping is the ideal behaviour for a PMMC movement.

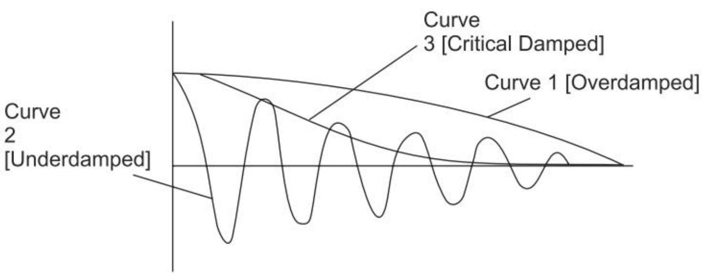


Fig. 2.4 Degree of damping

In practice, however, the instrument is usually slightly underdamped, causing the pointer to overshoot a little before coming to rest.

The various methods of damping are as follows.

One of the simplest methods is to attach an aluminium vane to the shaft of the moving coil. As the coil rotates, the vane moves in an air chamber, the amount of clearance between the chamber walls and the air vane effectively controls the degree of damping.

Some instruments use the principle of electromagnetic damping (Lenz's law), where the movable coil is wound on a light aluminium frame. The rotation of the coil in the magnetic field sets up a circulating current in the conductive frame, causing a retarding torque that opposes the motion of the coil.

A PMMC movement may also be damped by a resistor across the coil. When the coil rotates in the magnetic field, a voltage is generated in the coil, which circulates a current through it and the external resistance. This produces an opposing or retarding torque that damps the motion. In any galvanometer, the value of the external resistance that produces critical damping can be found. This resistance is called critically damping external resistance (CDRX). Most voltmeter coils are wound on metal frames to provide Electro-Magnetic damping. The metal frames constitute a short-circuit turn in a magnetic field.

Ammeters coils, are however wound in a non-conductive frame, because the coil turns are effectively shorted by the ammeter shunt. The coil itself provides the EM damping.

If low frequency alternating current is applied to the movable coil, the deflection of the pointer would be upscale for half the cycle of the input waveform and downscale (in the opposite direction) for the next half. At power line frequency (50 Hz) and above, the pointer cannot follow the rapid variations in direction and quivers slightly around the zero mark, seeking the average value of the ac (which equals zero). The PMMC instrument is therefore unsuitable for ac measurements, unless the current is rectified before reaching the coil.

Practical coil areas generally range from 0.5 – 2.5 cm^2.

The flux density for modern instruments usually ranges from 1500 – 5000 Wb/cm^2.

The power requirements of D'Arsonval movements are quite small, typically from 25 – 200 μW.

The accuracy of the instrument is generally of the order of 2 – 5% of full scale deflection.

The permanent magnet is made up of Alnico material.

Scale markings of basic dc PMMC instruments are usually linearly spaced, because the torque (and hence the pointer deflection) is directly proportional to the coil current. The basic PMMC instrument is therefore a linear-reading device.

The advantages and disadvantages of PMMC are as follows.

Advantages

1. They can be modified with the help of shunts and resistance to cover a wide range of currents and voltages.
2. They display no hysteresis.
3. Since operating fields of such instruments are very strong, they are not significantly affected by stray magnetic fields.

Disadvantages

1. Some errors may set in due to ageing of control springs and the permanent magnet.
2. Friction due to jewel-pivot suspension.

TAUT BAND INSTRUMENT 2.3

The taut band movement utilises the same principle as the D'Arsonval movable coil and fixed magnet. The primary difference between the two is the method of mounting the movable coil.

The taut band movement has the advantage of eliminating the friction caused by a jewel-pivot suspension. The meter has a coil mounted in a cradle and surrounded by a ring-bar magnet, as shown in Fig. 2.5. The cradle is secured to a support bracket, which in turn is suspended between two steel taut bands (ribbon), i.e. the movable coil is suspended by means of two taut torsion ribbons. The ribbons are placed under sufficient tension to eliminate any sag. This tension is provided by the tension spring, so that the instrument can be used in any position.

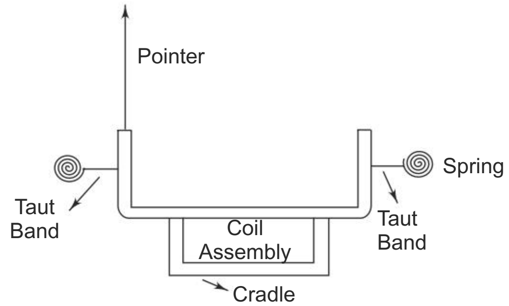


Fig. 2.5 (a) Taut band instrument (Side view)

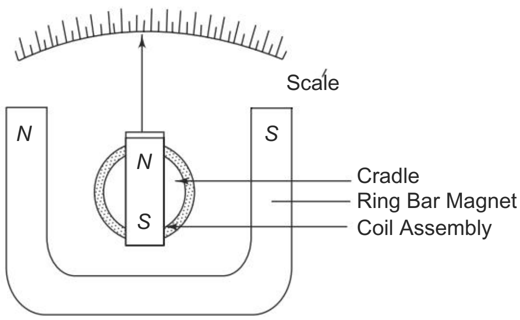


Fig. 2.5 (b) Taut band instrument (Top view)

The current to be measured is passed through the coil, thereby energising it. The interaction of the magnetic fields deflects the cradle to one side and moves the pointer along the scale.

The movement of the cradle exerts a twisting force on the steel bands. These twisted bands supply the torque to return the pointer to zero, when no current flows. There are no bearings, and there is a constant level of sensitivity throughout the range of movement.

Taut band instruments have a higher sensitivity than those using pivots and jewels. In addition taut band instruments are relatively insensitive to shock and temperature and are capable of withstanding greater overloads than PMMC or other types.

ELECTRODYNAMOMETER 2.4

The D'Arsonval movement responds to the average or dc value of the current flowing through the coil.

If ac current is sought to be measured, the current would flow through the coil with positive and negative half cycles, and hence the driving torque would be positive in one direction and negative in the other. If the frequency of the ac is very low, the pointer would swing back and forth around the zero point on the meter scale.

At higher frequencies, the inertia of the coil is so great that the pointer does not follow the rapid variations of the driving torque and vibrates around the zero mark.

Therefore, to measure ac on a D'Arsonval movement, a rectifier has to be used to produce a unidirectional torque. This rectifier converts ac into dc and the rectified current deflects the coil. Another method is to use the heating effect of ac current to produce an indication of its magnitude. This is done using an electrodynamometer (EDM).

An electrodynamometer is often used in accurate voltmeter and ammeters not only at power line frequency but also at low AF range. The electrodynamometer can be used by slightly modifying the PMMC movement. It may also serve as a transfer instrument, because it can be calibrated on dc and then used directly on ac thereby equating ac and dc measurements of voltage and current directly.

A movable coil is used to provide the magnetic field in an electrodynamometer, instead of a permanent magnet, as in the D'Arsonval movement. This movable coil rotates within the magnetic field. The EDM uses the current under measurement to produce the required field flux. A fixed coil, split into two equal halves provides the magnetic field in which the movable coil rotates, as shown in Fig. 2.6 (a). The coil halves are connected in series with the moving coil and are fed by the current being measured. The fixed coils are spaced far apart to allow passage for the shaft of the movable coil. The movable coil carries a pointer, which is balanced by counterweights. Its rotation is controlled by springs, similar to those in a D'Arsonval movement.

The complete assembly is surrounded by a laminated shield to protect the instrument from stray magnetic field which may affect its operation.

Damping is provided by aluminium air vanes moving in a sector shaped chamber. (The entire movement is very solid and rigidly constructed in order to keep its mechanical dimensions stable, and calibration intact.)

The operation of the instrument may be understood from the expression for the torque developed by a coil suspended in a magnetic field, i.e.

$$\tau = B \times A \times N \times I$$

indicating that the torque which deflects the movable coil is directly proportional to the coil constants (A and N), the strength of the magnetic field in which the coil moves (B), and the current (I) flowing through the coil.

In an EDM the flux density (B) depends on the current through the fixed coil and is therefore proportional to the deflection current (I). Since the coil constants are fixed quantities for any given meter, the developed torque becomes a function of the current squared (I^2).

If the EDM is used for dc measurement, the square law can be noticed by the crowding of the scale markings at low current values, progressively spreading at higher current values.

For ac measurement, the developed torque at any instant is proportional to the instantaneous current squared (i^2). The instantaneous values of i^2 are always positive and torque pulsations are therefore produced.

The meter movement, however, cannot follow rapid variations of the torque and take up a position in which the average torque is balanced by the torque of the control springs. The meter deflection is therefore a function of the mean of the squared current. The scale of the EDM is usually calibrated in terms of the square root of the average current squared, and therefore reads the effective or rms value of the ac.

The transfer properties of the EDM become apparent when we compare the effective value of the alternating current and the direct current in terms of their heating effect, or transfer of power.

(If the EDM is calibrated with a direct current of 500 mA and a mark is placed on the scale to indicate this value, then that ac current which causes the pointer to deflect to the same mark on the scale must have an rms value of 500 mA.)

The EDM has the disadvantage of high power consumption, due to its construction. The current under measurement must not only pass through the movable coil, but also provide the necessary field flux to get a sufficiently strong magnetic field. Hence high mmf is required and the source must have a high current and power.

In spite of this high power consumption the magnetic field is still weaker than that of the D'Arsonval movement because there is no iron in the path, the entire flux path consisting of air.

The EDM can be used to measure ac or dc voltage or current, as shown in Figs. 2.6 (a) and (b).

Typical values of EDM flux density are in the range of approximately 60 gauss as compared to the high flux densities (1000 – 4000 guass) of a good D'Arsonval movement. The low flux density of the EDM affects the developed torque and therefore the sensitivity of the instrument.

The addition of a series multiplier converts the basic EDM into a voltmeter [Fig. 2.6 (b)] which can be used for ac and dc measurements. The sensitivity of the EDM voltmeter is low, approximately 10 – 30 Ω/V, compared to 20 kΩ/V of the D'Arsonval movement. It is however very accurate at power line frequency and can be considered as a secondary standard.

The basic EDM shown in Fig. 2.6 (a) can be converted into an ammeter (even without a shunt), because it is difficult to design a moving coil which can carry more than approximately 100 mA.

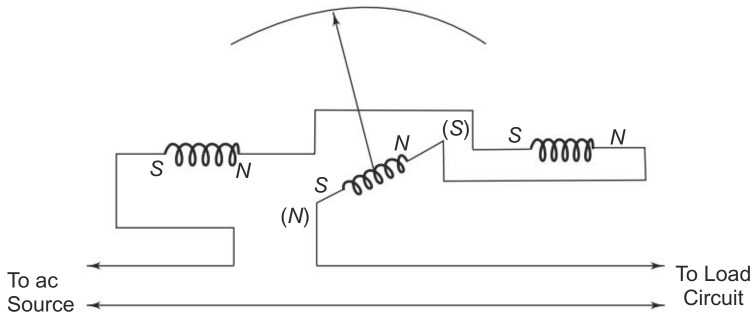


Fig. 2.6 (a) Basic EDM as an ammeter

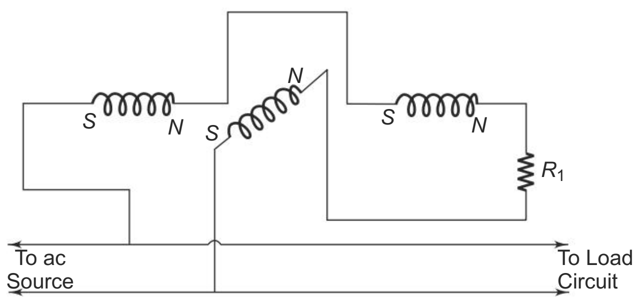


Fig. 2.6 (b) Basic EDM as a voltmeter

The EDM movement is extensively used to measure power, both dc and ac, for any waveform of voltage and current.

An EDM used as a voltmeter or ammeter has the fixed coils and movable coil connected in series, thereby reacting to I^2.

When an EDM is used as a single phase wattmeter, the coil arrangement is different, as shown in Fig. 2.7.

The fixed coils, shown in Fig. 2.7 as separate elements, are connected in series and carries the total line current. The movable coil located in the magnetic field of the fixed coils is connected in series with a current-limiting resistor across the power line, and carries a small current.

The deflection of the movable coil is proportional to the product of the instantaneous value of current in the movable coil and the total line current. The EDM wattmeter consumes some power for the maintenance of its magnetic field, but this is usually small compared to the load power.

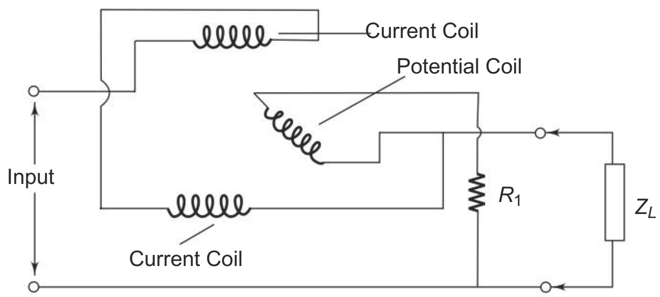


Fig. 2.7 EDM as a wattmeter

MOVING IRON TYPES INSTRUMENT 2.5

Moving iron instruments can be classified into attraction and repulsion types. Repulsion type instruments are the most commonly used.

Iron vane ammeters and voltmeters depend for their operations on the repulsion that exists between two like magnetic poles.

The movement consists of a stationary coil of many turns which carries the current to be measured. Two iron vanes are placed inside the coil. One vane is rigidly attached to the coil frame, while the other is connected to the instrument shaft which rotates freely. The current through the coil magnetises both the vanes with the same polarity, regardless of the instantaneous direction of current. The two magnetised vanes experience a repelling force, and since only one vane can move, its displacement is an indicator of the magnitude of the coil current. The repelling force is proportional to the current squared, but the effects of frequency and hysteresis tend to produce a pointer deflection that is not linear and that does not have a perfect square law relationship.

Figure 2.8 shows a radial vane repulsion instrument which is the most sensitive of the moving iron mechanisms and has the most linear scale. One of these like poles is created by the instrument coil and appears as an iron vane fixed in its position within the coil, as shown in Fig. 2.8. The other like pole is induced on the movable iron piece or vane, which is suspended in the induction field of the coil and to which the needle of the instrument is attached. Since the instrument is used on ac, the magnetic polarity of the coil changes with every half cycle and induces a corresponding amount of repulsion of the movable vane against the spring tension. The deflection of the instrument pointer is therefore always in the same direction, since there is always repulsion between the like poles of the fixed and the movable vane, even though the current in the inducing coil alternates.

The deflection of the pointer thus produced is effectively proportional to the actual current through the instrument. It can therefore be calibrated directly in amperes and volts.

The calibrations of a given instrument will however only be accurate for the ac frequency for which it is designed, because the impedance will be different at a new frequency.

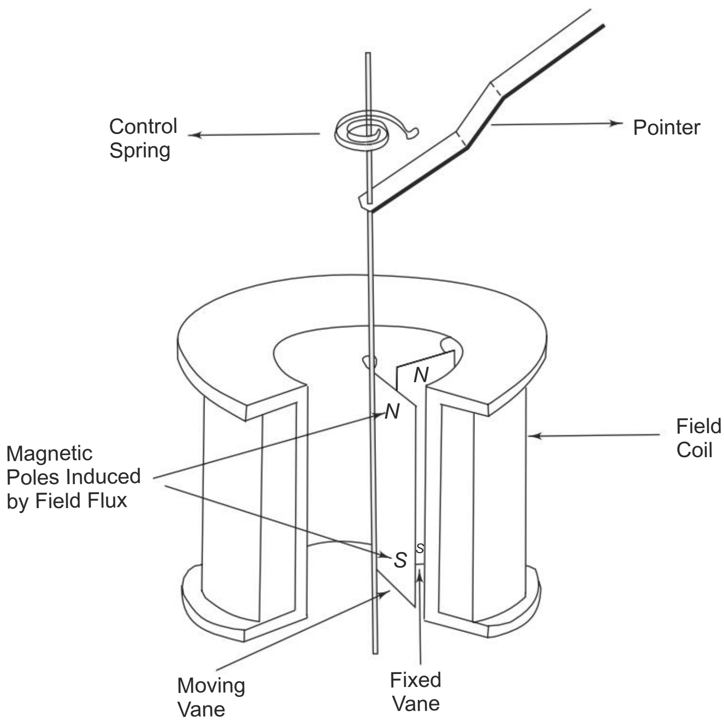


Fig. 2.8 Repulsion type AC meter (Radial vane type)

The moving coil or repulsion type of instrument is usually calibrated to read the effective value of amperes and volts, and is used primarily for rugged and inexpensive meters.

The iron vane or radial type is forced to turn within the fixed current carrying coil by the repulsion between like poles. The aluminium vanes, attached to the lower end of the pointer, acts as a damping vane, in its close fitting chamber, to bring the pointer quickly to rest.

CONCENTRIC VANE REPULSION TYPE (MOVING IRON TYPE) INSTRUMENT 2.6

A variation of the radial vane instrument is the concentric vane repulsion movement. The instrument has two concentric vanes.

One vane is rigidly attached to the coil frame while the other can rotate coaxially inside the stationary vane, as shown in Fig. 2.9. Both vanes are magnetised by the current in the coil to the same polarity, causing the vanes to slip laterally under repulsion. Because the moving vane is attached to a pivoted shaft, this repulsion results in a rotational force that is a function of the current in the coil. As in other mechanisms the final pointer position is a measure of the coil current. Since this movement, like all iron vane instruments, does not distinguish polarity, the concentric vane may be used on dc and ac, but it is most commonly used for the latter.

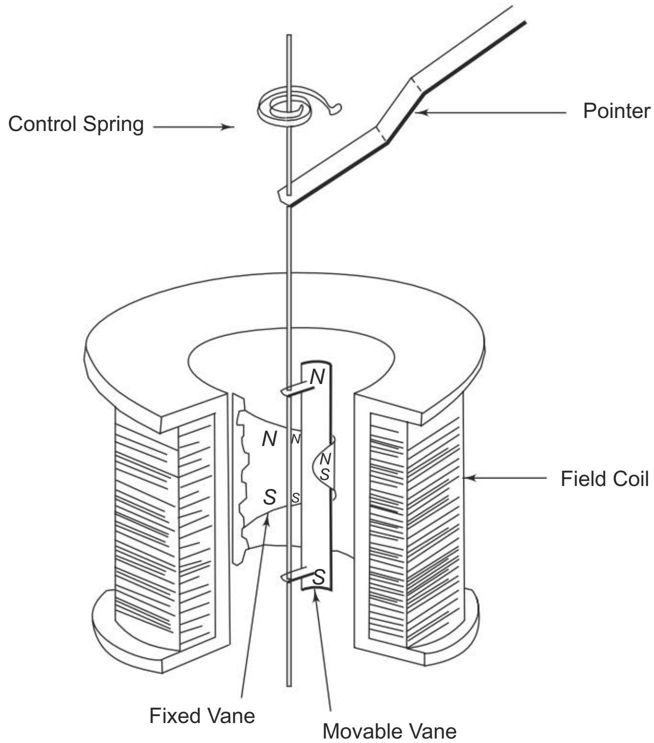


Fig. 2.9 Concentric iron vane (Repulsion type)

Damping is obtained by a light aluminium damping vane, rotating with small clearance in a closed air chamber. When used on ac, the actual operating torque is pulsating and this may cause vibration of the pointer. Rigid (trussed) pointer construction effectively eliminates such vibration and prevents bending of the pointer on heavy overloads. The concentric vane moving iron instrument is only moderately sensitive and has square law scale characteristics. The accuracy of the instrument is limited by several factors: (i) the magnetisation curve of the iron vane is non-linear. (ii) at low current values, the peak to peak of the ac produces a greater displacement per unit current than the average value, resulting in an ac reading that may be appreciably higher than the equivalent dc reading at the lower end of the scale. Similarly, at the higher end of the scale, the knee of the magnetisation curve is approached and the peak value of the ac produces less deflection per unit current than the average value, so that the ac reading is lower than the equivalent dc value.

(Hysteresis in iron and eddy currents in the vanes and other metal parts of the instrument further affect the accuracy of the reading.) The flux density is very small even at full scale values of current, so that the instrument has a low current sensitivity. There are no current carrying parts in the moving system, hence the iron vane meter is extremely rugged and reliable. It is not easily damaged even under severe overload conditions.

Adding a suitable multiplier converts the iron vane movement into a voltmeter; adding a shunt produces different current ranges. When an iron vane movement is used as an ac voltmeter, the frequency increases the impedance of the instrument and therefore a lower reading is obtained for a given applied voltage. An iron vane voltmeter should therefore always be calibrated at the frequency at which it is to be used. The usual commercial instrument may be used within its accuracy tolerance from 25–125 Hz.

DIGITAL DISPLAY SYSTEM AND INDICATORS 2.7

The rapid growth of electronic handling of numerical data has bought with it a great demand for simple systems to display the data in a readily understandable form. Display devices provide a visual display of numbers, letters, and symbols in response to electrical input, and serve as constituents of an electronic display system.

CLASSIFICATION OF DISPLAYS 2.8

Commonly used displays in the digital electronic field are as follows.

1. Cathode Ray Tube (CRT)
2. Light Emitting Diode (LED)
3. Liquid Crystal Display (LCD)
4. Gas discharge plasma displays (Cold cathode displays or Nixies)
5. Electro-Luminescent (EL) displays
6. Incandescent display
7. ElectroPhoretic Image Displays (EPID)
8. Liquid Vapour Display (LVD)

In general, displays are classified in a number of ways, as follows.

1. On methods of conversion of electrical data into visible light
   (a) Active displays
   (Light emitters – Incandescent, i.e. due to temperature, luminescence, i.e. due to non-thermal means or physio-thermal, and gas discharge-glow of light around the cathode.)
   — CRTs, Gas discharge plasma, LEDs, etc.
   (b) Passive displays
   Light controllers, LCDs, EPIDs, etc.
2. On the applications
   (a) Analog displays — Bar graph displays (CRT)
   (b) Digital displays — Nixies, Alphanumeric, LEDs, etc.
3. According to the display size and physical dimensions
   (a) Symbolic displays — Alphanumeric, Nixie tubes, LEDs, etc.
   (b) Console displays — CRTs, LEDs, etc.
   (c) Large screen display — Enlarged projection system
4. According to the display format
   (a) Direct view type (Flat panel planar) — Segmental, dotmatrix — CRTs

(b) Stacked electrode non-planar type — Nixie

5. In terms of resolution and legibility of characters
   (a) Simple single element indicator
   (b) Multi-element displays

DISPLAY DEVICES 2.9

When displaying large quantities of alphanumeric data, the read out system employed most commonly is a familiar CRT. Conventionally, CRTs form the basis of CROs and TV systems. To generate characters on the CRT, the generation system of characters on CRTs requires relatively simple electronic circuitry.

A typical CRT display has easy facilities for the control of digit size by controlling the deflection sensitivity of the system (either electromagnetic or electrostatic deflection). The number of characters displayed can be changed with the help of time shared deflection and modulator circuits.

Importantly, the intensity and brightness can be realised. with different gray scales, and the display can have different colour depending on the phosphor used in the screen. Generally the phosphor is chosen to be white or green.

Storage type CRTs facilitate storing a stationary pattern on the screen without flickering display and it is possible to retain the pattern for a long time, independent of the phosphor persistence.

LIGHT EMITTING DIODES (LED) 2.10

The LED, Fig. 2.10 (a) is basically a semiconductor PN junction diode capable of emitting electromagnetic radiation under forward conductions. The radiation emitted by LEDs can be either in the visible spectrum or in the infrared region, depending on the type of the semiconductor material used. Generally, infra-red emitting LED's are coated with Phosphor so that, by the excitation of phosphor visible light can be produced. LEDs are useful for electronics display and instrumentation. Figure 2.10 (b) shows the symbol of an LED.

The advantage of using LEDs in electronic displays are as follows.

1. LEDs are very small devices, and can be considered as point sources of light. They can therefore be stacked in a high-density matrix to serve as a numeric and alphanumeric display. (They can have a character density of several thousand per square metre).
2. The light output from an LED is function of the current flowing through it. An LED can therefore, be smoothly controlled by varying the current. This is particularly useful for operating LED displays under different ambient lighting conditions.
3. LEDs are highly efficient emitters of EM radiation. LEDs with light output of different colours, i.e. red, amber, green and yellow are commonly available.
4. LEDs are very fast devices, having a turn ON-OFF time of less than 1 ns.

5. The low supply voltage and current requirements of LEDs make them compatible with DTL and TTL, ICs.

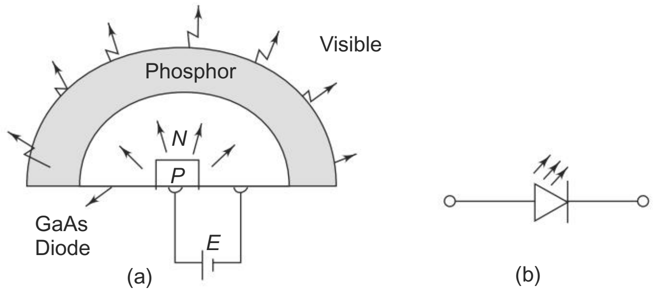


Fig. 2.10 (a) Structure of a visible emitter using GaAs PN junction (b) Symbol of LED

In germanium and silicon semiconductors, most of the energy is released in the form of heat. In Gallium Phosphide (GaP) and Gallium Arsenide Phosphide (GaAsP) most of the emitted photons have their wavelengths in the visible regions, and therefore these semiconductors are used for the construction of LEDs. The colour of light emitted depends upon the semiconductor material and doping level.

Different materials used for doping give out different colours.

1. Gallium Arsenide (GaAs) — red
2. Gallium Arsenide Phosphide (GaAsP) — red or yellow
3. Gallium Phosphide (GaP) — red or green

Alphanumeric displays using LEDs employ a number of square and oblong emitting areas, arranged either as dotmatrix or segmented bar matrix.

Alphanumeric LEDs are normally laid out on a single slice of semiconductor material, all the chips being enclosed in a package, similar to an IC, except that the packaging compound is transparent rather than opaque. Figure 2.10 (c) and (d) gives typical LED packages for single element LEDs.

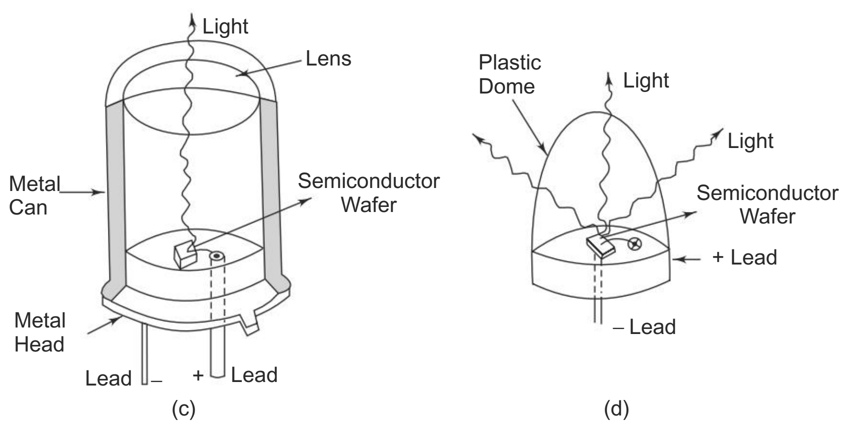


Fig. 2.10 (c) Metal can Tθ-5 type (d) Epoxy type

LIQUID CRYSTAL DISPLAY (LCD) 2.11

LCDs are passive displays characterised by very low power consumption and good contrast ratio. They have the following characteristics in common.

1. They are light scattering.
2. They can operate in a reflective or transmissive configuration.
3. They do not actively generate light and depend for their operation on ambient or back lighting.

A transmissive LCD has a better visual characteristic than a reflective LCD. The power required by an LCD to scatter or absorb light is extremely small, of the order of a few µW/cm. LCDs operate at low voltages, ranging from 1–15 V.

The operation of liquid crystals is based on the utilisation of a class of organic materials which remain a regular crystal-like structure even when they have melted. Two liquid crystal materials which are important in display technology are nematic and cholesteric, as shown in Fig. 2.11.

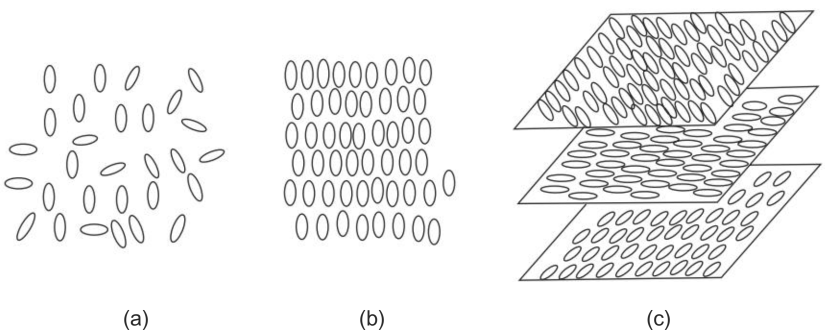


Fig. 2.11 Liquid crystal materials (a) Ordinary liquids (b) Nematic liquid crystal (c) Cholesteric liquid crystal

The most popular liquid crystal structure is the nematic liquid crystal (NLC). The liquid is normally transparent, but if it is subjected to a strong electric field, ions move through it and disrupt the well ordered crystal structure, causing the liquid to polarise and hence turn opaque. The removal of the applied field allows the crystals structure to reform and the material regains its transparency.

Basically, the LCD comprises of a thin layer of NLC fluid, about 10 µ thick, sandwiched between two glass plates having electrodes, at least one of which is transparent.

(If both are transparent, the LCD is of the transmissive type, whereas a reflective LCD has only one electrode transparent.)

The structure of a typical reflective LCD is shown in Fig. 2.12.

The NLC material in Fig. 2.12 has a homogeneous alignment of molecules. While the glass substrate supports the LCD and provides the required transparency, the electrode facilitates electrical connections for the display. The insulating spacers are the hermetic seal.

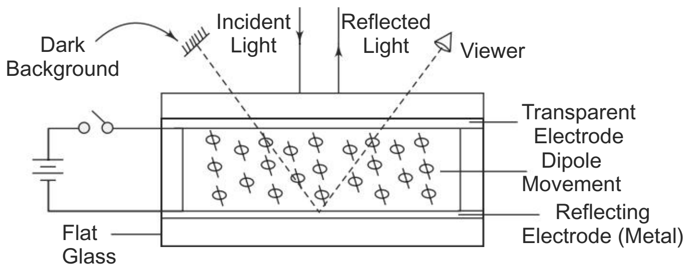


Fig. 2.12 Reflective display using NLC

The LCD material is held in the centre cell of a glass sandwich, the inner surface of which is coated with a very thin conducting layer of tin-oxide, which can be either transparent or reflective. The oxide coating on the front sheet of the indicator is etched to produce a single or multisegment pattern of characters and each segment of character is properly insulated from each other.

LCDs can be read easily in any situation, even when the ambient light is strong. If the read electrode is made transparent instead of reflective, back illumination is possible by a standard indicator lamp. Extending back illumination a step further by adding a lens arrangement. LCDs can be used as the slide in a projection system, to obtain an enlarged image.

Important Features of LCDs

1. The electric field required to activate LCDs is typically of the order of 10^4 V/cm. This is equivalent to an LCD terminal voltage of 10 V when the NLC layer is 10 μ thick.
2. NLC materials possess high resistivity > 10^{10} Ω. Therefore the current required for scattering light in an NLC is very marginal (typically 0.1 $\mu A/cm^2$).
3. Since the light source for a reflective LCD is the ambient light itself, the only power required is that needed to cause turbulence in the cell, which is very small, typically 1 μW/cm.
4. LCDs are very slow devices. They have a turn-on time of a few milli-seconds, and a turn-off time of tens of milliseconds.

To sum up, LCDs are characterised by low power dissipation, low cost, large area and low operating speed.

LCDs are usually of the seven segment type for numeric use and have one common back electrode and seven transparent front electrodes characters, as shown in Fig. 2.13.

The back electrode may be reflective or transmissive, depending on the mode of operation of the display device.

Generally arrays of such characters are simultaneously fabricated using thin-film or hybrid IC technology for segments and conductors on glass plates, and then filled in with NLC material, followed by hermetic sealing.

LCD arrays utilising a dot-matrix are also possible, but they are not popular because of their slow operation.

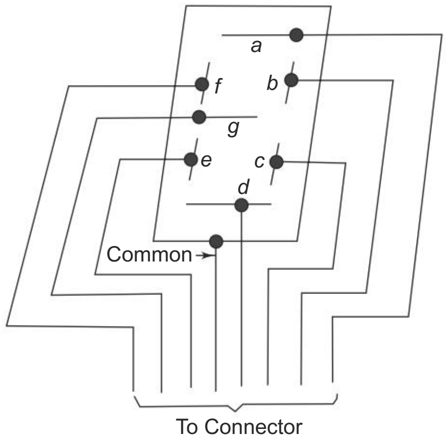


Fig. 2.13 Seven segment LCD character

OTHER DISPLAYS 2.12

Other important displays for use in electronic instrumentation are gas discharge plasma, electroluminescent, incandescent, electrophoretic, and liquid vapour display.

2.12.1 Gas Discharge Plasma Displays

These are the most well-known type of alphanumeric displays. Their operation is based on the emission of light in a cold cathode gas filled tube under breakdown condition.

These cold cathode numerical indicators are called Nixies (Numicators and Numbertrons).

This Nixie tube is a numeric indicator based on glow discharge in cold cathode gas filled tubes. It is essentially a multicathode tube filled with a gas such as neon and having a single anode, as shown in Fig. 2.14.

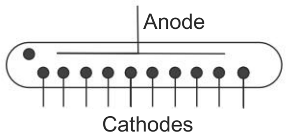


Fig. 2.14 Nixie tube—Symbolic representation

Each of the cathodes is made of a thin wire and is shaped in the form of characters to be displayed, for example, numerals 0 to 9. The anode is also in the form of a thin frame.

In its normal operation, the anode is returned to positive supply through a suitable current limiting resistor, the value of the supply being greater than the worst-case breakdown voltage of the gas within the tube. The gas in the vicinity of the appropriate cathode glows when the cathode is switched to ground potential.

(The characteristic orange red glow in the case of neon covers the selected cathode completely, thereby illuminating the character brightly.)

Since 10 cathodes have to be associated with a single anode inside the glass bulb, they have necessarily to be stacked in different planes. This requires different voltages for different cathodes to enable the glow discharge.

Many Nixie tubes also possess dot-cathodes either on the left or right of the character to serve as decimal points.

The standard Nixie is not the only format used with cold cathode technology—both bar and dot matrix versions are available. The bar types have a cathode which forms the segment and operates in a fashion similar to the standard neon tube. Identical supply voltage and drivers are required. In the dot type display, each dot is in matrix fashion and operates as an individual glow discharge light source. The required dots are selected by an *X-Y* addressing array of thin film metal lines, as shown in Fig. 2.15 (a).

Nixie tubes have the following important characteristics.

1. The numerals are usually large, typically 15–30 mm high, and appear in the same base line for in-line read-out.
2. Nixie tubes are single digit devices with or without a decimal point.
3. They are either side viewing or top viewing (as shown in Figs 2.15 (b) and (c)).

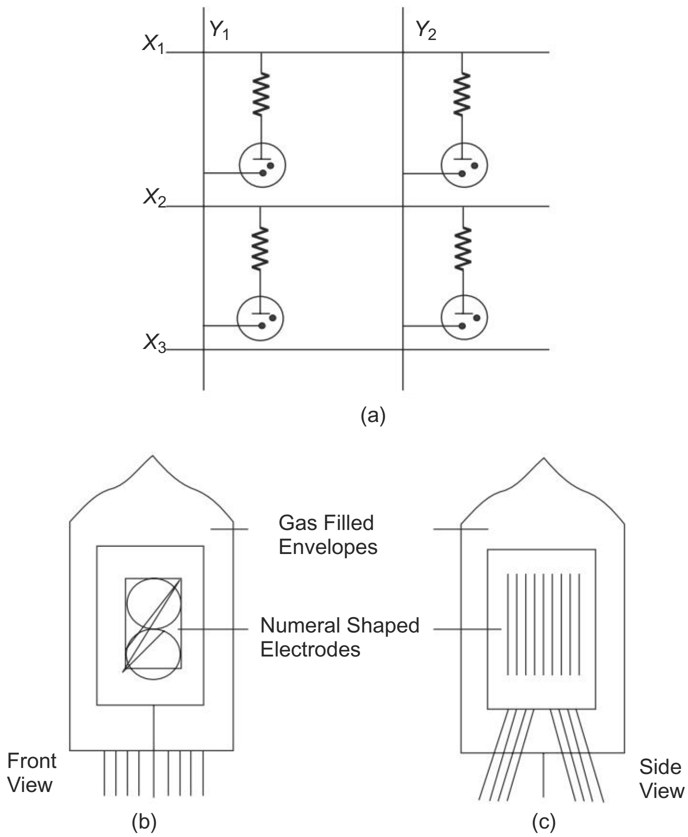


Fig. 2.15 (a) Matrix operation of display panel using gas filled devices (b) and (c) nixie tube

4. Most Nixie tubes require dc supply of 150–220 V, and the selected cathode carries current in the range of 1–5 mA.

5. The Nixie tube can be pulse operated and hence can be used in multiplexed displays.
6. Alphabetical symbols can also be introduced in the Nixie tube.

2.12.2 Segmented Gas Discharge Displays

Segmented gas discharge displays work on the principle of gas discharge glow, similar to the case of Nixie tubes. They are mostly available in 7 segment or 14 segment form, to display numeric and alphanumeric characters.

Since these devices require high voltages, special ICs are developed to drive them. The construction of a 7 segment Display is shown in Fig. 2.16. Each segment (decimal point) of the 7 segment display formed on a base has a separate cathode. The anode is common to each member of the 7 segment group which is deposited on the covering face plate. The space between the anodes and cathodes contains the gas. For each group of segments, a 'keep alive' cathode is also provided. For improving the switching speeds of the display a small constant current (a few micro amps) is passed through this keep alive cathode, which acts as a source of ions. Pins are connected to the electrodes at the rear of the base plate, with the help of which external connections can be made.

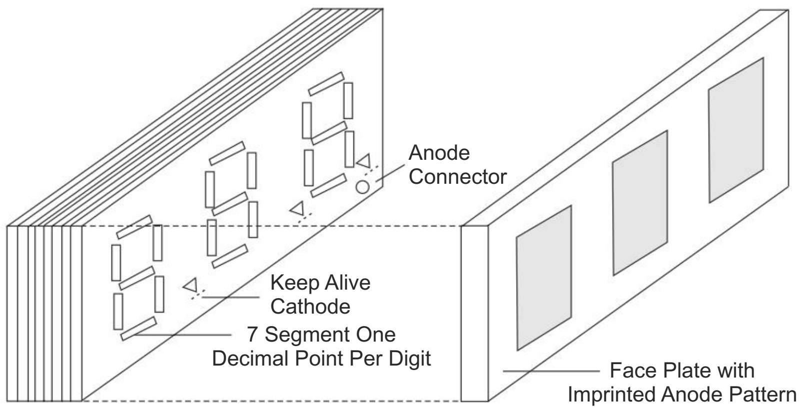


Fig. 2.16 Seven segment display using gaseous discharge

The major disadvantage of this gas discharge tube is that high voltage is required for operating it. Therefore, high voltage transistors, in the range of 150 – 200 V, are required as switches for the cathodes. A major advantage is that the power consumed is extremely small, because a bright display can be obtained even for currents as low as 200 µA.

This display follows a simple construction. Figure 2.17 gives the structure of a typical 7 segment display making use of a gas discharge plasma.

The device uses a glass substrate, shown in Fig. 2.17. Back electrodes of the thick film type serve as cathode segments, and front electrodes of the thin film type serve as transparent anodes. A gas, typically neon, is filled in the discharge space between the cathode and anode segment. The gas is struck between the cathode and anode of a chosen segment so that the cathode glow provides the illumination. All numeric characters can be displayed by activating the appropriate segment.

Display panels of rows or columns of such characters can be easily constructed by extending a single character. The power requirements of such devices are more or less in the same range as those for Nixie tubes.

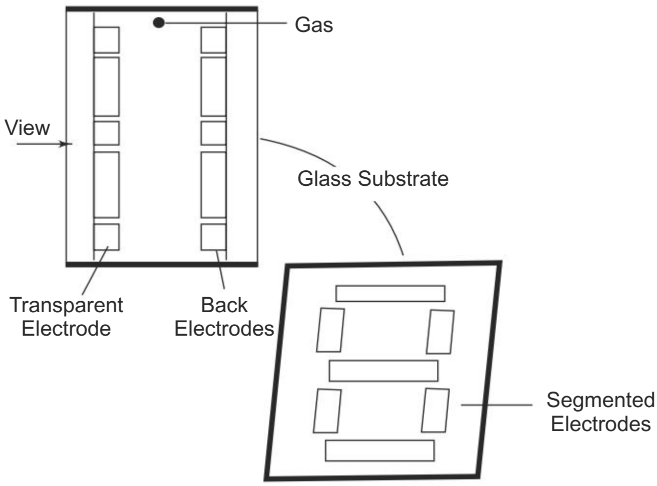


Fig. 2.17 Seven segment gas filled character

2.12.3 Segmental Displays using LEDs

In segmental displays, it is usual to employ a single LED for each segment.

For conventional 7 segment LED displays (including the decimal point, i.e. the 8th segment), the wiring pattern is simplified by making one terminal common to all LEDs and other terminals corresponding to different segments. The terminals can be either of the common anode (CA) form or common cathode (CC) form, shown in Figs 2.18 (b) and (c).

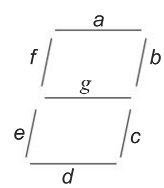


Fig. 2.18 (a) LED 7 segment format

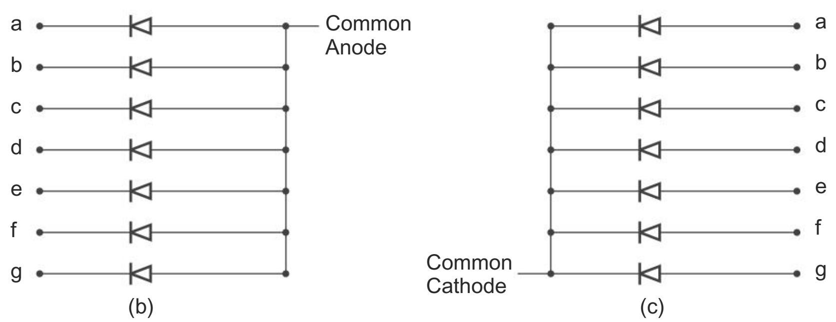


Fig. 2.18 (b) Common anode connections (c) Common cathode connections

A typical static single digit 7 segment LED display system and multi-digit are shown in Figs. 2.18 (a) and (d).

Multi-digit display system may be static or dynamic.

Common anode type displays require an active low (or current sinking) configuration for code converter circuitry, whereas an active high (or current sourcing) output circuit is necessary for common-cathode LED type display.

Both multi-digit and segmental displays require a code converter; one code converter per character for static display systems and a single code converter for time shared and multiplexed dynamic display systems, which are illuminated one at a time.

The typical circuit schemes described in the figures are only of the decimal numeric character. An 8 digit display system, operating on this principle and suitable for digital instrumentation is given in Fig. 2.18 (d).

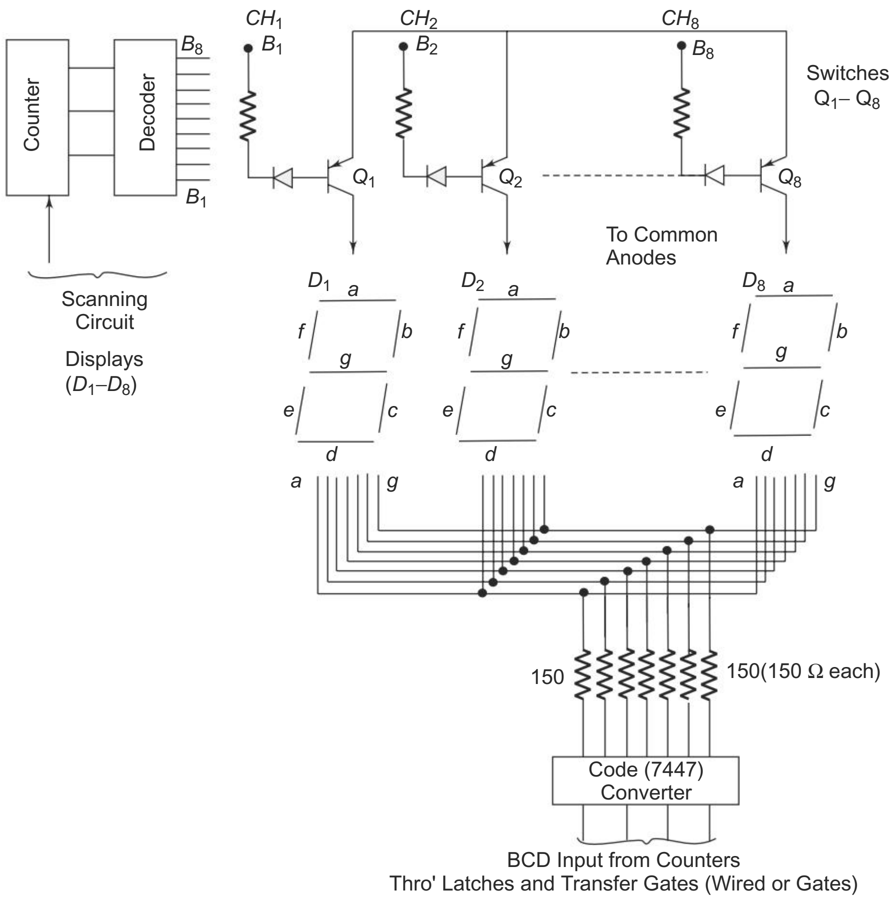


Fig. 2.18 (d) Multi-digit display system (8 digit) using LED 7 segment characters

It is also possible to generate hexadecimal numeric characters and conventional alphanumeric characters using 7 segment and 14 or 16 segment LED display units respectively, with a proper code converter. Both static and dynamic displays can be realised using LCDs, either in a common format (7 segment) or in single or multi character.

A chopped dc supply may be used, for simplicity, but conventionally an ac voltage is applied either to the common electrode or to the segment. Various segmental LCD driver circuits are displayed in Fig. 2.19.

Referring to Figs 2.19 (a) and (b), it is seen that an ac voltage (V_{ac}) is applied to either the common electrode or to the segment. High value resistances ($R > 1M$) are included in the circuit, as shown. The code converter controls the switches (S). V_{ac} is present across the selected segment and the common electrode when S is ON, and the voltage between any other segment (S-OFF) and the common electrode is zero. Hence the desired segments are energised, provided V_{ac} has a magnitude greater than or equal to the operating voltage of the LCD.

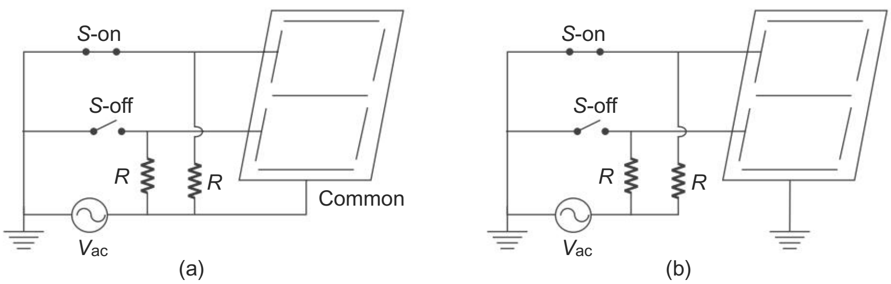


Fig. 2.19 (a) Segments driving circuits for LCD, switching method common electrode (b) Segments Driving circuits for LCD, switching method

The basic operation of the phase shift method for driving the segment is shown in Fig. 2.19 (c). In this circuit, ac voltages of the same amplitude and frequency (not necessarily same phase) are supplied to the common electrode as well as the segments.

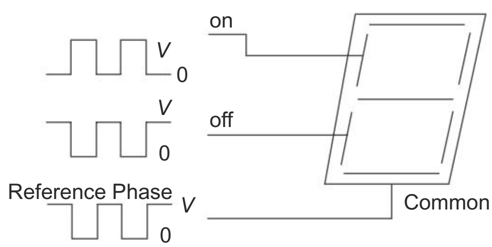


Fig. 2.19 (c) Segments driving circuits for LCD, using phase shift method

There will be a finite voltage drop between a segment and the common electrode only when the ac voltages applied are out of phase, and thus the selected segment is energised. On the other hand, when in-phase voltages are present, the voltage drop between a segment and the common electrode is zero, leading to the off state.

2.12.4 Dot Matrix Displays

Excellent alphanumeric characters can be displayed by using dot matrix LEDs with an LED at each dot location. Commonly used dot matrices for the display of prominent characters are 5 × 7, 5 × 8, and 7 × 9, of which 5 × 7 shown in

Fig. 2.20 (a), is very popular due to economic considerations. The two wiring patterns of dotmatrix displays are as follows.

1. Common anode or common cathode connection (uneconomical).
2. $X - Y$ array connection (economical and can be extended vertically or horizontally using a minimum number of wires, Fig. 2.20 (b)).

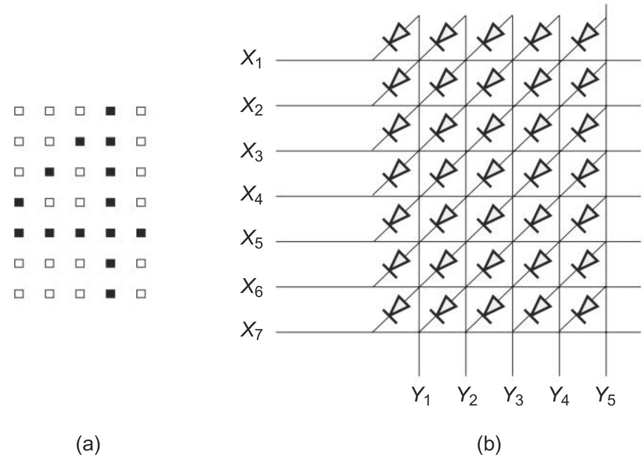


Fig. 2.20 (a) 5×7 dot matrix character using LED (b) Wiring pattern for 5×7 LED character

A typical 3 digit alphanumeric character display system using 5×7 dot matrix LEDs is shown in Fig. 2.21.

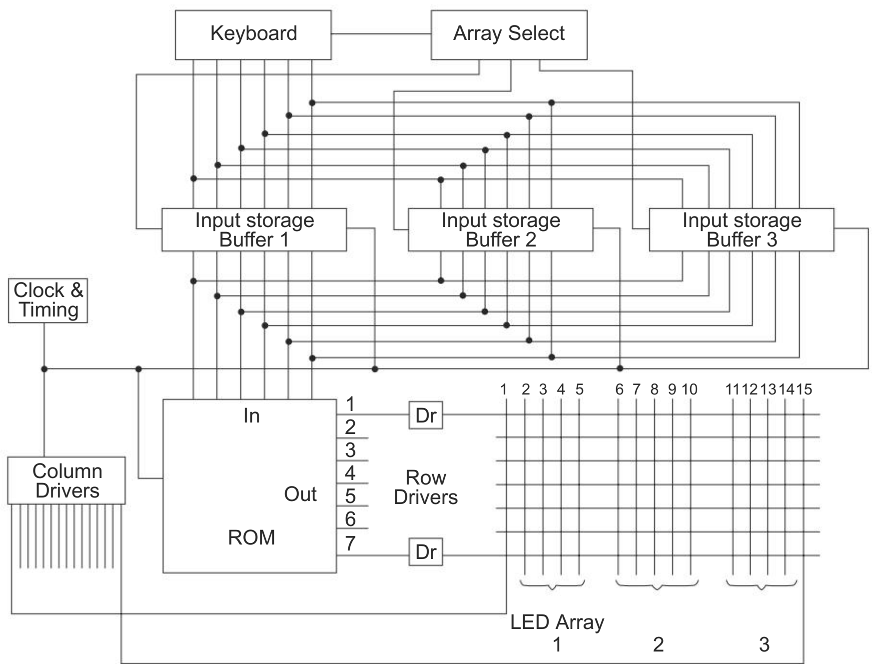


Fig. 2.21 A 3 digit alphanumeric display system using 5×7 characters

2.12.5 Bar Graph Displays

Bar graph displays are analogue displays which are an alternative to conventional D'Arsonval moving coil meters. They use a closely packed linear array or column of display elements, i.e. "DOT-LED'S", which are independently driven so that the length of the array (or the height of the column) corresponds to the voltage or current being measured. These displays are generally used in the panel meters to accept analog input signals and produce an equivalent display of the input signal level by illuminating the corresponding LEDs, as shown in Fig. 2.22.

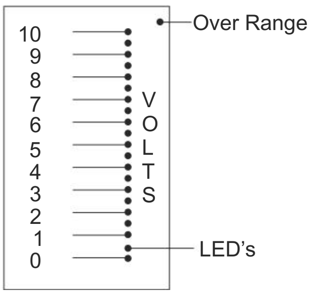


Fig. 2.22 Analog meter using bar graph of LEDs

2.12.6 Electro Luminescent (EL) Displays

Electro luminescent displays are an important means of light generation. They can be fabricated using polycrystalline semiconductors, and in view of their simple technology, brightness of display and possibility of different colours, are rapidly gaining in popularity.

The semiconductors used for EL displays are essentially phosphor powder or film type structures.

The powder type consist of powder phosphor with some binding material e.g. organic liquids deposited on a sheet of glass. The glass has transparent conductive segments (e.g. 7 segment displays) or dots (dot matrix display) along with the required conductive leads on the side on which phosphor is coated for electrical connections.

A metallic electrode, usually aluminium, is placed over the phosphor in a pressure cell by vacuum evaporation, so as to form an electrical connection on the other side of the phosphor. The resulting device is capacitive, because of poor conduction paths in phosphor. An ac field applied across the chosen segment (or dot) and aluminium electrode excites the phosphor, resulting in emission of light. In film type structures the EL powder structure is replaced by a polycrystalline phosphor film which is deposited on a glass substrate using a vacuum or pressure cell. These devices can be operated by ac as well as dc.

2.12.7 Incandescent Display

Incandescence has been a basic process of light generation for several decades. This process is now down in fully integrated electronic displays.

Incandescent displays using 16 segment as well as 5×7 dot matrix formats fabricated using thin film micro electronics are now available for alphanumeric characters. Such displays are characterised by simple technology, bright output and compatibility with ICs, but at very low operating speeds.

A thin film of tungsten can be made to emit light if its temperature is raised to about 1200°C by electrical excitation. A 5×7 character array is formed on a ceramic substrate employing such films in a matrix form and is used as an integrated electronic display unit. Figure 2.23 gives a typical tungsten film or filament suitable for a dot location in the display.

An array of such filaments can be formed on ceramic substrates using conventional thin film technology commonly used in semiconductor fabrication.

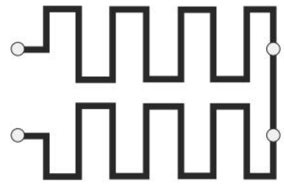


Fig. 2.23 Tungsten filament suitable for incandescent display

Considering the filament dimensions and the dimensions of commonly available substrates, an array of three characters can be located on a 2.5 cm ceramic substrate.

16 segment incandescent displays are also available, but their display is slow, because of the large thermal time constant associated with the filaments.

2.12.8 Electrophoretic Image Display (EPID)

Electrophoresis is the movement of charged pigment particles suspended in a liquid under the influence of an electric field. This pheonomenon has been utilised in electrophoretic image displays, as shown in Fig. 2.24.

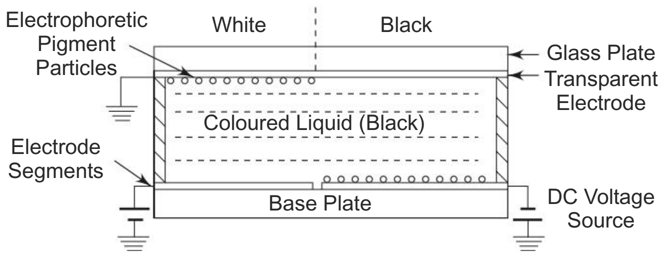


Fig. 2.24 Structure of an EPID

The basic principle; fabrication and operating characteristics of a reflective type Electrophoretic Image Display (EPID) panel are as follows. These displays are characterised by large character size, low power dissipation and internal memory.

The relatively slow speed of these displays is a major limitation, particularly for use as a dynamic display. The life span of an EPID is a few thousand hours only.

The EPID panel makes use of the electrophoretic migration of charged pigment particles in a supension. The suspension, 25 – 100 µ thick, which largely contains the pigment particles and a suspending liquid, is sandwiched between a pair of electrodes, one of which is transparent.

The application of a dc electric field, across the electrodes, as shown in Fig. 2.24 moves the particles electrophoretically towards either electrode, the movement depending mainly on the polarity of the charge on the particles. The reflective colour of the suspension layer changes on account of this migration.

EPID panels generally follow a segmented character format – typically 7 segment for numeric characters.

It is usual to have the transparent electrode as a common electrode. The back electrodes are generally segmented. Two such segments are shown in Fig. 2.24.

During the normal operation of the display, the transparent electrode is maintained at ground potential and the segmented electrodes at the back are given different potentials.

If the pigment particles are white and positively charged in the black suspending liquid, the application of a positive voltage to the chosen segment moves the pigment particles away from it and towards the transparent electrode. This is shown on the left side of Fig. 2.24. Pigment particles appear white in reflective colour as viewed through the transparent electrodes.

On the other hand, when a segment has a negative voltage with reference to the transparent electrode, the white pigment particles go towards it and get immersed in the black suspension. In this case, the viewer sees the reflection from the black liquid itself.

Colour combinations of both the pigment particles and the suspending liquid can be used to achieve a desired colour display.

Moreover, the colours between the displayed pattern and its background can be reversed, by changing the polarities of segment voltages.

In addition, the EPID panel has a memory, because the pigment particles deposited on an electrode surface remain there even after the applied voltage is removed.

2.12.9 Liquid Vapour Display (LVD)

LVDs are the latest in economical display technology. They employ a new reflective passive display principle and depend on the presence of ambient lights for their operation. Figure 2.25 gives the structure of a typical LVD cell.

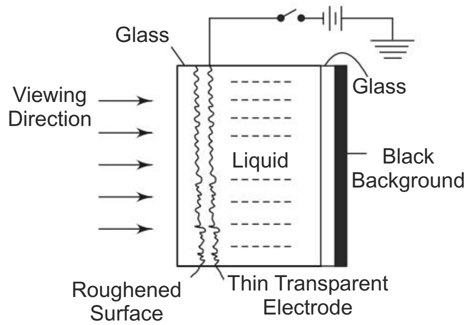


Fig. 2.25 Structure of an LVD Cell

It consists of a transparent volatile liquid encased between two glass plates and side spacers. The rear glass plate has a black background and the front glass surface in contact with the liquid is roughened, so that the liquid wets it, i.e. in its simplest form, an LVD consists of a roughened glass surface wetted with a transparent volatile liquid of the same refractive index as that of the glass. The rear surface is blackened.

The transparent electrode is heated by using a voltage drive, which is the basis for the display function.

In the OFF condition of display with no voltage applied across the transparent electrode, the viewer sees the black background through the front transparent glass electrode and the liquid.

To achieve an ON condition of the display, a voltage is applied to the transparent electrode. This causes sufficient heat in the electrode, which evaporates the liquid in contact with it, and a combination of vapour film and vapour bubbles is formed around the roughened glass surface. As the refractive index of vapour is approximately 1, there is a discontinuity established at the interface between the front glass plate and the liquid, which gives rise to light scattering. This makes it a simple display device.

The organic liquid selected for LVD should have the following features.

1. Refractive index close to that of the glass plate.
2. Minimum energy for vapourising the liquid in contact with the roughened surface.

The electrical heating of a thin film of liquid adjacent to the roughened surface using transparent electrodes and the applied voltage, makes it an unusually good display with a better contrast ratio than an LCD. The speed of operation of LVDs is low.

A summary of some important display devices is given in Tables 2.1 and 2.2.

Table 2.1 Some Popular Display Devices

Display devices	*Applications*	*Advantages*	*Disadvantages*
1. CRTs	Large display, small and large group viewing, console display.	Bright, efficient, uniform, planar display—all colours, high reliability	Bulky, high voltage, non-digital address, high initial cost.
2. LEDs	Indicators and small displays, individual viewing, flat-panel.	Bright, efficient, red, yellow, amber, green colours, compatible with ICs, small size	High cost per element, limited reliability, low switching speed
3. LCDs	do	Good contrast in bright ambient light, low power, compatible with ICs, low cost element	Limited temperature range (0 – 60°C) Limited reliability, ac operation necessary, low switching speed.
4. NIXIEs	Indicators, small, medium and large displays, small group viewing.	Bright, range of colours, low cost element, compatible with ICs	High drive power
5. ELs	Indicator and small display, flat panel	Low cost element, many columns	Not compatible with ICs

Table 2.2 Typical applications of digital display

Field of applications	Displays
1. *Industrial Electronics*	
Meters, positioner and instrumentation, test equipment, guages and counters	Incandescent, LED, LCD, CRT, Nixie
2. *Medical*	
Digital thermometer, Pulse rate meter, manometer, patient monitoring	CRT, LED's
3. *Computers, Commerce and Business*	
Peripheral and ALU status, calculators and cash register	LED, CRT, EL, Nixie, LCD
4. *Domestic*	
Electronic Oven, telephone, dial indicator, TV channel indicators, clock and calendars, video games	LED, CRT, LCD, Nixie
5. *Military and Space Research*	
Situation indicators	Traditional

PRINTERS 2.13

Character printers and graphic plotters are the two devices used to prepare a permanent (or hard copy) record of computer output.

The basic difference between printers and plotters is that the former are devices whose purpose is to print letters, numbers and similar characters in text readable form, while the latter print diagrams with continuous lines.

CLASSIFICATION OF PRINTERS 2.14

Printers used in computers are classified in the following three broad categories.

1. Impact and Non-impact Printers

Impact printers form characters on a paper by striking the paper with a print head and squeezing an inked ribbon between the print head and the paper.

Non-impact printers form characters without engaging the print mechanism with the print surface, e.g. by heating sensitised paper or by spraying ink from a jet.

2. Fully Formed Character and Dot Matrix Printer

Fully formed characters are like those made by a standard typewriter—all parts of characters are embossed in the reverse on the type bars of the typewriter. When printed, all type elements appear connected or fully formed.

Dot matrix characters are shaped by combinations of dots that form a group representing a letter or number when viewed together.

3. Character at a Time and Line at a Time Printer

Character at a time printers (character printers or serial printers), print each character serially, and virtually instantaneously.

Line at a time printers (line printers), print each line virtually instantaneously.

(Some advanced printers, e.g. those using lasers and xerographic methods, print lines so rapidly that they virtually print a page at a time, and are therefore called page printers. They are rarely used in mini computers and microcomputers, for special purposes like phototype setting.)

PRINTER CHARACTER SET 2.15

Most printers used with mini or micro computers use ASCII codes. Printers are specified as using the 48 character set, the 64 character set, the 96 character set or the 128 character set.

The 48 and 64 character sets include commonly used special symbols, numbers, a space, and upper case (capital) English alphabets.

The 96 ASCII character set includes the lower case English alphabet and several additional special symbols. Of the 96 characters, 'space' and 'delete' do not print, leaving only 94 printable characters.

The entire 128 character ASCII set contains 32 characters normally used for communication and control. These characters usually do not print, but correspond to expandable functions, such as communication and control.

CHARACTER AT A TIME IMPACT PRINTERS FOR FULLY FORMED CHARACTERS (DRUM WHEEL) 2.16

The typewriter is the classic example of this printer, with characters fully formed because they are embossed on each type bar.

Ordinary type bar typewriters cannot be used with computers, because they lack a computer coding interface for easy communications.

(The classic printer used with mini and micro computers in the past was the Teletype Model 33 printer. The ready availability and low cost of these printers, plus their relatively easy interfacing, made them natural for use in small computers. The model 33 prints at a rate of 10 characters per second which is slow compared to today's printer of 55 characters per second for similar printers.) The print mechanism is a vertical cylinder. Characters are embossed in several rows and columns around the cylinder, as shown in Fig. 2.26 (a).

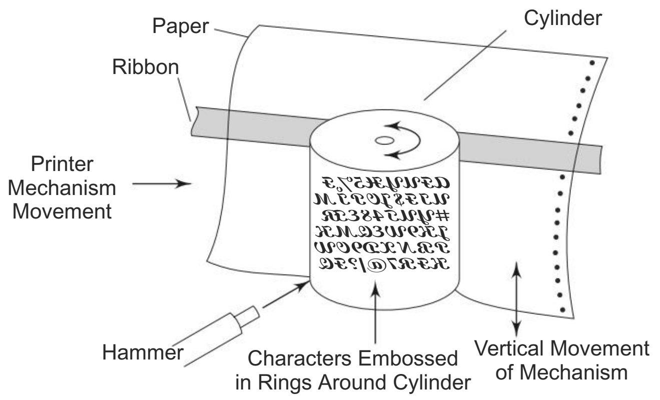


Fig. 2.26 (a) Drum wheel printer

The ASCII character code sent to the printer, is translated into motion that rotates the cylinder, so that the column containing the desired character faces the paper. The cylinder is then raised or lowered (depending on the ASCII code) to present the column containing the desired character to be printed directly to the paper. A hammer mechanism propels (hits) the cylinder towards the paper, where only the positioned character strikes the ribbon, creating the printed impression of the character on the paper.

These printers are interfaced with small computers by a 20 – 60 mA current used to transmit ASCII coded bits serially.

Another type of fully formed character printer, designed for computer use, has characters mounted on the periphery of a spinning print head, known as a daisy wheel printer, and is shown in Fig. 2.26 (b).

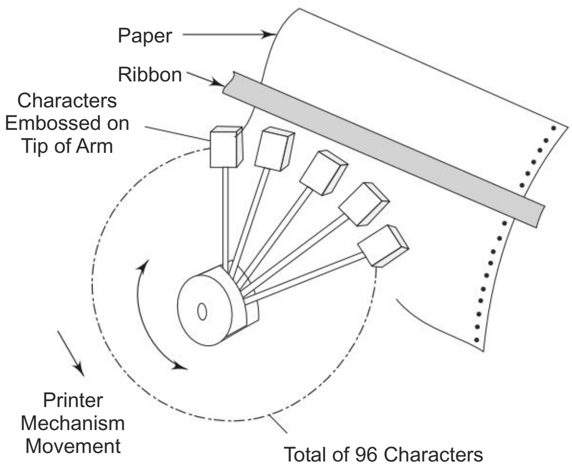


Fig. 2.26 (b) Daisy wheel printer

A daisy wheel print head is mounted on a rotating disk with flexible flower like petals similar to a daisy flower. Each petal contains the embossed character in reverse. As the daisy wheel spins, a hammer strikes the desired flexible petal containing the character, in turn impacting the paper with the embossed character through an inked ribbon.

To print a letter, the wheel is rotated until the desired letter is in position over the paper. A solenoid driven hammer then hits the petal against the ribbon to print the letter.

Daisy wheel printers are slow, with a speed of about 50 characters per second (cps). The advantage of the daisy wheel mechanism is high print quality, and interchangeable fonts.

Character at a time printing follows the following sequence of steps; left to right printing to the end of the line, stop, return carriage and start a second line, and again print left to right. It is unidirectional.

Spinning wheel printers are capable of bidirectional printing. The second line is stored in a buffer memory within the printer control circuitry and can be printed in either direction, depending on which takes the least printer time.

LINE AT A TIME IMPACT PRINTERS FOR FULLY FORMED CHARACTERS (LINE PRINTERS) 2.17

In line printers, characters or spaces constituting printable lines (typically 132 character positions wide) are printed simultaneously across the entire line. Paper is spaced up and the next line is printed. Speeds for line printers range from several hundreds to thousands of lines per minute.

Line printers are used for high volumes of printed output and less frequently in micro computers, because of their high equipment cost relative to character at a time printers.

An embossed type font is positioned across a line for printing by using embossed type, either on a carrier consisting of a chain, train or band moving horizontally across the paper and print line, or a drum rotating in front of the paper with characters embossed. Typically, there are 132 columns on the drum. As the drum rotates, the column of characters pan vertically across the paper and the print line (shown in Fig. 2.27). In both methods, hammers (one for each of 132 print positions) strikes when the correct character is positioned, imprinting the character on the paper with an inked ribbon.

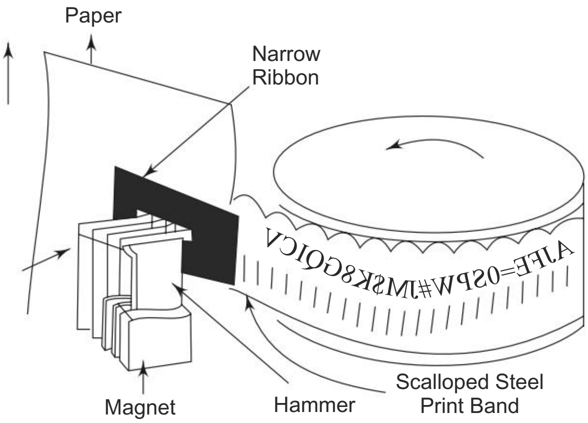


Fig. 2.27 Band printers (Line printer)

Print characters are embossed on the band. The band revolves between two capstans, passing in front of the paper. An inked ribbon is positioned between the moving band and the paper. As the print characters on the band move by 132 horizontal print positions, the 132 corresponding print hammers behind the paper strike the band at the appropriate time, causing the line of characters to print each desired character in 132 print positions.

In band printers, a metallic or plastic band has a fully formed etched character on it. The band rotates at high speed. There is one hammer for one print position, because several hammers can strike simultaneously for many print positions. These printers are faster than dot matrix printers. These line printers have speeds varying from 75 to 4000 lines per minute (1 pm). These printers are both noisier and costlier than dot-matrix printers.

A band always contains more than one character set. This reduces access time needed to match the characters, thereby reducing the printing time. Below the characters are the timing marks which are sensed by the printers electronic circuitry. It compares the character to be printed with the character corresponding to the timing mark, senses it and if a match occurs, fires the corresponding hammer.

A chain printer is similar to a band printer, except that in the former the characters sets are held in a metal or rubber chain and rotated across the paper along a print line.

A chain revolves in front of the ribbon and paper. Each link in the chain is designed to hold a pallet on which type characters are embossed. Hammers are located behind the paper and each of 132 hammers strikes the moving type pallet when the desired character passes the position in which it is timed to print.

DRUM PRINTER 2.18

Figure 2.28 illustrates a drum printer. Each of the 64 or 96 characters used is embossed in 132 columns around the drum, corresponding to the print positions. The drum rotates in front of the paper and ribbon. Print hammers strike the paper, imprinting characters from the drum through the ribbon and forming an impression on the paper.

The drum printer uses a cylindrical drum which contains characters embossed around it. There is one complete character set for each print position. To print characters, magnetically driven hammers in each character position strike the paper and ribbon against the spinning drum. An entire line of characters can be printed during each rotation of the drum. Printing speeds of drum printers vary from 200 – 300 lpm. The drawbacks of drum printers are that the fonts are not easily changeable, and the print lines may be wavy.

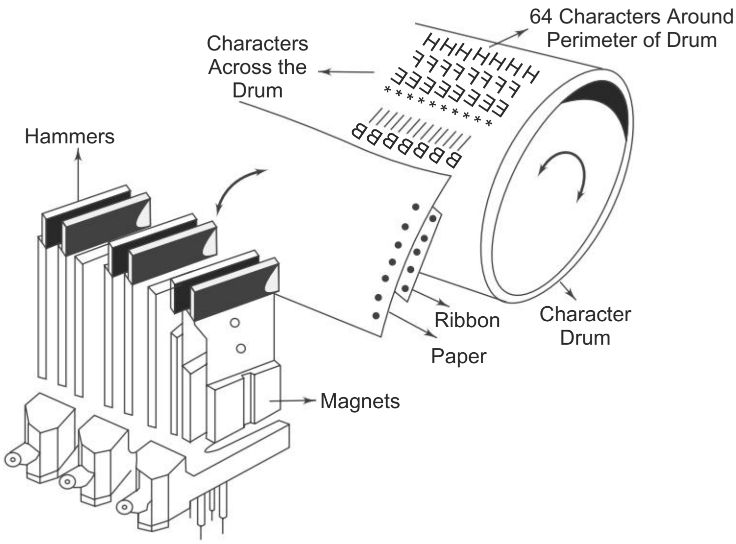


Fig. 2.28 Drum printers

DOT-MATRIX PRINTERS 2.19

Dot matrix characters are formed by printing a group of dots to form a letter, number or other symbol. This method is widely used with mini and micro computers.

Dots are formed both by impact and non impact print methods and are both character at a time and line at a time printers.

Figure 2.29 shows the letter 'A' formed by a dot-matrix, five dots wide and seven dots high (5×7) and in a 9×7 matrix. A 5×7 dot-matrix is frequently used when all letters are acceptable, in upper case.

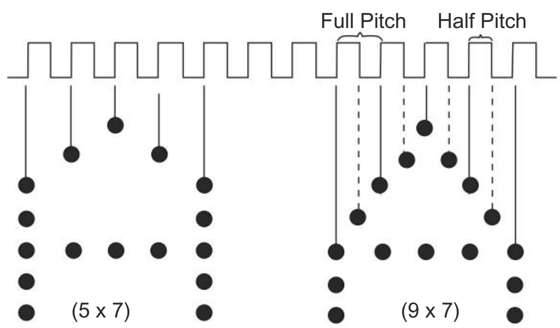


Fig. 2.29 Dot matrix sizes

Dot-matrix printers can print any combination of dots with all available print positions in the matrix. The character is printed when one of 128 ASCII codes is signalled and controlled by the ROM (read only memory) chip, which in turn controls the patterns of the dots. By changing the ROM chip a character set for any language or graphic character set can be used by the printer.

CHARACTER AT A TIME DOT-MATRIX IMPACT PRINTER 2.20

The print head for an impact dot matrix character is usually composed for an array of wires (or pins) arranged in a tabular form, that impact the character through an inked ribbon, as shown in Fig. 2.30. For this reason, these printers are sometimes also called wire printers.

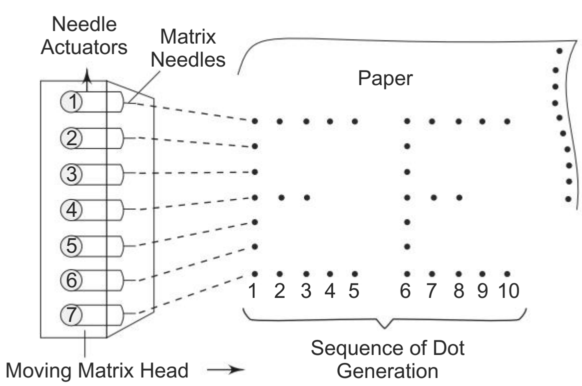
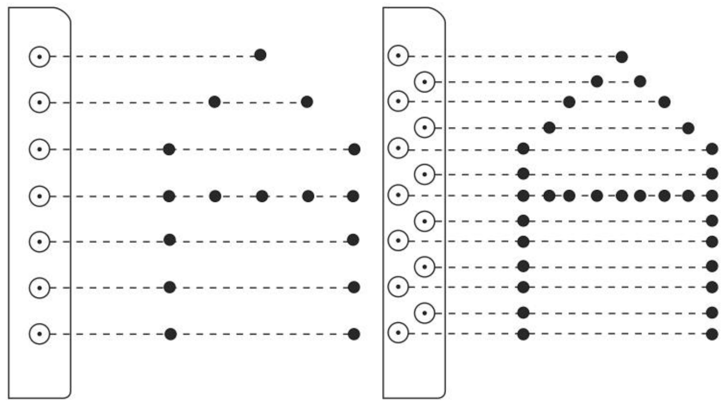


Fig. 2.30 Impact dot matrix print head

The print head often contains a single column seven wires high, though it may be two or more columns wide (Fig. 2.31 (a)).



(a) (b)

Fig. 2.31 (a) 5 × 7 single element (b) 5 × 7 double element

For purpose of illustration, assume that the print head contains a single column seven wires high. The seven wires are thrust from the print head (usually electromagnetically) in whatever combination the print controller requires to create a character. The wire strikes the ribbon and in turn impacts the paper, printing one vertical column of a single character.

The dot-matrix print head contains wires (or pins) arranged in tabular form. Characters are printed as a matrix of dots. The thin wire, driven by solenoids at the rear of the print head, strikes the ribbon against the paper to produce dots. The print wires are arranged in a vertical column, so that characters are printed out one dot column at a time as the print head moves on a line.

For a 5 × 7 full step dot-matrix character, the print head spaces one step, prints the second column of dots and repeats the pricess until all five columns are printed.

If the printer is designed to print dots in half steps, the same process is used, except that five horizontal print steps are used to form the characters (the five normal steps, plus four intervening half steps), thereby forming a 9 × 7 half step dot matrix character.

The dot-matrix character printer, strictly speaking, does not actually print a character at a time, but one column of a dot-matrix character at a time. However, the print speeds of a dot-matrix printer are very high, up to 180 characters per second.

Early dot-matrix print heads had only seven print wires, and consequently poor print quality. Currently available dot-matrix printers use 9, 14, 18 or even 24 print wires in the print head. Using a large number of print wires and/or printing a line twice with the dots for the second printing offset slightly from those of the first, ensures a better quality of print (Fig. 2.31 b).

Common speeds of dot-matrix printers range from 50 – 200 cps, but printers with speed as high as 300 cps are also available.

The dot-matrix codes of the characters are stored in EPROM. The fonts or print graphics can be changed under program control. This is the main advantage of dot-matrix printers.

The font of dot-matrix printers can be changed during printing by including the desired formats in RAM or ROM. Hence, it is possible to include standard ASCII characters, italics, subscripts, etc. on the same line. Special graphics can also be programmed into the printer.

NON-IMPACT DOT-MATRIX (NIDM) PRINTERS 2.21

Non-impact dot-matrix printers cause a mark without directly touching the paper. They are therefore quiet compared to impact printers.

They cannot make carbon copies, however, as there is no force to impress the character through multiple carbon copies. NIDM printers are useful for printing single copies of computer output, for recording the output of printing calculators and video displays, and for logging industrial data.

There are four types of NIDM printers thermal, electrosensitive, electrostatic, and ink jet.

Review Questions

1. Explain the basic principle of a D'Arsonval movement.
2. Draw and explain the construction of a PMMC movement.
3. State the functions of counterweights in a PMMC movement.
4. Describe the operation of a practical PMMC movement.
5. Explain the basic construction of a taut band movement.
6. Compare a PMMC movement with a taut band movement.
7. State the operating principle of an electrodynamometer (EDM).
8. Describe the operation of an electro-dynamometer.
9. Why is an electrodynamometer called a square law device?
10. What is a transfer instrument? Why is an electrodynamometer referred to as a transfer instrument?
11. State the principle of a moving iron type instrument.
12. Describe with a diagram the construction of a radial vane type movement.
13. Explain the operation of a repulsion type ac meter.
14. Describe with a diagram the construction of a concentric vane repulsion type movement. Explain the operation of a concentric vane repulsion type movement.
15. Differentiate between moving iron and moving coil movement.
16. State the difference between radial and concentric vane movement
17. How can an electrodynamometer be converted into an ammeter? How can an electrodynamometer be converted into a voltmeter? How can an electrodynamometer be converted into a wattmeter?
18. State the difference between analog and digital indicators.
19. How are the displays classified. List different types of display devices.

20. Draw the structure of an LED and explain its operation.
21. What are the conditions to be satisfied by the device for emission of visible light?
22. State the advantages and disadvantages of using LED in electronic display.
23. List different materials used to radiate different colours.
24. Explain the construction and operation of a seven-segment display.
25. Discuss with a neat diagram, a method of realizing a 7-segment numeric display using LEDs.
26. Bring out the important differences between the common anode and common cathode type circuit arrangements for a 7-segment numeric display using LEDs.
27. State the operating principle of LCD display.
28. State different types of liquid crystal used for LCD display.
29. Explain with diagram the operation of a Nematic Liquid Crystal (NLC).
30. Explain the basic differences between transmissive and reflective type LCD.
31. Explain with a diagram the operation of a reflective display using NLC.
32. State the important features of LCDs.
33. State the advantages of LCD display over LED display.
34. Compare LCD and LED display.
35. State the principle of a gas discharge plasma display.
36. What do you understand by Nixie?
37. Explain with a diagram the operation of a seven-segment display using gaseous discharge.
38. Give reasons for the following:
    (i) Dot matrix presentation is more popular than bar matrix in character generation in CRT.
    (ii) Reflective LCDs have many advantages over transmissive LCDs.
    (iii) Bar graphs have unique role in electronic display system.
39. What is an electro-luminscent display? Explain its operation.
40. Compare the relative performance of the following displays in numeric display application:
    (i) Electrophoretic image display
    (ii) Liquid vapour display
    (iii) Nixie tube
    (iv) Flat panel alphanumeric CRT
41. Explain with a diagram the construction and working of an electrophoretic image display.
42. Describe with diagram the operation of a liquid vapour tube.
43. What are printers? Where are they used? Describe different types of printers.
44. How is character-at-a-time printing done?
45. How is line at a time printing done?
46. What do you mean by impact and non-impact printers?
47. State the different methods of character-at-a-time printing.
48. Explain with a diagram how printing is done using a drum wheel.
49. Explain with a diagram the operation of a band printer.
50. Explain the principle of operation of a dot matrix printer.
51. Differentiate between (5×7) and (9×7) matrix.
52. Which matrix is most commonly used? Why?
53. Describe with a diagram the operation of an impact dot matrix printer.
54. Explain in brief the working of a non-impact dot matrix printer.
55. What is a daisy wheel? Explain with a diagram the operation of a daisy wheel.
56. What are dot matrix printers? How is printing done by them? How can the quality of printing be improved.
57. What are the main advantages of a dot matrix printer over other printers.
58. What are half steps in a Dot matrix? Why are they used?

Multiple Choice Questions

1. A D'Arsonval movement is
   (a) taut band
   (b) PMMC
   (c) electrodynamometer
   (d) moving iron type
2. A PMMC uses a
   (a) taut band
   (b) moving coil
   (c) electrodynamometer
   (d) moving iron type.
3. A taut band movement uses a/an
   (a) ribbon
   (b) moving coil
   (c) electrodynamometer
   (d) moving iron type
4. A moving iron movement uses
   (a) ribbon
   (b) moving coil
   (c) electrodynamometer
   (d) radial vane
5. Less power is consumed by the following devices:
   (a) LED (b) LCD
   (c) neon lamps (d) nixie tube
6. LED is based on the principle of
   (a) scattering (b) illumination
   (c) absorption (d) transmission
7. The liquid used in LCDs are
   (a) nematic (b) tantalum
   (c) oil (d) electrolytic

Further Reading

1. B.S. Sonde, *Transducers and Display Systems*, Tata McGraw-Hill, 1979.
2. Philco Technological Centre, *Electronic Precision Measurement Techniques and Experiment*.
3. C. Louis Hohenstein, 'Computer Peripherals for Mini Computer', *Micro-processor and P.C.*, McGraw-Hill, 1980.
4. A.K. Sawhney, *Electronic and Electrical Measurements*, Khanna Publishers.

Chapter 3

Ammeters

DC AMMETER 3.1

The PMMC galvanometer constitutes the basic movement of a dc ammeter. Since the coil winding of a basic movement is small and light, it can carry only very small currents. When large currents are to be measured, it is necessary to bypass a major part of the current through a resistance called a shunt, as shown in Fig. 3.1. The resistance of shunt can be calculated using conventional circuit analysis.

Referring to Fig. 3.1

R_m = internal resistance of the movement.
I_{sh} = shunt current
I_m = full scale deflection current of the movement
I = full scale current of the ammeter + shunt (i.e. total current)

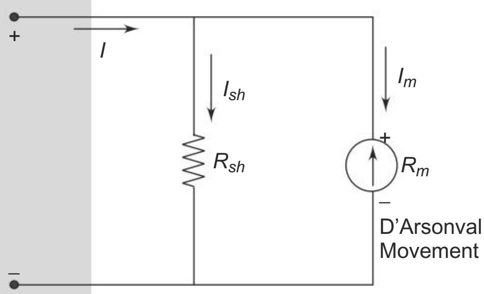


Fig. 3.1 Basic dc Ammeter

Since the shunt resistance is in parallel with the meter movement, the voltage drop across the shunt and movement must be the same.

Therefore $V_{sh} = V_m$

$\therefore \quad I_{sh} R_{sh} = I_m R_m, \qquad R_{sh} = \dfrac{I_m R_m}{I_{sh}}$

But $\quad I_{sh} = I - I_m$

hence $\quad R_{sh} = \dfrac{I_m R_m}{I - I_m}$

For each required value of full scale meter current, we can determine the value of shunt resistance.

Example 3.1 (a) *A 1 mA meter movement with an internal resistance of 100 Ω is to be converted into a 0 – 100 mA. Calculate the value of shunt resistance required.*

Solution Given $R_m = 100\ \Omega$, $I_m = 1$ mA, $I = 100$ mA

$$R_{sh} = \frac{I_m R_m}{I - I_m} = \frac{1\text{ mA} \times 100\ \Omega}{99\text{ mA}} = \frac{100\text{ mA}\Omega}{99\text{ mA}} = \frac{100\ \Omega}{99} = 1.01\ \Omega$$

The shunt resistance used with a basic movement may consist of a length of constant temperature resistance wire within the case of the instrument. Alternatively, there may be an external (manganin or constantan) shunt having very low resistance.

The general requirements of a shunt are as follows.

1. The temperature coefficients of the shunt and instrument should be low and nearly identical.
2. The resistance of the shunt should not vary with time.
3. It should carry the current without excessive temperature rise.
4. It should have a low thermal emf.

Manganin is usually used as a shunt for dc instruments, since it gives a low value of thermal emf with copper.

Constantan is a useful material for ac circuits, since it's comparatively high thermal emf, being unidirectional, is ineffective on the these circuits.

Shunt for low current are enclosed in the meter casing, while for currents above 200 A, they are mounted separately.

Example 3.1 (b) *A 100 μA meter movement with an internal resistance of 500 Ω is to be used in a 0 – 100 mA Ammeter. Find the value of the required shunt.*

Solution The shunt can also be determined by considering current I to be 'n' times larger than I_m. This is called a multiplying factor and relates the total current and meter current.

Therefore $\quad I = n\, I_m$

Therefore the equation for

$$R_{sh} = \frac{I_m R_m}{I - I_m} = \frac{I_m R_m}{nI_m - I_m} = \frac{I_m R_m}{I_m(n-1)} = \frac{R_m}{(n-1)}$$

Given: $I_m = 100\ \mu$A and $R_m = 500\ \Omega$

Step 1: $$n = \frac{I}{I_m} = \frac{100\text{ mA}}{100\ \mu\text{A}} = 1000$$

Step 2: $$R_{sh} = \frac{R_m}{(n-1)} = \frac{500\ \Omega}{1000-1} = \frac{500}{999} = 0.50\ \Omega$$

MULTIRANGE AMMETERS 3.2

The current range of the dc ammeter may be further extended by a number of shunts, selected by a range switch. Such a meter is called a multirange ammeter, shown in Fig. 3.2.

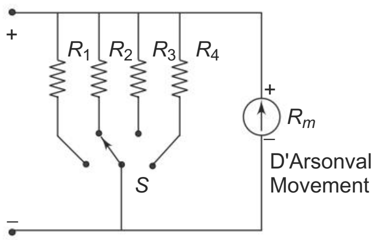


Fig. 3.2 Multirange ammeter

The circuit has four shunts R_1, R_2, R_3 and R_4, which can be placed in parallel with the movement to give four different current ranges. Switch *S* is a multiposition switch, (having low contact resistance and high current carrying capacity, since its contacts are in series with low resistance shunts). Make before break type switch is used for range changing. This switch protects the meter movement from being damaged without a shunt during range changing.

If we use an ordinary switch for range changing, the meter does not have any shunt in parallel while the range is being changed, and hence full current passes through the meter movement, damaging the movement. Hence a make before break type switch is used. The switch is so designed that when the switch position is changed, it makes contact with the next terminal (range) before breaking contact with the previous terminal. Therefore the meter movement is never left unprotected. Multirange ammeters are used for ranges up to 50A. When using a multirange ammeter, first use the highest current range, then decrease the range until good upscale reading is obtained. The resistance used for the various ranges are of very high precision values, hence the cost of the meter increases.

Example 3.2 *A 1 mA meter movement having an internal resistance of 100 Ω is used to convert into a multirange ammeter having the range 0–10 mA, 0–20 mA and 0–50 mA. Determine the value of the shunt resistance required.*

Solution Given $I_m = 1$ mA and $R_m = 100\ \Omega$

Case 1: For the range 0 – 10 mA

$$\text{Given } R_{sh1} = \frac{I_m \cdot R_m}{I - I_m} = \frac{1\text{ mA} \times 100}{10\text{ mA} - 1\text{ mA}} = \frac{100}{9} = 11.11\ \Omega$$

Case 2: For the range 0 – 20 mA

$$\text{Given } R_{sh2} = \frac{I_m \cdot R_m}{I - I_m} = \frac{1\text{ mA} \times 100}{20\text{ mA} - 1\text{ mA}} = \frac{100}{19} = 5.2\ \Omega$$

Case 3: For the range 0 – 50 mA

$$\text{Given } R_{sh3} = \frac{I_m \cdot R_m}{I - I_m} = \frac{1\text{ mA} \times 100}{50\text{ mA} - 1\text{ mA}} = \frac{100}{49} = 2.041\ \Omega$$

Example 3.3 *Design a multirange ammeter with range of 0–1 A, 5 A and 10 A employing individual shunt in each A D'Arsonval movement with an internal resistance of 500 Ω and a full scale deflection of 10 mA is available.*

Solution

Given I_m = 10 mA and R_m = 500 Ω

Case 1 : For the range 0 – 1A, i.e, 1000 mA

$$\text{Given } R_{sh1} = \frac{I_m \cdot R_m}{I - I_m} = \frac{10 \text{ mA} \times 500}{1000 \text{ mA} - 10 \text{ mA}} = \frac{5000}{990} = 5.05\ \Omega$$

Case 2 : For the range 0 – 5A, i.e, 5000 mA

$$\text{Given } R_{sh2} = \frac{I_m \cdot R_m}{I - I_m} = \frac{10 \text{ mA} \times 500}{5000 \text{ mA} - 10 \text{ mA}} = \frac{5000}{4990} = 1.002\ \Omega$$

Case 3 : For the range 0 – 10A, i.e, 10000 mA

$$\text{Given } R_{sh3} = \frac{I_m \cdot R_m}{I - I_m} = \frac{10 \text{ mA} \times 500}{10000 \text{ mA} - 10 \text{ mA}} = \frac{5000}{99990} = 0.050\ \Omega$$

Hence the values of shunt resistances are 5.05 Ω , 1.002 Ω and 0.050 Ω.

THE ARYTON SHUNT OR UNIVERSAL SHUNT 3.3

The Aryton shunt eliminates the possibility of having the meter in the circuit without a shunt. This advantage is gained at the price of slightly higher overall resistance. Figure 3.3 shows a circuit of an Aryton shunt ammeter. In this circuit, when the switch is in position "1", resistance R_a is in parallel with the series combination of R_b, R_c and the meter movement. Hence the current through the shunt is more than the current through the meter movement, thereby protecting the meter movement and reducing its sensitivity. If the switch is connected to position "2", resistance R_a and R_b are together in parallel with the series combination of R_c and the meter movement. Now the current through the meter is more than the current through the shunt resistance.

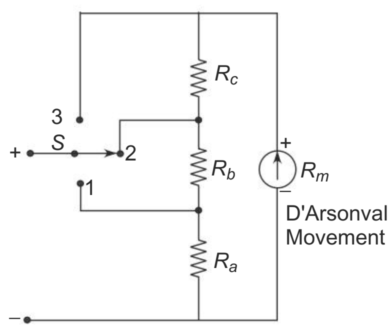


Fig. 3.3 Aryton shunt

If the switch is connected to position "3" R_a, R_b and R_c are together in parallel with the meter. Hence maximum current flows through the meter movement and very little through the shunt. This increases the sensitivity.

Example 3.4 (a) *Design an Aryton shunt (Fig. 3.4) to provide an ammeter with a current range of 0 – 1 mA, 10 mA, 50 mA and 100 mA. A D'Arsonval movement with an internal resistance of 100 Ω and full scale current of 50 μA is used.*

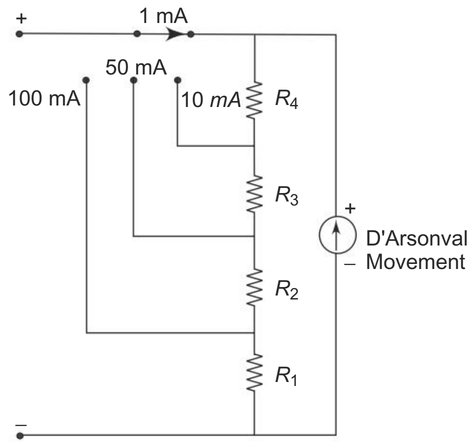


Fig. 3.4(a) For Example 3.4(a)

Solution Given $R_m = 100\ \Omega$, $I_m = 50\ \mu A$.

For 0 – 1 mA range

$$I_{sh} R_{sh} = I_m R_m$$

$$\therefore \quad 950\ \mu A\ (R_1 + R_2 + R_3 + R_4) = 50\ \mu A \times 100.$$

$$\therefore \quad R_1 + R_2 + R_3 + R_4 = \frac{50\ \mu A \times 100}{950\ \mu A} = \frac{5000}{950} = 5.26\ \Omega \tag{3.1}$$

For 0 – 10 mA

$$9950\ \mu A\ (R_1 + R_2 + R_3) = 50\ \mu A \cdot (100 + R_4) \tag{3.2}$$

For 0 – 50 mA

$$49950\ \mu A\ (R_1 + R_2) = 50\ \mu A \cdot (100 + R_3 + R_4) \tag{3.3}$$

For 0 – 100 mA

$$99950\ \mu A\ (R_1) = 50\ \mu A\ (100 + R_2 + R_3 + R_4) \tag{3.4}$$

But $R_1 + R_2 + R_3 = 5.26 - R_4$. Substituting in Eq. 3.2, we have

$$9950\ \mu A\ (5.26 - R_4) = 50\ \mu A\ (100 + R_4)$$

$$9950\ \mu A \times 5.26 - 9950\ \mu A \times R_4 = 5000\ \mu A + 50\ \mu A\ R_4$$

$$(9950\ \mu A \times 5.26 - 5000\ \mu A) = 9950\ \mu A\ R_4 + 50\ \mu A\ R_4$$

Therefore $$R_4 = \frac{9950\ \mu A \times 5.26 - 5000\ \mu A}{10\ mA} = \frac{47377\ \mu A}{10\ mA} = 4.737\ \Omega$$

$R_4 = 4.74\ \Omega$

In Eq. 3.1, substituting for R_4 we get

$$R_1 + R_2 + R_3 = 5.26 - 4.74 = 0.52$$

$$\therefore \quad R_1 + R_2 = 0.52 - R_3$$

Substituting in Eq. 3.3, we have

$$49950\ \mu\text{A}\ (0.52 - R_3) = 50\ \mu\text{A}\ (R_3 + 4.74 + 100)$$

$$49950\ \mu\text{A} \times 0.52 - 49950\ \mu\text{A} \times R_3 = 50\ \mu\text{A} \times R_3 + 50\ \mu\text{A} \times 4.74 + 50\ \mu\text{A} \times 100$$

$$49950\ \mu\text{A} \times 0.52 - 50\ \mu\text{A} \times 4.74 = 49950\ \mu\text{A} \times R_3 + 50\ \mu\text{A} \times R_3 + 5000\ \mu\text{A}$$

$$(25974 - 237)\ \mu\text{A} = 50\ \text{mA} \times R_3 + 5000\ \mu\text{A}$$

$$25737\ \mu\text{A} = 50\ \text{mA} \times R_3 + 5000\ \mu\text{A}$$

$$R_3 = \frac{25737\ \mu\text{A} - 5000\ \mu\text{A}}{50\ \text{mA}} = \frac{20737\ \mu\text{A}}{50\ \text{mA}}$$

$$R_3 = 0.4147 = 0.42\ \Omega$$

But $\quad R_1 + R_2 = 0.52 - R_3$

$\therefore \quad R_1 + R_2 = 0.52 - 0.4147 = 0.10526$

Therefore $\quad R_2 = 0.10526 - R_1 \qquad (3.5)$

From Eq. 3.4

$$99950\ \mu\text{A}\ (R_1) = 50\ \mu\text{A} \times (100 + R_2 + R_3 + R_4)$$

But $\quad R_2 + R_3 + R_4 = 5.26 - R_1 \qquad$ (from Eq. 3.1)

Substituting in Eq. 3.4

$$99950\ \mu\text{A} \times R_1 = 50\ \mu\text{A} \times (100 + 5.26 - R_1)$$

$$99950\ \mu\text{A} \times R_1 = 5000\ \mu\text{A} + (50\ \mu\text{A} \times 5.26) - (R_1 \times 50\ \mu\text{A})$$

$$99950\ \mu\text{A} \times R_1 + 50\ \mu\text{A} \times R_1 = 5000\ \mu\text{A} + 50\ \mu\text{A} \times 5.26$$

$$(99950\ \mu\text{A} + 50\ \mu\text{A})\ R_1 = 5000\ \mu\text{A} + 263\ \mu\text{A}$$

$$100\ \text{mA} \times R_1 = 5263\ \mu\text{A}$$

$$R_1 = \frac{5263\ \mu\text{A}}{100\ \text{mA}} = 0.05263$$

Therefore $\quad R_1 = 0.05263\ \Omega$

From Eq. 3.5, we have

$$R_2 = 0.10526 - R_1 = 0.10526 - 0.05263 = 0.05263\ \Omega$$

Hence the value of shunts are

$$R_1 = 0.05263\ \Omega\ ; R_2 = 0.05263\ \Omega$$

$$R_3 = 0.4147\ \Omega\ ;\ R_4 = 4.74\ \Omega$$

Example 3.4 (b) *Calculate the value of the shunt resistors for the circuit shown below.*

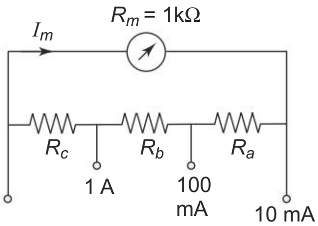


Fig. 3.4(b) For Example 3.4(b)

Solution The total shunt resistance R_{sh} is determined by

$$R_{sh} = \frac{R_m}{(n-1)} \quad \text{where } n = I/I_m$$

Given I_m = 100 μA and R_m = 1000 Ω

Step 1: For 10 mA range:

$$n = \frac{I}{I_m} = \frac{10 \text{ mA}}{100 \text{ μA}} = 100$$

$$R_{sh} = \frac{R_m}{(n-1)} = \frac{1000\ \Omega}{100-1} = \frac{1000}{99} = 10.1\ \Omega$$

Step 2: When the meter is set on the 100 mA range, the resistance R_b and R_c provides the shunt.

The shunt can be found from the equation

$$R_{sh_2} = (R_b + R_c) = \frac{I_m(R_m + R_{sh})}{I} = \frac{100 \text{ μA } (10.1 + 1000)}{100 \text{ mA}} = 1.01\ \Omega$$

Step 3: The resistor which provides the shunt resistance on the 1A range can be found from the equation

$$R_c = \frac{I_m(R_m + R_{sh})}{I} = \frac{100 \text{ μA } (10.1 + 1000)}{1000 \text{ mA}} = 0.101\ \Omega$$

Step 4: But $R_b + R_c$ = 1.01 Ω

$$R_b = 1.01 - R_c = 1.01 - 0.101\ \Omega = 0.909\ \Omega$$

Step 5: Resistor R_a is found by

$$R_a = R_{sh} - (R_b + R_c) = 10.1 - (0.909 + .101)\ \Omega$$
$$= 10.1 - 1.01\ \Omega$$

$= 9.09\Omega$

Hence $R_a = 9.09\ \Omega$, $R_b = 0.909\ \Omega$ and $R_c = 0.101\ \Omega$

REQUIREMENTS OF A SHUNT 3.4

The type of material that should be used to join the shunts should have two main properties.

1. Minimum Thermo Dielectric Voltage Drop

Soldering of joint should not cause a voltage drop.

2. Solderability

Resistance of different sizes and values must be soldered with minimum change in value.

The following precautions should be observed when using an ammeter for measurement.

1. Never connect an ammeter across a source of emf. Because of its low resistance it would draw a high current and destroy the movement. Always connect an ammeter in series with a load capable of limiting the current.
2. Observe the correct polarity. Reverse polarity causes the meter to deflect against the mechanical stopper, which may damage the pointer.
3. When using a multirange meter, first use the highest current range, then decrease the current range until substantial deflection is obtained. To increase the accuracy use the range that will give a reading as near full scale as possible.

EXTENDING OF AMMETER RANGES 3.5

The range of an ammeter can be extended to measure high current values by using external shunts connected to the basic meter movement (usually the lowest current range), as given in Fig. 3.5.

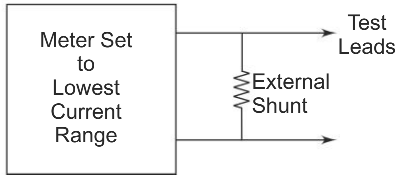


Fig. 3.5 Extending of ammeters

Note that the range of the basic meter movement cannot be lowered. (For example, if a 100 µA movement with 100 scale division is used to measure 1 µA, the meter will deflect by only one division. Hence ranges lower than the basic range are not practically possible.)

RF AMMETER (THERMOCOUPLE) 3.6

3.6.1 Thermocouple Instruments

Thermocouples consists of a junction of two dissimilar wires, so chosen that a voltage is generated by heating the junction. The output of a thermocouple is delivered to a sensitive dc microammeter.

(Calibration is made with dc or with a low frequency, such as 50 cycles, and applies for all frequencies for which the skin effect in the heater is not appreciable. Thermocouple instruments are the standard means for measuring current at radio frequencies.)

The generation of dc voltage by heating the junction is called thermoelectric action and the device is called a thermocouple.

3.6.2 Different Types of Thermocouples

In a thermocouple instrument, the current to be measured is used to heat the junction of two metals. These two metals form a thermocouple and they have the property that when the junction is heated it produces a voltage proportional to the heating effect. This output voltage drives a sensitive dc microammeter, giving a reading proportional to the magnitude of the ac input.

The alternating current heats the junction; the heating effect is the same for both half cycles of the ac, because the direction of potential drop (or polarity) is always be the same. The various types of thermocouples are as follows.

Mutual Type (Fig. 3.6 (a)) In this type, the alternating current passes through the thermocouple itself and not through a heater wire. It has the disadvantages that the meter shunts the thermocouple.

Contact Type (Fig. 3.6 (b)) This is less sensitive than the mutual type. In the contact type there are separate thermocouple leads which conduct away the heat from the heater wire.

Separate Heater Type (Fig. 3.6 (c)) In this arrangement, the thermocouple is held near the heater, but insulated from it by a glass bead. This makes the instrument sluggish and also less sensitive because of temperature drop in the glass bead. The separate type is useful for certain applications, like RF current measurements. To avoid loss of heat by radiation, the thermocouple arrangement is placed in a vacuum in order to increase its sensitivity.

Bridge Type (Fig. 3.6 (d)) This has the high sensitivity of the mutual type and yet avoids the shunting effect of the microammeter.

The sensitivity of a thermocouple is increased by placing it in a vacuum since loss of heat by conduction is avoided, and the absence of oxygen permits operation at a much higher temperature. A vacuum thermocouple can be designed to give a full scale deflection of approximately 1 mA. A similar bridge arrangement in air would require about 100 mA for full scale deflection.

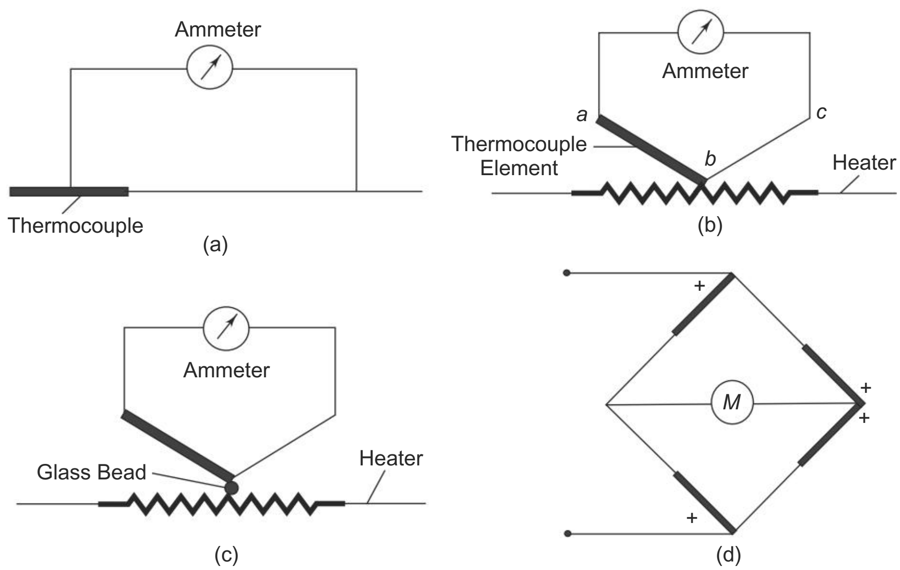


Fig. 3.6 (a) Mutual type (b) Contact type (c) Separate heater type (d) Bridge type thermocouple

Material commonly used to form a thermocouple are constantan against copper, manganin or a platinum alloy. Such a junction gives a thermal emf of approximately 45 μV/°C.

The heating element of open air heaters is typically a non-corroding platinum alloy. Carbon filament heaters are used in vacuum type.

Thermocouple heaters operate so close to the burnout point under normal conditions, that they can withstand only small overloads without damage, commonly up to 50%. This is one of the limitations of the thermocouple instrument.

(Commonly used metal combinations are copper-constantan, iron-constantan, chromel-constantan, chromel-alumel, and platinum-rhodium. Tables are available that show the voltages produced by each of the various metal combination at specific temperatures.)

LIMITATIONS OF THERMOCOUPLES 3.7

Following are the limitations of thermocouples

1. Heaters can stand only small overload.
2. A rise in temperature (higher operating temperatures) causes a change in the resistance of the heater.
3. Presence of harmonics changes meter reading, because the heating effect is proportional to the square of current.

This can be understood by the following example.

The effective value of input wave is

$$= \sqrt{I_1^2 + I_2^2 + I_3^2 + I_4^2 + \cdots}$$

where I_1 is the fundamental
I_2 is second harmonic
I_3 is third harmonic

If 20% harmonics are present, then $I_2 = \dfrac{I_1}{5}$.

Therefore, the error in the current reading if 20% harmonics is present, is calculated as follows. Therefore, effective value of input wave

$$= \sqrt{I_1^2 + I_2^2} = \sqrt{I_1^2 + \left(\frac{I_1}{5}\right)^2}$$

$$= \sqrt{I_1^2 + \frac{I_1^2}{25}} = \sqrt{\frac{26}{25} I_1^2} = \sqrt{1.04\, I_1^2}$$

$$= 1.02\, I_1 = I_1 + 0.02\, I_1$$

But 0.02 = 2%. Hence 20% harmonics increase the error by 2%.

EFFECT OF FREQUENCY ON CALIBRATION 3.8

The frequency effect arises because of various factors such as:

1. Skin effect
2. Non uniform distribution of current along the heater wires
3. Spurious capacitive currents

1. Skin Effect The skin effect causes a higher reading at higher frequencies, especially if the heater wire is small. A low current instrument with a circular cross-section, used in vacuum, may have a skin effect error of less than 1% at frequencies up to 30,000 MHz. Ribbon heaters are often used for large currents, but they have larger skin effects. Solid wire, and better still hollow conductors are ideal with a view to minimising the skin effect.

Calibration done with dc or low frequency as such as 50 Hz for which the skin effect of the heater is not appreciable. Accuracy can be as high as 1% for frequencies up to 50 MHz. For this reason, thermocouple instruments are classified as RF instruments.

Above 50 MHz the skin effect forces the current to the outer surface of the conductor, increasing the effective resistance of the heating wire and reducing the instrument's accuracy. For small currents of up to 3 A, the heating wire should be solid and very thin. Above 3 A the heating element should be hollow and tubular in design to reduce the skin effect.

2. Non-uniform Distribution of Current This occurs at frequencies where the heater length is of the order of a fraction of a wavelength (magnitude of one wavelength).

The current distribution along the heater is not uniform and the meter indication is uncertain. Hence to avoid this the heater length and its associated leads should be less than 1/10th of a wavelength.

3. Spurious Capacitive Currents These occur when the thermocouple instrument is connected in such a manner that both terminals are at a potential above ground. As the frequency is increased, a large current flows through the capacitance formed by the thermocouple leads, with the meter acting as one electrode and the ground as the other. To avoid this, proper shielding of the instrument should be provided.

The calibration of a thermocouple is reasonably permanent. When calibrating Contact and Mutual with dc, it is always necessary to reverse the polarity to take the average reading. This is because of the resistance drop in the heater at the contact may cause a small amount of dc-current to flow; reversing the calibrating current averages out this effect.

MEASUREMENTS OF VERY LARGE CURRENTS BY THERMOCOUPLES 3.9

Thermocouples instruments with heaters large enough to carry very large currents may have an excessive skin effect. Ordinary shunts cannot be used because the shunting ratio will be affected by the relative inductance and resistance, resulting in a frequency effect.

One solution to this problem consists of minimising the skin effect by employing a heater, which is a tube of large diameter, but with very thin walls.

Another consists of employing an array of shunts of identical resistance arranged symmetrically as shown in Fig. 3.7 (a).

In Fig. 3.7 (a) each filament of wire has the same inductance, so that the inductance causes the current to divide at high frequencies, in the same way as does the resistance at low frequencies. In Fig. 3.7 (b) the condensor shunt is used such that the current divides between the two parallel capacitors proportional to their capacitance, and maintains this ratio independent of frequency, as long as the capacitor that is in series with the thermocouple has a higher impedance than the thermocouple heater and the lead inductance is inversely proportional to the capacitances.

In Fig. 3.7 (c) the current transformer is used to measure very large RF currents at low and moderate frequencies using a thermocouple instrument of ordinary range. Such transformers generally use a magnetic dust core. The current ratio is given by

$$\frac{\text{Primary Current}}{\text{Secondary Current}} = \frac{1}{K}\sqrt{\frac{L_s}{L_p}}\sqrt{1+\frac{1}{Q_s}}$$

where L_s = secondary inductance
L_p = primary inductance
K = coefficient of coupling between Lp and Ls
r_s = resistance of secondary, including meter resistance

$Q_s = \omega L_s / r_s = Q$ of the secondary circuit taking into account meter resistance

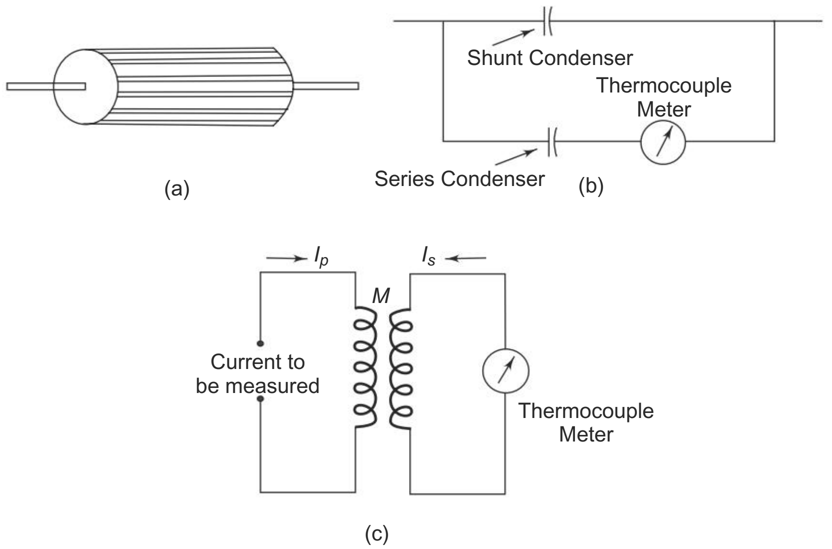


Fig. 3.7 (a) Array of shunts (b) Condenser shunt (c) Current transformer

If Q of the secondary winding is appreciable (i.e. greater than 5), the transformation ratio is independent of frequency.

A current ratio of 1000 or more can be obtained at low and moderate RF by using a many turn secondary wound on a toroidal ring.

Review Questions

1. Explain with a diagram how a PMMC can be used as an ammeter.
2. What are the requirements of a shunt? How can a basic ammeter be converted into a multirange ammeter?
3. What are the limitations of a multi-range ammeter. How is it overcome?
4. State the precautions to be observed when using an ammeter.
5. Explain with a diagram the operation of an ayrton shunt.
6. State the advantages of an Ayrton shunt ammeter over a multirange ammeter.
7. How is current in the RF range measured?
8. Why is a thermocouple used in RF measurement of current?
9. Explain the construction and working of a thermocouple measuring instrument. State the limitations of a thermocouple instruments.
10. Why is a thermocouple measuring instrument classified as an RF instrument?
11. State different types of thermocouples used for current measurement. Explain each one in brief.
12. How is a large current measured using a thermocouple?
13. What are the effects of frequency on the calibration of a thermocouple?
14. Explain with a diagram how a current transformer can be used to measure large RF currents.

Multiple Choice Questions

1. The instrument required to measure current is a/an
   (a) voltmeter (b) ammeter
   (c) wattmeter (d) ohmmeter
2 A D'Arsonval movement is
   (a) taut band
   (b) p1mmc
   (c) electrodynamometer
   (d) moving iron type
3. To select the range, a multirange ammeter uses a
   (a) double pole double throw switch
   (b) make before break type switch
   (c) single pole double throw switch
   (d) simple switch
4. To select a range , the Aryton shunt uses a
   (a) double pole double throw switch
   (b) make before break type switch
   (c) single pole double throw switch
   (d) simple switch
5. Current in the RF range is measured by
   (a) simple ammeter
   (b) ammeter using thermocouples.
   (c) multirange ammeters.
   (d) aryton shunt.
6. Large current in RF range at low moderate frequencies is measured by
   (a) simple ammeter
   (b) ammeter using thermocouples.
   (c) using a current transformer
   (d) using Aryton shunt.
7. To minimize skin effect at high RF range
   (a) inductance is used
   (b) array of Shunts are used
   (c) dielectric material is used
   (d) aryton shunt is used
8. At low and moderate RF using a secondary wound on a torridal ring, a current ratio is obtained.
   (a) 500 (b) 1000
   (c) 2000 (d) 5000

Practice Problems

1. Calculate the value of shunt resistance required for using a 50 µA meter movement having an internal resistance of 100 Ω for measuring current in the range of 0–250 mA.
2. What value of shunt resistance is required for using 50 µA meter movement having an internal resistance of 250 Ω for measuring current in the range of 0–500 mA?
3. Design a multirange ammeter with ranges of 0–100 mA, 0–200 mA, 0–500 mA, 0–1 A employing individual shunts for each range. A D'Arsonval movement with an internal resistance of 500 Ω and a full scale current of 100 µA is available.
4. Design a multirange ammeter with ranges of 0–1 A, 5 A, 25 A, 125 A employing individual shunts for each range. A D'Arsonval movement with an internal resistance of 730 Ω and a full scale current of 5 mA is available
5. Design an Ayrton shunt to provide an ammeter with current ranges of 0–1 mA, 5 mA, 20 mA and 50 mA, using a D'Arsonval movement having internal resistance of 50 Ω and a full scale current of 100 µA.
6. Design an Ayrton shunt to provide an ammeter with current ranges of 0–1 mA, 10 mA, 50 mA and 100 mA, using a D'Arsonval movement having internal resistance of 100 Ω and a full scale current of 50 µA.
7. Design an Ayrton shunt to provide an ammeter with current ranges of 0–100 mA, 500 mA, 1A, using a D'Arsonval movement having internal resistance of 50 Ω and a full scale current of 1mA.

Further Reading

1. Terman and Petit, *Electronic Measurements*, McGraw-Hill Book Co, New York, 1952.
2. Sol. D. Prensky, *Electronic Instrumentation*, Prentice-Hall of India, 1963.
3. John. H. Fasal, *Simplified Electronic Measurements*, Hayden Book Co. Inc., Mumbai, 1971.
4. Larry. D. Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley and Sons, New York, 1987.
5. W.D. Copper and A.D. Helfrick, *Electronic Instrumentation and Measurements Techniques*, 3rd Edition, Prentice-Hall of India, 1985.

Voltmeters and Multimeters

Chapter 4

INTRODUCTION 4.1

The most commonly used dc meter is based on the fundamental principle of the motor. The motor action is produced by the flow of a small amount of current through a moving coil which is positioned in a permanent magnetic field. This basic moving system, often called the D'Arsonval movement, is also referred to as the basic meter.

Different instrument forms may be obtained by starting with the basic meter movement and adding various elements, as follows.

1. The basic meter movement becomes a dc instrument, measuring
   (i) dc current, by adding a shunt resistance, forming a microammeter, a milliammeter or an ammeter.
   (ii) dc voltage, by adding a multiplier resistance, forming a milli voltmeter, voltmeter or kilovoltmeter.
   (iii) resistance, by adding a battery and resistive network, forming an ohmmeter.
2. The basic meter movement becomes an ac instrument, measuring
   (i) ac voltage or current, by adding a rectifier, forming a rectifier type meter for power and audio frequencies.
   (ii) RF voltage or current, by adding a thermocouple-type meter for RF.
   (iii) Expanded scale for power line voltage, by adding a thermistor in a resistive bridge network, forming an expanded scale (100 – 140 V) ac meter for power line monitoring.

BASIC METER AS A DC VOLTMETER 4.2

To use the basic meter as a dc voltmeter, it is necessary to know the amount of current required to deflect the basic meter to full scale. This current is known as full scale deflection current (I_{fsd}). For example, suppose a 50 µA current is required for full scale deflection.

This full scale value will produce a voltmeter with a sensitivity of 20,000 Ω per V.

The sensitivity is based on the fact that the full scale current of 50 µA results whenever 20,000 Ω of resistance is present in the meter circuit for each voltage applied.

$$\text{Sensitivity} = 1/I_{fsd} = 1/50\ \mu\text{A} = 20\ \text{k}\Omega/\text{V}$$

Hence, a 0 – 1 mA would have a sensitivity of 1 V/1 mA = 1 kΩ/V or 1000 Ω.

Example 4.1 *Calculate the sensitivity of a 200 μA meter movement which is to be used as a dc voltmeter.*

Solution The sensitivity

$$S = \frac{1}{(I_{fsd})} = \frac{1}{200\ \mu\text{A}}$$

Therefore $S = 5\ \text{k}\Omega/\text{V}$

DC VOLTMETER 4.3

A basic D'Arsonval movement can be converted into a dc voltmeter by adding a series resistor known as multiplier, as shown in Fig. 4.1. The function of the multiplier is to limit the current through the movement so that the current does not exceed the full scale deflection value. A dc voltmeter measures the potential difference between two points in a dc circuit or a circuit component.

To measure the potential difference between two points in a dc circuit or a circuit component, a dc voltmeter is always connected across them with the proper polarity.

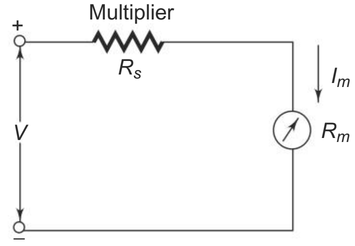


Fig. 4.1 Basic dc voltmeter

The value of the multiplier required is calculated as follows. Referring to Fig. 4.1,

I_m = full scale deflection current of the movement (I_{fsd})
R_m = internal resistance of movement
R_s = multiplier resistance
V = full range voltage of the instrument

From the circuit of Fig. 4.1

$$V = I_m (R_s + R_m)$$

$$R_s = \frac{V - I_m R_m}{I_m} = \frac{V}{I_m} - R_m$$

Therefore
$$R_s = \frac{V}{I_m} - R_m$$

The multiplier limits the current through the movement, so as to not exceed the value of the full scale deflection I_{fsd}.

The above equation is also used to further extend the range in DC voltmeter.

Example 4.2 (a) *A basic D'Arsonval movement with a full scale deflection of 50 μA and internal resistance of 500 Ω is used as a voltmeter. Determine the value of the multiplier resistance needed to measure a voltage range of 0 – 10 V.*

Solution Given

$$R_s = \frac{V}{I_m} - R_m = \frac{10}{50\ \mu\text{A}} - 500$$
$$= 0.2 \times 10^6 - 500 = 200\text{ k} - 500$$
$$= 199.5\text{ k}\Omega$$

Example 4.2 (b) *Calculate the value of multiplier resistance on the 50V range of a dc voltmeter that uses a 500 μA meter movement with an internal resistance of 1 kΩ.*

Solution

Step 1: The sensitivity of 500 μA meter movement is given by

$$S = 1/I_m = 1/500\ \mu\text{A} = 2\text{ k}\Omega/\text{V}.$$

Step 2: The value of the multiplier resistance can be calculated by

$$R_s = S \times \text{range} - R_m$$
$$R_s = 2\text{ k}\Omega/\text{V} \times 50\text{ V} - 1\text{ k}\Omega$$
$$= 100\text{ k}\Omega - 1\text{ k}\Omega = 99\text{ k}\Omega$$

MULTIRANGE VOLTMETER 4.4

As in the case of an ammeter, to obtain a multirange ammeter, a number of shunts are connected across the movement with a multi-position switch. Similarly, a dc voltmeter can be converted into a multirange voltmeter by connecting a number of resistors (multipliers) along with a range switch to provide a greater number of workable ranges.

Figure 4.2 shows a multirange voltmeter using a three position switch and three multipliers R_1, R_2, and R_3 for voltage values V_1, V_2, and V_3.

Figure 4.2 can be further modified to Fig. 4.3, which is a more practical arrangement of the multiplier resistors of a multirange voltmeter.

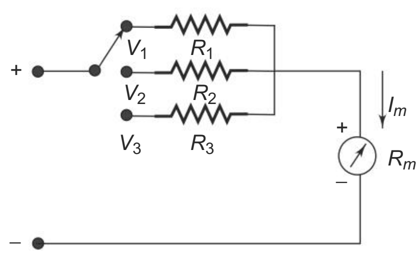


Fig. 4.2 Multirange voltmeter

In this arrangement, the multipliers are connected in a series string, and the range selector selects the appropriate amount of resistance required in series with the movement.

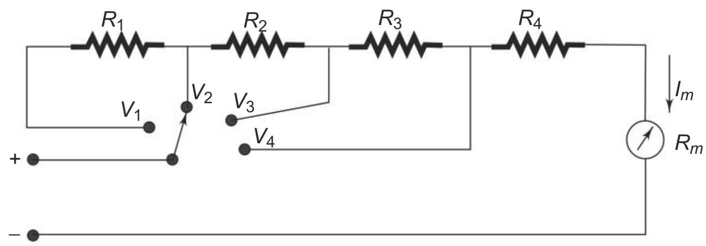


Fig. 4.3 Multipliers connected in series string

This arrangement is advantageous compared to the previous one, because all multiplier resistances except the first have the standard resistance value and are also easily available in precision tolerances.

The first resistor or low range multiplier, R_4, is the only special resistor which has to be specially manufactured to meet the circuit requirements.

Example 4.3 *A D'Arsonval movement with a full scale deflection current of 50 μA and internal resistance of 500 Ω is to be converted into a multirange voltmeter. Determine the value of multiplier required for 0–20 V, 0–50 V and 0–100 V.*

Solution Given I_m = 50 μA and R_m = 500 Ω

Case 1: For range 0 – 20V

$$R_s = \frac{V}{I_m} - R_m = \frac{20}{50 \times 10^{-6}} - 500 = 0.4 \times 10^6 - 500 = 400\text{ K} - 500 = 399.5\text{ k}\Omega$$

Case 2: For range 0 – 50 V

$$R_s = \frac{V}{I_m} - R_m = \frac{50}{50 \times 10^{-6}} - 500 = 1 \times 10^6 - 500 = 1000\text{ K} - 500 = 999.5\text{ k}\Omega$$

Case 3: For range 0 – 100 V

$$R_s = \frac{V}{I_m} - R_m = \frac{100}{50 \times 10^{-6}} - 500 = 2 \times 10^6 - 500 = 2000\text{ K} - 500 = 1999.5\text{ k}\Omega$$

Example 4.4 *A D'Arsonval movement with a full scale deflection current of 10 mA and internal resistance of 500 Ω is to be converted into a multirange voltmeter. Determine the value of multiplier required for 0–20 V, 0–50 V and 0–100 V.*

Solution Given I_m = 10 mA and R_m = 500 Ω

Case 1: For range 0 – 20V

$$R_s = \frac{V}{I_m} - R_m = \frac{20}{10 \times 10^{-3}} - 500 = 2 \times 10^3 - 500 = 2000 - 500 = 1.5\text{ k}\Omega$$

Case 2: For range 0 – 50V

$$R_s = \frac{V}{I_m} - R_m = \frac{50}{10 \times 10^{-3}} - 500 = 5 \times 10^3 - 500 = 5000 - 500 = 4.5\ \text{k}\Omega$$

Case 3: For range 0 – 100V

$$R_s = \frac{V}{I_m} - R_m = \frac{100}{10 \times 10^{-3}} - 500 = 10 \times 10^3 - 500 = 10\text{K} - 500 = 9.5\ \text{k}\Omega$$

Example 4.5 *Convert a basic D'Arsonval movement with an internal resistance of 100 Ω and a full scale deflection of 10 mA into a multirange dc voltmeter with ranges from 0 – 5 V, 0 – 50 V and 0 – 100 V.*

Solution Given $I_m = 10$ mA, $R_m = 100\ \Omega$

Step 1: For a 5 V (V_3) the total circuit resistance is

$$R_t = \frac{V}{I_{fsd}} = \frac{5}{10\text{ mA}} = 0.5\ \text{k}\Omega$$

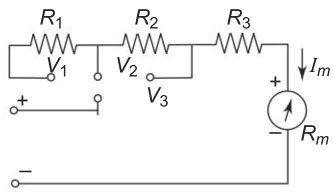


Fig. 4.3(a)

Therefore $R_3 = R_t - R_m = 500\ \Omega - 100\ \Omega$
$= 400\ \Omega$

Step 2: For a 50 V (V_2) position

$$R_t = \frac{V}{I_{fsd}} = \frac{50}{10\text{ mA}} = 5\ \text{k}\Omega$$

Therefore $R_2 = R_t - (R_3 + R_m) = 5\ \text{k}\Omega - (400\ \Omega + 100\ \Omega)$
$= 5\ \text{k}\Omega - 500\ \Omega = 4.5\ \text{K}\Omega$

Step 3: For a 100 V range (V_1) position

$$R_t = \frac{V}{I_{fsd}} = \frac{100}{10\text{ mA}} = 10\ \text{k}\Omega$$

Therefore $R_1 = R_t - (R_2 + R_3 + R_m) = 10\ \text{k}\Omega - (4.5\ \text{k}\Omega + 400\ \Omega + 100\ \Omega)$
$= 10\ \text{k}\Omega - 5\ \Omega = 5\ \text{k}\Omega$

Hence it can be seen that R_3 has a non-standard value.

Example 4.6 *Convert a basic D'Arsonval movement with an internal resistance of 50 Ω and a full scale deflection current of 2 mA into a multirange dc voltmeter with voltage ranges of 0 – 10 V, 0 – 50 V, 0 – 100 V and 0 – 250 V. Refer to Fig. 4.3.*

Solution For a 10 V range (V_4 position of switch), the total circuit resistance is

$$R_t = \frac{V}{I_{fsd}} = \frac{10}{2\text{ mA}} = 5\ \text{k}\Omega$$

Therefore $R_4 = R_t - R_m = 5\text{ k} - 50 = 4950\ \Omega.$

For 50 V range (V_3 position of switch), the total circuit resistance is

$$R_t = \frac{V}{I_{fsd}} = \frac{50}{2\,\text{mA}} = 25\ \text{k}\Omega$$

Therefore $R_3 = R_t - (R_4 + R_m) = 25\ \text{k} - (4950 + 50) = 25\ \text{k} - 5\ \text{k}$

$\therefore$ $R_3 = 20\ \text{k}\Omega$

For 100 V range (V_2 position of switch), the total circuit resistance is

$$R_t = \frac{V}{I_{fsd}} = \frac{100}{2\,\text{mA}} = 50\ \text{k}\Omega$$

Therefore,
$$R_2 = R_t - (R_3 + R_4 + R_m)$$
$$= 50\ \text{k} - (20\ \text{k} + 4950 + 50)$$

$\therefore$ $R_2 = 50\ \text{k} - 25\ \text{k} = 25\ \text{k}\Omega$

For 250 V range, (V_1 position of switch), the total circuit resistance is

$$R_t = \frac{V}{I_{fsd}} = \frac{250}{2\,\text{mA}} = 125\ \text{k}\Omega$$

Therefore
$$R_1 = R_t - (R_2 + R_3 + R_4 + R_m)$$
$$= 125\ \text{k} - (25\ \text{k} + 20\ \text{k} + 4950 + 50)$$
$$= 125\ \text{k} - 50\ \text{k}$$
$$= 75\ \text{k}\Omega$$

Only the resistance R_4 (low range multiplier) has a non-standard value.

EXTENDING VOLTMETER RANGES 4.5

The range of a voltmeter can be extended to measure high voltages, by using a high voltage probe or by using an external multiplier resistor, as shown in Fig. 4.4. In most meters the basic movement is used on the lowest current range. Values for multipliers can be determined using the procedure of Section 4.4.

The basic meter movement can be used to measure very low voltages. However, great care must be used not to exceed the voltage drop required for full scale deflection of the basic movement.

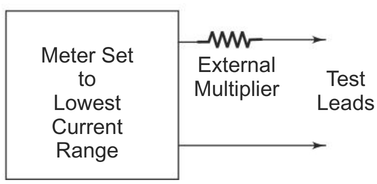


Fig. 4.4 Extending voltage range

Sensitivity The sensitivity or Ohms per Volt rating of a voltmeter is the ratio of the total circuit resistance R_t to the voltage range. Sensitivity is essentially the reciprocal of the full scale deflection current of the basic movement. Therefore, $S = 1/I_{fsd}\ \Omega/V$.

The sensitivity 'S' of the voltmeter has the advantage that it can be used to calculate the value of multiplier resistors in a dc voltmeter. As,

R_t = total circuit resistance $[R_t = R_s + R_m]$
S = sensitivity of voltmeter in ohms per volt
V = voltage range as set by range switch
R_m = internal resistance of the movement

Since $R_s = R_t - R_m$ and $R_t = S \times V$

$\therefore$ $R_s = (S \times V) - R_m$

Example 4.7 *Calculate the value of the multiplier resistance on the 50 V range of a dc voltmeter, that uses a 200 μA meter movement with an internal resistance of 100 Ω.*

Solution As $R_s = S \times$ Range – internal resistance, and $S = 1/I_{fsd}$.

$\therefore$ The sensitivity of the meter movement is

$$S = 1/I_{fsd} = 1/200\ \mu A = 5\ k\Omega/V.$$

The value of multiplier R_s is calculated as

$$R_s = S \times \text{Range} - \text{internal resistance} = S \times V - R_m$$
$$= 5\ k \times 50 - 100$$
$$= 250\ k - 100$$
$$= 249.9\ k\Omega$$

Example 4.8 *Calculate the value of multiplier resistance for the multiple range dc voltmeter circuit shown in Fig. 4.5 (a).*

Solution The sensitivity of the meter movement is given as follows.

$$S = 1/I_{fsd} = 1/50\ \mu A = 20\ k\Omega/V$$

The value of the multiplier resistance can be calculated as follows.

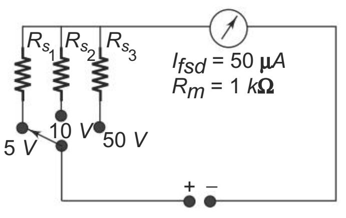


Fig. 4.5 (a)

For 5 V range

$$R_{s_1} = S \times V - R_m$$
$$= 20\ k \times 5 - 1\ k$$
$$= 100\ k - 1\ k = 99\ k\Omega$$

For 10 V range

$$R_{s_2} = S \times V - R_m$$
$$= 20\ k \times 10 - 1\ k$$
$$= 200\ k - 1\ k = 199\ k\Omega$$

For 50 V range

$$R_{s_3} = S \times V - R_m$$
$$= 20\ k \times 50 - 1\ k$$
$$= 1000\ k - 1\ k = 999\ k\Omega$$

Example 4.9 *Calculate the value of multiplier resistance for the multirange dc voltmeter as shown in Fig 4.5(b).*

Solution

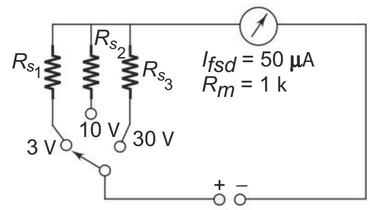


Fig. 4.5 (b)

Step 1: The sensitivity of 50 μA meter movement is given by

$S = 1/I_m = 1/50\ \mu\text{A} = 20\ \text{k}\Omega/\text{V}$.

The value of the multiplier resistance can be calculated by

Step 2: The value of the multiplier for 3 V range

$$R_s = S \times \text{range} - R_m$$
$$R_s = 20\ \text{k}\Omega/\text{V} \times 3\text{V} - 1\ \text{k}\Omega$$
$$= 60\ \text{k}\Omega - 1\ \text{k}\Omega = 59\ \text{k}\Omega.$$

Step 3: The value of the multiplier resistance For 10 V range can be calculated by

$$R_s = S \times \text{range} - R_m$$
$$R_s = 20\ \text{k}\Omega./\text{V} \times 10\ \text{V} - 1\ \text{k}\Omega$$
$$= 200\ \text{k}\Omega - 1\ \text{k}\Omega = 199\ \text{k}\Omega.$$

Step 4: The value of the multiplier resistance For 30V range can be calculated by

$$R_s = S \times \text{range} - R_m$$
$$R_s = 20\ \text{k}\Omega/\text{V} \times 30\ \text{V} - 1\ \text{k}\Omega$$
$$= 600\ \text{k}\Omega - 1\ \text{k}\Omega = 599\ \text{k}\Omega$$

Example 4.10 *A moving coil instrument gives a full scale deflection of 20 mA when the potential difference across its terminals is 100 mV. Calculate*

(a) Shunt resistance for a full scale deflection corresponding of 50 A.

(b) The series resistance for a full scale reading with 500 V. Also calculate the power dissipation in each case.

Solution Given meter current I_m = 20 mA and voltage = 100 mV

Step 1: Meter resistance $R_m = \dfrac{100\ \text{mV}}{20\ \text{mA}} = 5\ \Omega$

Step 2: Shunt resistance is given by

$$R_{sh} = \frac{I_m R_m}{I - I_m} = \frac{20\ \text{mA} \times 5\ \Omega}{50000\ \text{mA} - 20\ \text{mA}} = \frac{100\ \text{mA}}{49980\ \text{mA}} = .002\ \Omega$$

Step 3: Voltage Multiplier

$$R_{sh} = \frac{V}{I_m} - R_m = \frac{500\ \text{V}}{20\ \text{mA}} - 5\ \Omega = 25 \times 10^3 - 5\ \Omega$$
$$= 24995\ \Omega \approx 25\ \text{k}\Omega$$

Power = $V_m \cdot I_m$ = 500 × 20 mA = 10 W

LOADING 4.6

When selecting a meter for a certain voltage measurement, it is important to consider the sensitivity of a dc voltmeter. A low sensitivity meter may give a correct reading when measuring voltages in a low resistance circuit, but it is certain to produce unreliable readings in a high resistance circuit. A Voltmeter when connected across two points in a highly resistive circuits, acts as a shunt for that portion of the circuit, reducing the total equivalent resistance of that portion as shown in Fig. 4.6. The meter then indicates a lower reading than what existed before the meter was connected. This is called the loading effect of an instrument and is caused mainly by low sensitivity instruments.

Example 4.11 *Figure 4.6 shows a simple series circuit of R_1 and R_2 connected to a 100 V dc source. If the voltage across R_2 is to be measured by voltmeters having*

(a) a sensitivity of 1000 Ω/V, and

(b) a sensitivity of 20,000 Ω/V, find which voltmeter will read the accurate value of voltage across R_2. Both the meters are used on the 50 V range.

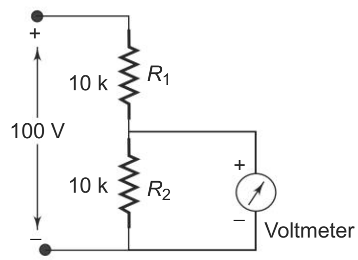


Fig. 4.6 Example on loading effect

Solution Inspection of the circuit indicates that the voltage across the R_2 resistance is

$$\frac{10\text{ k}}{10\text{ k} + 10\text{ k}} \times 100\text{ V} = 50\text{ V}$$

This is the true voltage across R_2.

Case 1

Using a voltmeter having a sensitivity of 1000 Ω/V.

It has a resistance of 1000 × 50 = 50 kΩ on its 50 V range.

Connecting the meter across R_2 causes an equivalent parallel resistance given by

$$R_{eq} = \frac{10\text{ k} \times 50\text{ k}}{10\text{ k} + 50\text{ k}} = \frac{500\text{ M}}{60\text{ k}} = 8.33\text{ k}\Omega$$

Now the voltage across the total combination is given by

$$V_1 = \frac{R_{eq}}{R_1 + R_{eq}} \times V$$

$$V_1 = \frac{8.33\text{ k}}{10\text{ k} + 8.33\text{ k}} \times 100\text{ V} = 45.43\text{ V}$$

Hence this voltmeter indicates 45.43 V.

Case 2

Using a voltmeter having a sensitivity of 20,000 Ω/V. Therefore it has a resistance of

$$20{,}000 \times 50 = 1000 \text{ k} = 1 \text{ M}\Omega$$

This voltmeter when connected across R_2 produces an equivalent parallel resistance given by

$$R_{eq} = \frac{10 \text{ k} \times 1 \text{ M}}{10 \text{ k} + 1 \text{ M}} = \frac{10^9}{1.01 \text{ M}} = \frac{10 \text{ k}}{1.01} = 9.9 \text{ k}\Omega$$

Now the voltage across the total combination is given by

$$V_2 = \frac{9.9 \text{ k}}{10 \text{ k} + 9.9 \text{ k}} \times 100 \text{ V} = 49.74 \text{ V}$$

Hence this voltmeter will read 49.74 V.

This example shows that a high sensitivity voltmeter should be used to get accurate readings.

Example 4.12 *Two different voltmeters are used to measure the voltage across R_b in the circuit of Fig. 4.7.*

The meters are as follows.

Meter 1: $S = 1$ kΩ/V, $R_m = 0.2$ k, range 10 V

Meter 2: $S = 20$ kΩ/V, $R_m = 1.5$ k, range 10 V

Calculate (i) voltage across R_b without any meter across it, (ii) voltage across R_b when the meter 1 is used (iii) voltage across R_b when the meter 2 is used, and (iv) error in the voltmeters.

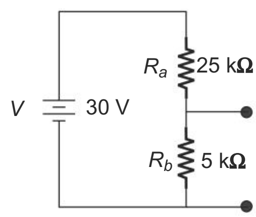


Fig. 4.7

Solution (i) The voltage across the resistance R_b, without either meter connected, is calculated using the voltage divider formula.

Therefore,
$$VR_b = \frac{5 \text{ k}}{25 \text{ k} + 5 \text{ k}} \times 30 = \frac{150 \text{ k}}{30 \text{ k}} = 5 \text{ V}$$

(ii) Starting with meter 1, having sensitivity $S = 1$ kΩ/V

Therefore the total resistance it presents to the circuit

$$R_{m_1} = S \times \text{range} = 1 \text{ k}\Omega/\text{V} \times 10 = 10 \text{ k}\Omega$$

The total resistance across R_b is, R_b in parallel with meter resistance R_{m1}

$$R_{eq} = \frac{R_b \times R_{m_1}}{R_b + R_{m_1}} = \frac{5 \text{ k} \times 10 \text{ k}}{5 \text{ k} + 10 \text{ k}} = 3.33 \text{ k}\Omega$$

Therefore, the voltage reading obtained with meter 1 using the voltage divider equation is

$$VR_b = \frac{R_{eq}}{R_{eq} + R_a} \times V = \frac{3.33\text{ k}}{3.33\text{ k} + 25\text{ k}} \times 30 = 3.53\text{ V}$$

(iii) The total resistance that meter 2 presents to the circuit is

$$R_{m_2} = S \times \text{range} = 20\text{ k}\Omega/\text{V} \times 10\text{ V} = 200\text{ k}\Omega$$

The parallel combination of R_b and meter 2 gives

$$R_{eq} = \frac{R_b \times R_{m_2}}{R_b + R_{m_2}} = \frac{5\text{ k} \times 200\text{ k}}{5\text{ k} + 200\text{ k}} = \frac{1000\text{ k} \times 1\text{ k}}{205\text{ k}} = 4.88\text{ k}\Omega$$

Therefore the voltage reading obtained with meter 2, using the voltage divider equation is

$$VR_b = \frac{4.88\text{ k}}{25\text{ k} + 4.88\text{ k}} \times 30 = \frac{4.88\text{ k}}{29.88\text{ k}} \times 30 = 4.9\text{ V}$$

(iv) The error in the reading of the voltmeter is given as:

$$\%\text{ Error} = \frac{\text{Actual voltage} - \text{Voltage reading observed in meter}}{\text{Actual voltage}} \times 100\%$$

$$\therefore \quad \text{voltmeter 1 error} = \frac{5\text{ V} - 3.33\text{ V}}{5\text{ V}} \times 100\% = 33.4\%$$

Similarly

$$\text{voltmeter 2 error} = \frac{5\text{ V} - 4.9\text{ V}}{5\text{ V}} \times 100\% = 2\%$$

Example 4.13 *Find the voltage reading and % error of each reading obtained with a voltmeter on (i) 5 V range, (ii) 10 V range and (iii) 30 V range, if the instrument has a 20 kΩ/V sensitivity and is connected across R_b of Fig. 4.8 (a).*

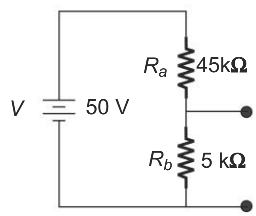


Fig. 4.8 (a)

Solution The voltage drop across R_b without the voltmeter connected is calculated using the voltage equation

$$VR_b = \frac{R_b}{R_a + R_b} \times V = \frac{5\text{ k}}{45\text{ k} + 5\text{ k}} \times 50 = \frac{50 \times 5\text{ k}}{50\text{ k}} = 5\text{ V}$$

On the 5 V range

$$R_m = S \times \text{range} = 20\text{ k}\Omega \times 5\text{ V} = 100\text{ k}\Omega$$

$$\therefore \quad R_{eq} = \frac{R_m \times R_b}{R_m + R_b} = \frac{100\text{ k} \times 5\text{ k}}{100\text{ k} + 5\text{ k}} = \frac{500\text{ k}}{105\text{ k}} = 4.76\text{ k}\Omega$$

The voltmeter reading is

$$VR_b = \frac{R_{eq}}{R_a + R_{eq}} \times V = \frac{4.76\ \text{k}}{45\ \text{k} + 4.76\ \text{k}} \times 50 = 4.782\ \text{V}$$

The % error on the 5 V range is

$$\%\ \text{Error} = \frac{\text{Actual voltage} - \text{Voltage reading in meter}}{\text{Actual voltage}}$$

$$= \frac{5\ \text{V} - 4.782\ \text{V}}{5\ \text{V}} \times 100 = \frac{0.217\ \text{V}}{5\ \text{V}} \times 100 = 4.34\ \%$$

On 10 V range

$$R_m = S \times \text{range} = 20\ \text{k}\Omega/\text{V} \times 10\ \text{V} = 200\ \text{k}\Omega$$

$$\therefore \quad R_{eq} = \frac{R_m \times R_b}{R_m + R_b} = \frac{200\ \text{k} \times 5\ \text{k}}{200\ \text{k} + 5\ \text{k}} = 4.87\ \text{k}\Omega$$

The voltmeter reading is

$$VR_b = \frac{R_{eq}}{R_{eq} + R_a} \times V = \frac{4.87\ \text{k}}{4.87\ \text{k} + 45\ \text{k}} \times 50 = 4.88\ \text{V}$$

The % error on the 10 V range = $\dfrac{5\ \text{V} - 4.88\ \text{V}}{5\ \text{V}} \times 100 = 2.34\%$

On 30 V range

$$R_m = S \times \text{range} = 20\ \text{k}\Omega/\text{V} \times 30\ \text{V} = 600\ \text{k}$$

$$\therefore \quad R_{eq} = \frac{R_m \times R_b}{R_m + R_b} = \frac{600\ \text{k} \times 5\ \text{k}}{600\ \text{k} + 5\ \text{k}} = \frac{3000\ \text{k} \times 1\ \text{k}}{605\ \text{k}} = 4.95\ \text{k}$$

The voltmeter reading on the 30 V range

$$VR_b = \frac{R_{eq}}{R_{eq} + R_a} \times V = \frac{4.95\ \text{k}}{45\ \text{k} + 4.95\ \text{k}} \times 50 = 4.95\ \text{V}$$

The % error on the 30 V range

$$= \frac{5\ \text{V} - 4.95\ \text{V}}{5\ \text{V}} \times 100 = \frac{0.05}{5\ \text{V}} \times 100 = 1\ \%$$

In the above example, the 30 V range introduces the least error due to loading. However, the voltage being measured causes only a 10% full scale deflection, whereas on the 10 V range the applied voltage causes approximately a one third of the full scale deflection with less than 3% error.

Example 4.14 *A current meter that has an internal resistance of 100 Ω is used to measure the current thro resistor R_3 in Fig 4.8(b) given below. Determine the % of the reading due to ammeter loading.*

Solution

Step 1: The current meter will be connected in to the circuit as shown in Fig 4.8 (a).

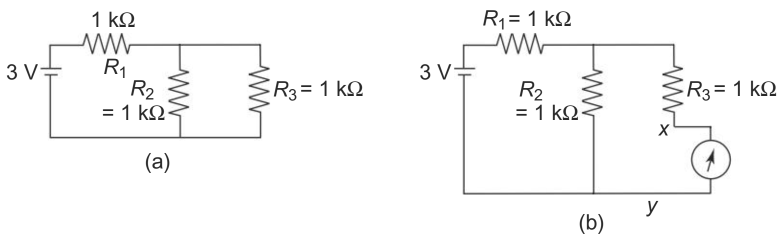


Fig. 4.8

Looking back into terminals x and y and using Thevinin's equivalent resistance,

$$R_t = R_1 + \frac{R_2 \times R_3}{R_2 + R_3} = 1\text{ k} + \frac{1\text{ k} \times 1\text{ k}}{1\text{ k}} = 1.5\text{ k}\Omega$$

Step 2: The ratio of the meter current to the expected current is

$$\frac{I_m}{I} = \frac{R_t}{R_t + R_m}$$

Therefore $$I_m = \frac{1.5\text{ k}\Omega}{1.5\text{ k}\Omega + 100\ \Omega} = \frac{1.5\text{ k}}{1.6\text{ k}} = 0.938$$

Therefore $$I_m = 0.938 \times I$$

The current thro the meter is 93.8% of the expected current, therefore the meter current caused a 6.2% error due to effects of loading .

TRANSISTOR VOLTMETER (TVM) 4.7

Direct coupled amplifiers are economical and hence used widely in general purpose low priced VTVM's. Figure 4.9 gives a simplified schematic diagram of a dc coupled amplifier with an indicating meter. The dc input is applied to a range attenuator to provide input voltage levels which can be accommodated by the dc amplifier. The input stage of the amplifier consists of a FET which provides high input impedance to effectively isolate the meter circuit from the circuit under measurement. The input impedance of a FET is greater than 10 MΩ. The bridge is balanced, so that for zero input the dial indicates zero.

The two transistors, Q_1 and Q_2 forms a dc coupled amplifier driving the meter movement. Within the dynamic range of the amplifier, the meter deflection is proportional to the magnitude of the applied input voltage. The input overload does not burn the meter because the amplifier saturates, limiting the maximum current through the meter. The gain of the dc amplifier allows the instrument to be used for measurement of voltages in the mV range. Instruments in the µV range of measurement require a high gain dc amplifier to supply sufficient current for driving the meter movement. In order to avoid the drift problems of dc amplifiers, chopper type dc amplifiers are commonly used in high sensitivity voltmeters.

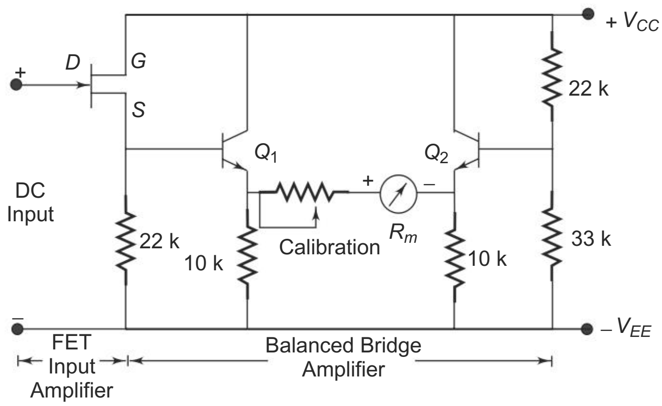


Fig. 4.9 Transistor voltmeter

CHOPPER TYPE DC AMPLIFIER VOLTMETER (MICROVOLTMETER) 4.8

In a chopper type amplifier the dc input voltage is converted into an ac voltage, amplified by an ac amplifier and then converted back into a dc voltage proportional to the original input signal.

The balanced bridge voltmeter has limitations caused by drift problems in dc amplifier. Any fluctuations of voltage supply or variation in the '*Q*' characteristics due to ageing or rise in temperature causes a change in the zero setting or balance. This drift in the steady state conditions of a dc amplifier causes the output indications to change as if the signal input had changed. This drift problem limits the minimum voltage that can be measured. To measure small voltages, a chopper type dc amplifier is used.

A chopper amplifier is normally used for the first stage of amplification in very sensitive instruments of a few µV range. In such an amplifier the dc voltage is chopped to a low frequency of 100 – 300 Hz. It is passed through a blocking capacitor, amplified and then passed through another blocking capacitor, in order to remove the dc drift or offset of the amplified signal.

The principle of operation is as given in Fig. 4.10. An ac amplifier which has a very small drift compared to a dc amplifier is used. The chopper may be mechanical or electronic. Photo diodes are used as nonmechanical choppers for modulation (conversion of dc to ac) and demodulation (conversion of ac to dc). Photo conductors have a low resistance, ranging from a few hundreds to a few thousand ohms, when they are illuminated by a neon or incandescent lamp. The photo conductor resistance increases sharply, usually to several Mega ohms when not illuminated.

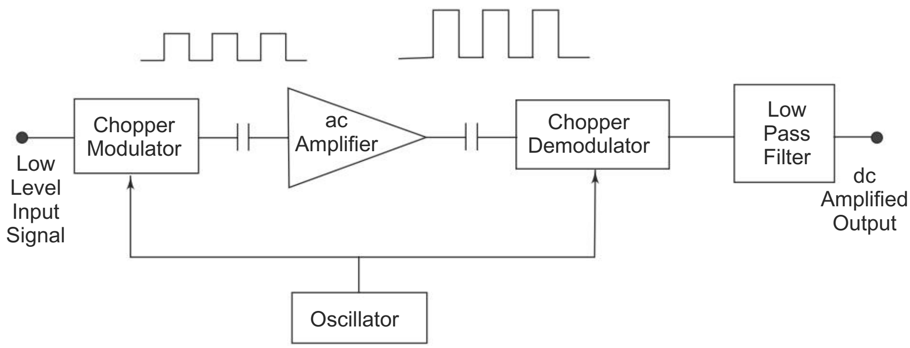


Fig. 4.10 Principle of operation (Chopper type voltmeter)

Figure 4.11 (a) shows a simple circuit for an basic principle of an electronic modulator.

A flashing light source, whose intensity varies from maximum to minimum almost instantaneously, causes the photo diode resistance to change from R_{min} to R_{max} quickly. Therefore the output voltage is an ac, because the photo diode has a high output when its resistance is high and a low output when its resistance is low, as shown in Fig. 4.11 (b).

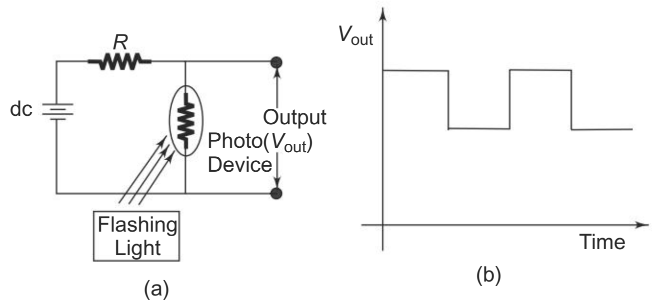


Fig. 4.11 (a) Basic principle of an electronic modulator (b) Output obtained from circuit in Fig. 4.11 (a) (Output voltage waveform)

In the circuit diagram of Fig. 4.12, an oscillator drives two neon lamps into illumination on alternating half cycles of oscillation. The oscillator frequency is limited to a few 100 cycles, because the transition time required for the photo diode to change from high resistance to low resistance limits the chopping range.

Each neon lamp illuminates one photo diode in the input circuit of the amplifier and one in the output circuit. The two photo diodes form a series shunt half wave modulator or chopper. When one photo diode or the input has maximum resistance, the other has minimum resistance. The same conditions exist at the output circuit. Together they act like a switch across the input to the amplifier, alternatively opening and closing at a rate determined by the frequency of the neon oscillator.

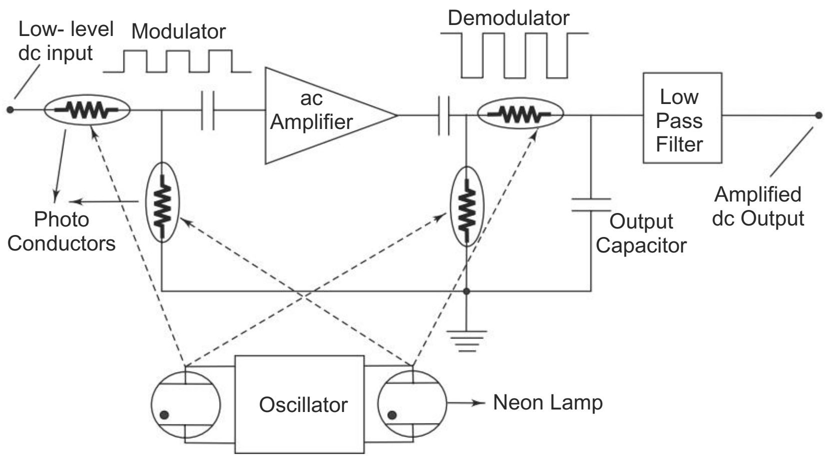


Fig. 4.12 Chopper type voltmeter

The input signal to the amplifier is a square wave whose amplitude is proportional to the input voltage with a frequency equal to the oscillators frequency. The ac amplifier delivers an amplified square wave at its output terminals. The photo diodes (demodulator) in the output circuit operate in antisynchronously with the input chopper, recovering the dc signal by a demodulating action. The dc output signal is then passed to a low pass filter to remove any residual ac component. This amplified drift free dc output is then applied to a PMMC movement for measurement.

The chopper eliminates the need for a high gain dc amplifier with its inherent drift and stability problems.

The input impedance of the chopper amplifier dc voltmeter is usually of the order of 10 MΩ or higher, except in low input ranges. In order to eliminate errors caused by high source impedance, an arrangement for nulling is included in the meter circuit. A control facility provided on the front panel of the instrument permits the input voltage to be nullified with a bucking voltage. When the bucking voltage is equal to the input voltage, a null is indicated, the meter exhibits infinite impedance, and therefore loading effects are totally eliminated. The input voltage is then removed and a bucking voltage equal to the input voltage is indicated by the meter.

A commercially available instrument using a photo chopper amplifier has an input impedance of 100 MΩ, a resolution of 0.1 μV and input ranges from 3 μV full scale to 1 kV full scale with an accuracy of ± 2% of full scale deflection.

Advantages of Chopper Voltmeters

(i) The input impedance of a Chopper Amplifier is usually of the order of 10 MΩ or higher, except on very low input ranges.

(ii) The drift in an ordinary dc amplifier is of the order of mV. The full scale range of an ordinary dc amplifier is limited to measuring input signal of 1 – 100 mV. In a chopper modulator system with the use of ac amplifier, drift can be cut down by a factor of 100, thus allowing an input signal range of about 0.01 mV = 10 μV full scale to be handled.

SOLID STATE VOLTMETER 4.9

Figure 4.13 shows the circuit of an electronic voltmeter using an IC OpAmp 741C. This is a directly coupled very high gain amplifier. The gain of the OpAmp can be adjusted to any suitable lower value by providing appropriate resistance between its output terminal, Pin No. 6, and inverting input, Pin No. 2, to provide a negative feedback. The ratio R_2/R_1 determines the gain, i.e. 101 in this case, provided by the OpAmp. The 0.1 μF capacitor across the 100 k resistance R_2 is for stability under stray pick-ups. Terminals 1 and 5 are called offset null terminals. A 10 kΩ potentiometer is connected between these two offset null terminals with its centre tap connected to a – 5V supply. This potentiometer is called zero set and is used for adjusting zero output for zero input conditions.

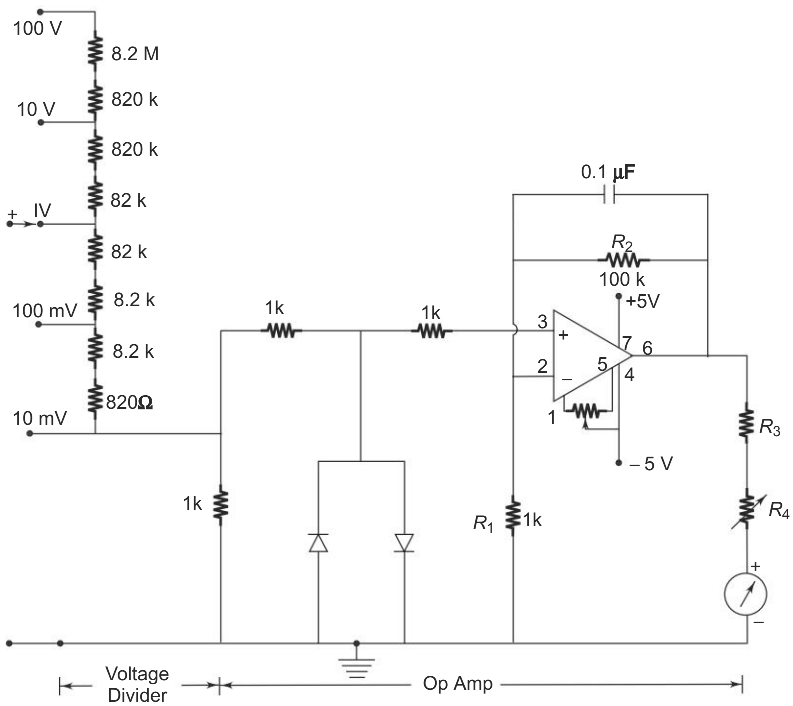


Fig. 4.13 Solid state mV voltmeter using OpAmp

The two diodes used are for IC protection. Under normal conditions, they are non-conducting, as the maximum voltage across them is 10 mV. If an excessive voltage, say more than 100 mV appears across them, then depending upon the polarity of the voltage, one of the diodes conducts and protects the IC. A μA scale of 50 – 1000 μA full scale deflection can be used as an indicator. R_4 is adjusted to get maximum full scale deflection.

DIFFERENTIAL VOLTMETER 4.10

Basic Differential Measurement The differential voltmeter technique, is one of the most common and accurate methods of measuring unknown voltages. In this technique, the voltmeter is used to indicate the difference between known and unknown voltages, i.e., an unknown voltage is compared to a known voltage.

Figure 4.14 (a) shows a basic circuit of a differential voltmeter based on the potentiometric method; hence it is sometimes also called a potentiometric voltmeter.

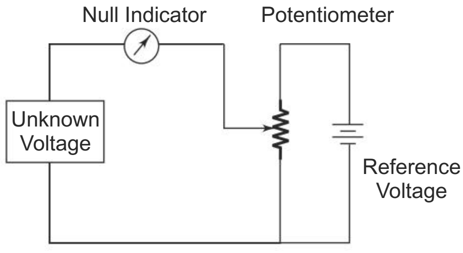


Fig. 4.14 (a) Basic differential voltmeter

In this method, the potentiometer is varied until the voltage across it equals the unknown voltage, which is indicated by the null indicator reading zero. Under null conditions, the meter draws current from neither the reference source nor the unknown voltage source, and hence the differential voltmeter presents an infinite impedance to the unknown source. (The null meter serves as an indicator only.)

To detect small differences the meter movement must be sensitive, but it need not be calibrated, since only zero has to be indicated.

The reference source used is usually a 1 V dc standard source or a zener controlled precision supply. A high voltage reference supply is used for measuring high voltages.

The usual practice, however, is to employ voltage dividers or attenuators across an unknown source to reduce the voltage. The input voltage divider has a relatively low input impedance, especially for unknown voltages much higher than the reference standard. The attenuation will have a loading effect and the input resistance of voltmeter is not infinity when an attenuator is used.

In order to measure ac voltages, the ac voltage must be converted into dc by incorporating a precision rectifier circuit. A block diagram of an ac differential voltmeter is shown in Fig. 4.14 (b).

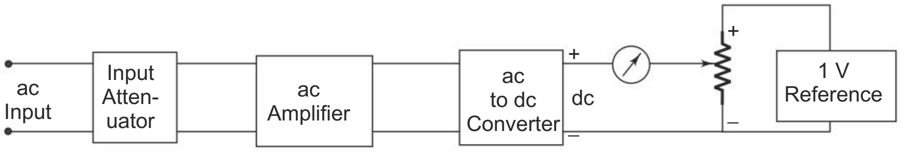


Fig. 4.14 (b) Block diagram of an ac differential voltmeter

DC STANDARD/DIFFERENCE VOLTMETER 4.11

Multifunction Laboratory Instrument

A basic dc standard differential voltmeter can be operated in different modes. The three basic modes of operation are (i) as a dc voltage standard, (ii) as a dc differential voltmeter, and (iii) as a dc voltmeter (conventional).

1. DC Voltage Standard *A + 1 V dc stable supply is obtained from a temperature controlled reference supply, which is applied to a decimal divider network. With the help of switches on the front panel of the voltmeter, the divider ratios can be controlled (varied), allowing the reference supply to be adjusted in steps of 1 μV each, from 0 to 1 V. A low level dc amplifier is used to amplify the low level voltages obtained from the decimal divider to a sufficient level. This reference output voltage is then applied to a high gain dc amplifier with positive feedback to obtain precisely controlled gain characteristics.*

Figure 4.15 (a) illustrates the standard mode of operation. In the standard mode of operation, the differential voltmeter is used to provide a standard reference source in the laboratory, where the instrument generates a precision output voltage from 0 – 1000 V as a reference source.

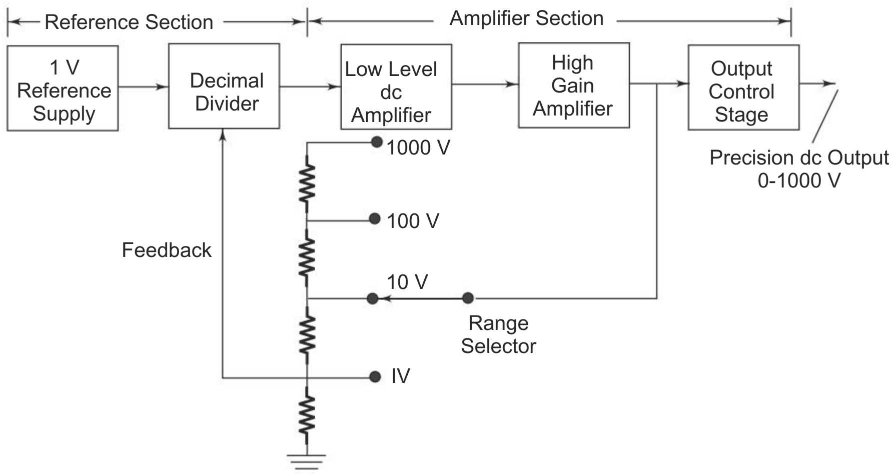


Fig. 4.15 (a) Block diagram of dc standard/differential voltmeter

The dc amplifier consists of several stages in cascade, providing an open loop gain (A) of 10 or higher. (The feedback network monitors the actual output voltage and feeds a controlled fraction of the output back to the amplifier input.)

The closed loop gain of the feedback amplifier is given by

$$G = \frac{A}{1 + A\beta}$$

where G = closed loop gain (voltage gain with feedback)
A = open loop gain (voltage gain without feedback)
β = fraction of the output used as degenerative feedback

If the open loop gain is very high and $A\ \beta$ is much greater than 1, $G = 1/\beta$ implying that the gain of the amplifier depends only on the amount of degenerative feedback.

From the equation ($G = 1/\beta$), it can be seen that as β decreases, the closed loop gain of the amplifier increases. The value of β in turn depends upon the accuracy of the voltage dividers used, i.e. the precision value of the resistors used for the voltage dividers (wire wound precision resistors). The following ranges of standard voltage are available at the output.

0 – 1 V in 1 μV steps (1 V range)
0 – 10 V in 10 μV steps (10 V range)
0 – 100 V in 100 μV steps (100 V range)
0 – 1000 V in 1 mV steps (1000 V range)

2. DC Differential Voltmeter The unknown dc voltage is applied to the input of the amplifier section and a part of the output voltage is fed back to the input stage with the help of one divider network, which controls the closed loop gain of the amplifier. The other section of the voltage divider network applies a fraction of the output voltage to the differential input of the meter amplifier.

The meter circuit measures the difference between the feedback voltage and the reference voltage, indicating a null deflection when the two voltages are equal. The range selector (on the front panel) controls both the feedback voltage and the voltage that is applied in opposition to the reference divider output, such that 1 V capacity of the reference supply is never exceeded.

A differential voltmeter is shown in Fig. 4.15 (b).

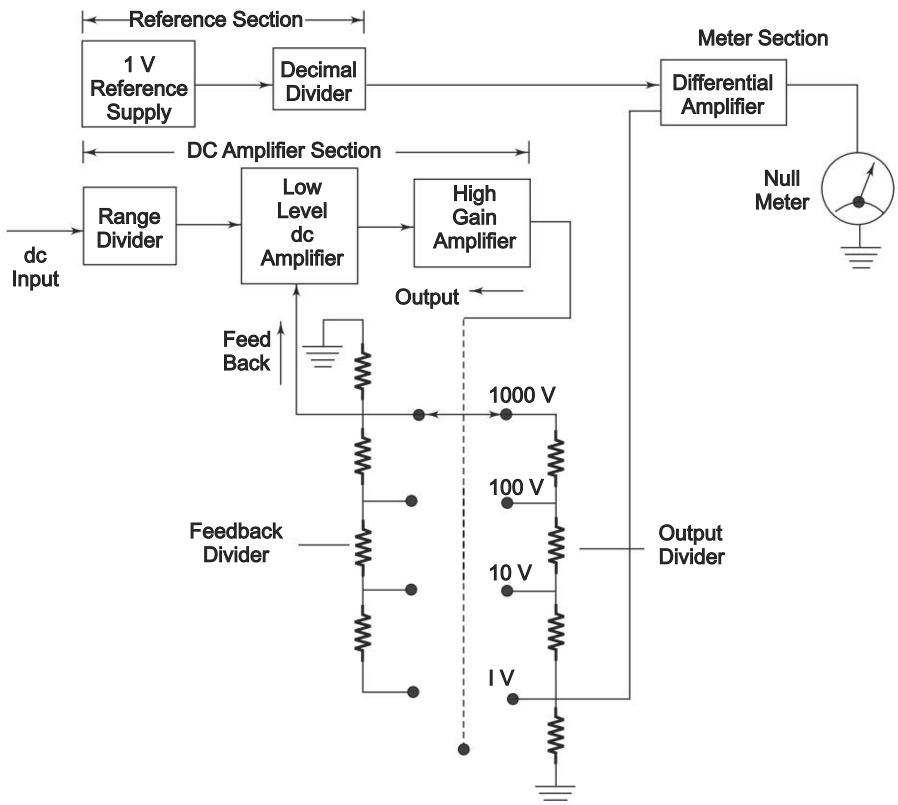


Fig. 4.15 (b) Differential voltmeter

3. DC Voltmeter In this mode of operation the instrument is connected as a dc voltmeter. High input impedance to the unknown voltage source is provided by the dc amplifier, which act as a buffer stage.

The input voltage is amplified and the dc output voltage is applied directly to the meter circuit. The meter circuit involves a feedback controlled amplifier and allows a selection of the sensitivity.

AC VOLTMETER USING RECTIFIERS 4.12

Rectifier type instruments generally use a PMMC movement along with a rectifier arrangement. Silicon diodes are preferred because of their low reverse current and high forward current ratings. Figure 4.16 (a) gives an ac voltmeter circuit consisting of a multiplier, a bridge rectifier and a PMMC movement.

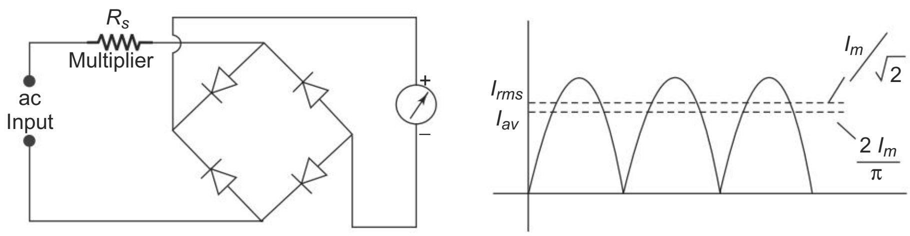


Fig. 4.16 (a) ac voltmeter (b) Average and RMS value of current

The bridge rectifier provides a full wave pulsating dc. Due to the inertia of the movable coil, the meter indicates a steady deflection proportional to the average value of the current (Fig. 4.16 (b)). The meter scale is usually calibrated to give the RMS value of an alternating sine wave input.

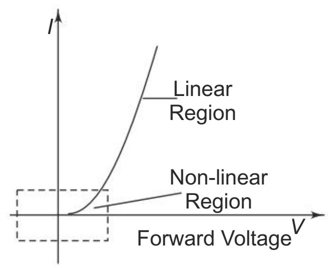


Fig. 4.16 (c) Diode characteristics (Forward)

Practical rectifiers are non-linear devices particularly at low values of forward current (Fig. 4.16 (c)). Hence the meter scale is non-linear and is generally crowded at the lower end of a low range voltmeter. In this part the meter has low sensitivity because of the high forward resistance of the diode. Also, the diode resistance depends on the temperature.

The rectifier exhibits capacitance properties when reverse biased, and tends to bypass higher frequencies. The meter reading may be in error by as much as 0.5% decrease for every 1 kHz rise in frequency.

A general rectifier type ac voltmeter arrangement is given in Fig. 4.17.

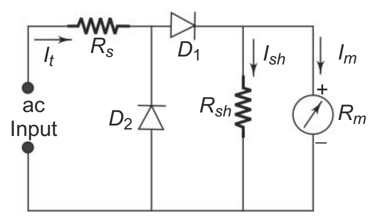


Fig. 4.17 General rectifier type ac voltmeter

Diode D_1 conducts during the positive half of the input cycle and causes the meter to deflect according to the average value of this half cycle. The meter movement is shunted by a resistor, R_{sh}, in order to draw more current through the diode D_1 and move the operating point into the linear portion of the characteristic curve. In the negative half cycle, diode D_2 conducts and the current through the measuring circuit, which is in an opposite direction, bypasses the meter movement.

AC VOLTMETER USING HALF WAVE RECTIFIER 4.13

If a diode D_1 is added to the dc voltmeter, as shown in Fig. 4.18, we have an ac voltmeter using half wave rectifier circuit capable of measuring ac voltages. The sensitivity of the dc voltmeter is given by

$$S_{dc} = 1/I_{fsd} = 1/1 \text{ mA} = 1 \text{ k}\Omega$$

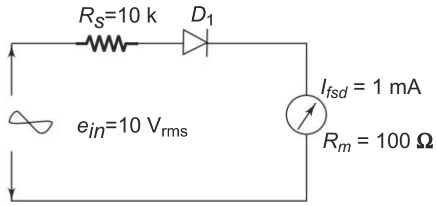


Fig. 4.18 ac voltmeter using half wave rectifier

A multiple of 10 times this value means a 10 V dc input would cause exactly full scale deflection when connected with proper polarity. Assume D_1 to be an ideal diode with negligible forward bias resistance. If this dc input is replaced by a 10 V rms sine wave input. The voltages appearing at the output is due to the +ve half cycle due to rectifying action.

The peak value of 10 V rms sine wave is

$$E_p = 10 \text{ V rms} \times 1.414 = 14.14 \text{ V peak}$$

The dc will respond to the average value of the ac input, therefore

$$E_{av} = E_p \times 0.636 = 14.14 \times 0.636 = 8.99 \text{ V}$$

Since the diode conducts only during the positive half cycle, the average value over the entire cycle is one half the average value of 8.99 V, i.e. about 4.5 V.

Therefore, the pointer will deflect for a full scale if 10 V dc is applied and 4.5 V when a 10 Vrms sinusoidal signal is applied. This means that an ac voltmeter is not as sensitive as a dc voltmeter.

As $$E_{dc} = 0.45 \times E_{rms}$$

∴ The value of the multiplier resistor can be calculated as

$$R_s = \frac{E_{dc}}{I_{dc}} - R_m = \frac{0.45 \times E_{rms}}{I_{dc}} - R_m$$

Example 4.15 *Calculate the value of the multiplier resistor for a 10 V rms range on the voltmeter shown in Fig. 4.19.*

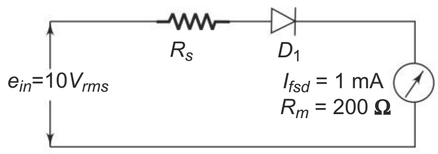


Fig. 4.19

Solution

Method 1 Sensitivity of the meter movement is

$$S_{dc} = 1/I_{fsd} = 1/1 \text{ mA} = 1 \text{ k}\Omega$$
$$R_s = S_{dc} \times \text{range} - R_m = 1 \text{ k}\Omega/\text{V} \times 0.45\ E_{rms} - R_m$$
$$= 1 \text{ k}\Omega/\text{V} \times 0.45 \text{ V} \times 10 \text{ V} - 200\ \Omega$$
$$= 4500 - 200$$
$$= 4.3 \text{ k}\Omega$$

Method 2

$$R_s = \frac{0.45 \times E_{rms}}{1 \text{ mA}} - R_m = \frac{0.45 \times 10}{1 \text{ mA}} - 200$$
$$= 4.5 \text{ k} - 0.2 \text{ k}$$
$$= 4.3 \text{ k}\Omega$$

AC VOLTMETER USING FULL WAVE RECTIFIER 4.14

Consider the circuit shown in Fig. 4.20. The peak value of a 10 V rms signal is

$$E_p = 1.414 \times E_{\text{rms}}$$
$$= 1.414 \times 10 = 14.14 \text{ V peak}$$

Average value is

$$E_{av} = 0.636 \times E_{\text{peak}}$$
$$= 14.14 \times 0.636 = 8.99 \text{ V}$$
$$\cong 9 \text{ V}$$

Therefore, we can see that a 10 V rms voltage is equal to a 9 V dc for full scale deflection, i.e. the pointer will deflect to 90% of full scale, or

Sensitivity (ac) = 0.9 × Sensitivity (dc)

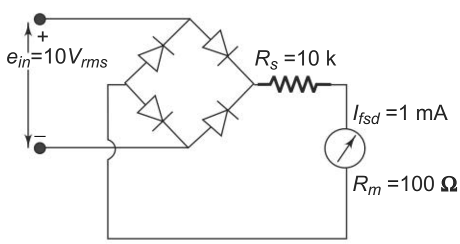


Fig. 4.20 ac voltmeter using full wave rectifier

Example 4.16 *Calculate the value of the multiplier resistor required for a 100 Vrms range on the voltmeter shown in Fig 4.21.*

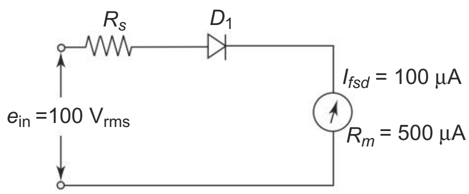


Fig. 4.21

Solution

Method 1 Sensitivity of the voltmeter is given by

$$S_{dc} = 1/I_{fsd} = 1/100\ \mu A = 10^6/100 = 10\ k\Omega/V$$

$$R_s = S_{dc} \times \text{range} - R_m = 10\ k\Omega/V \times 0.45\ V \times 100 - 500\ \Omega$$
$$= 450\ k\Omega - 500\ \Omega = 449.5\ k\Omega$$

Method 2

$$R_s = \frac{0.45 \times E_{rms}}{I_{dc}} - R_m = \frac{0.45 \times 100}{100\ \mu A} - 500\ \Omega$$
$$= 0.45 \times 10^{6} - 500\ \Omega$$
$$= 450\ k\Omega - 500\ \Omega = 449.5\Omega$$

Example 4.17 *Calculate the value of the multiplier resistor for a 50 Vrms ac range on the voltmeter as shown in Fig 4.22.*

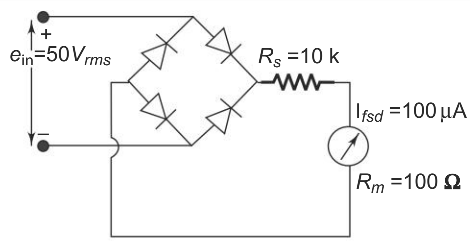


Fig. 4.22

Solution Given I_{fsd} = 100 μA, R_m = 100 Ω

Step 1: The dc sensitivity is given by

$$S_{dc} = 1/I_{fsd} = 1/100\ \mu A = 10^6/100 = 10\ k\Omega/V$$

Step 2: Ac sensitivity = 0.9 × dc sensitivity

$$S_{ac} = 0.9 \times S_{dc} = 0.9 \times 10\ k\Omega/V = 9\ k\Omega/V$$

Step 3: The multiplier resistor is given by

$$R_s = S_{ac} \times \text{range} - R_m$$
$$= 9\ k\Omega/V \times 50\ \text{Vrms} - 100\Omega$$
$$= 450\ K - 100\ \Omega = 449.9\ k\Omega$$

Example 4.18 *Calculate the value of the multiplier resistor for a 10 V rms ac range on the voltmeter in Fig. 4.23.*

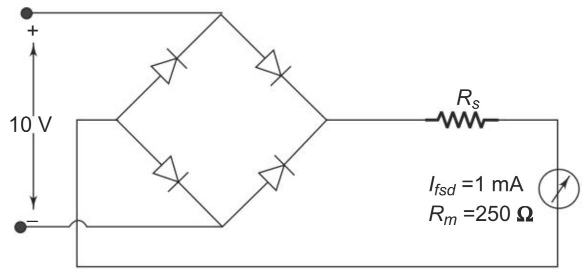


Fig. 4.23

Solution The dc sensitivity is given by

$$S_{dc} = 1/I_{fsd} = 1/1 \text{ mA} = 1 \text{ k}\Omega/\text{V}$$

Therefore

$$\text{AC sensitivity} = 0.9 \times \text{dc sensitivity}$$

$$S_{ac} = 0.9 \times 1 \text{ k}\Omega/\text{V} = 0.9 \text{ k}\Omega/\text{V}$$

The multiplier resistor is given by

$$\begin{aligned} R_s &= S_{\text{ac}} \times \text{range} - R_m = 0.9 \text{ k}\Omega/\text{V} \times 10 \text{ V} - 250 \\ &= 900 \times 10 - 250 \\ &= 9000 - 250 \\ &= 8750 \\ &= 8.75 \text{ k}\Omega \end{aligned}$$

Example 4.19 *Determine the reading obtained with a dc voltmeter in the circuit Fig. Ex4.19, when the switch is set to position A, then set the switch to position B and determine the reading obtained with a half-wave rectifier and fullwave rectifier ac voltmeter.*

All meters use a 100 µA full scale deflection meter movement and are set on 10 V dc or rms ranges.

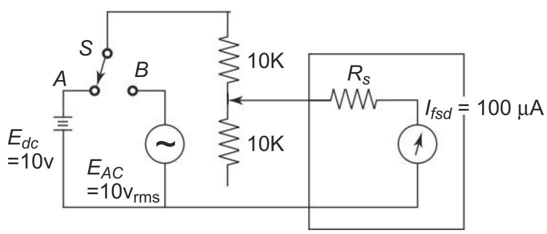


Fig. Ex4.19

Solution

Step 1: The sensitivity with a dc voltmeter is calculated as follows:

$$S_{dc} = 1/I_{fsd} = 1/100 \text{ µA} = 10 \text{ k}\Omega/\text{V}$$

Step 2: The multiplier resistor R_s can be calculated as follows:

$$R_s = S_{dc} \times \text{range} = 10 \text{ k}\Omega/\text{V} \times 10\text{V} = 100 \text{ k}\Omega$$

Step 3: The voltage across resistor R_2 read by the voltmeter can be obtained as follows:

$$ER_2 = \frac{R_2 // R_s}{R_1 + R_2 // R_s} \times e = \frac{10\text{ k} // 100\text{ k}}{10\text{ k} + 10\text{ k} // 100\text{ k}} \times 10\text{ V}$$

$$= \frac{9.09\text{ k}}{10\text{ k} + 9.09\text{ k}} \times 10\text{ V} = \frac{9.09\text{ k}}{19.09\text{ k}} \times 10\text{ V} = 4.76\text{ V}$$

Step 4: The reading obtained with ac voltmeter using a half-wave rectifier is calculated as follows:

$$S_{hw} = 0.45 \times S_{dc} = 0.45 \times 10\text{ k}\Omega/\text{V} = 4.5\text{ k}\Omega/\text{V}$$

$$R_s = S_{dc} \times \text{range} = 4.5\text{ k}\Omega/\text{V} \times 10\text{V} = 45\text{ k}\Omega$$

Step 5: The voltage read by the ac voltmeter is calculated as follows:

$$E = \frac{R_2 // R_s}{R_1 + R_2 // R_s} \times E = \frac{45\text{ k} // 10\text{ k}}{10\text{ k} + 10\text{ k} // 45\text{ k}} \times 10\text{ V}$$

$$= \frac{8.18\text{ k}}{10\text{ k} + 8.18\text{ k}} \times 10\text{ V} = \frac{8.18\text{ k}}{18.18\text{ k}} \times 10\text{ V} = 4.499\text{ V}$$

Step 6: And finally the reading obtained with ac voltmeter using full-wave rectifier is calculated as follows:

$$S_{fw} = \frac{0.90 \times S_{dc}}{V} = 0.90 \times 10\text{ k}\Omega/\text{V} = 9.0\text{ k}\Omega/\text{V}$$

$$R_s = S_{fw} \times \text{range} = 9.0\text{ k}\Omega/\text{V} \times 10\text{V} = 90\text{ k}\Omega$$

Step 7: The voltage read by the ac voltmeter is calculated as follows:

$$E = \frac{R_2 // R_s}{R_1 + R_2 // R_s} \times E = \frac{90\text{ K} // 10\text{ k}}{10\text{ k} + 10\text{ k} // 90\text{ k}} \times 10\text{ V}$$

$$= \frac{9\text{ k}}{10\text{ k} + 9\text{ k}} \times 10\text{ V} = \frac{9\text{ k}}{19\text{ k}} \times 10\text{ V} = 4.73\text{ V}$$

As can be seen an ac voltmeter using half wave or full wave rectifier has more loading effect than dc voltmeter.

MULTIRANGE AC VOLTMETER 4.15

Figure 4.24 is circuit for measuring ac voltages for different ranges. Resistances R_1, R_2, R_3 and R_4 form a chain of multipliers for voltage ranges of 1000 V, 250 V, 50 V, and 10 V respectively.

On the 2.5 V range, resistance R_5 acts as a multiplier and corresponds to the multiplier R_s shown in Fig. 4.17.

R_{sh} is the meter shunt and acts to improve the rectifier operation.

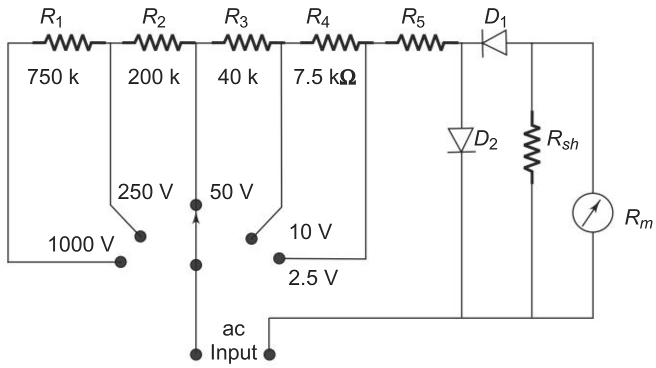


Fig. 4.24 Multirange ac voltmeter

AVERAGE RESPONDING VOLTMETER 4.16

A simplified version of a circuit used in a typical average responding voltmeters is given in Fig. 4.25.

The applied waveform is amplified in a high gain stabilised amplifier to a reasonably high level and then rectified and fed to a dc mA meter calibrated in terms of rms input voltage. In this meter instrument, the rectified current is averaged by a filter to produce a steady deflection of the meter pointer. A dc component in the applied voltage is excluded from the measurement by an input blocking capacitor preceding the high gain amplifier.

The ac amplifier has a large amount of negative feedback, which ensures gain stability for measurement accuracy, and an increased frequency range of the instrument. The inclusion of the meter in the feedback path minimises the effect of diode non-linearity and meter impedance variations on the circuit performance.

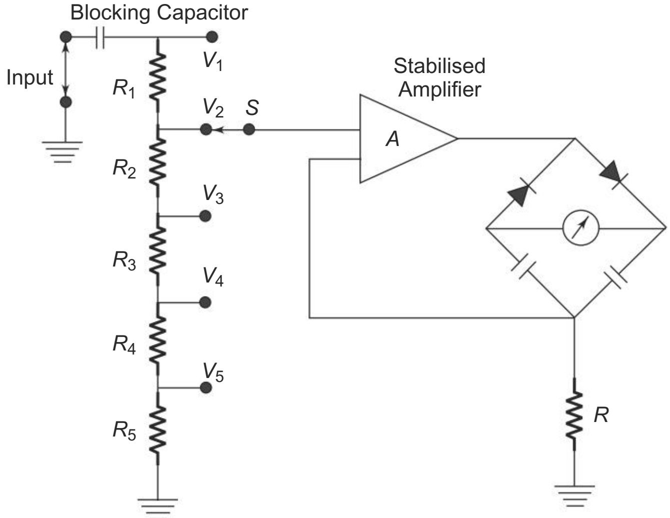


Fig. 4.25 Block diagram of average responding voltmeter

Capacitors in the meter circuit tend to act as storage or filter capacitors for the rectifier diodes as well as coupling capacitors for the feedback signal. The diodes acts as switches to maintain unidirectional meter current despite changes in the instantaneous polarity of the input voltage.

Errors in the reading of an average responding voltmeter may be due to the application of complex waveforms, i.e. a distorted or nonsinusoidal input or the presence of hum or noise.

The accuracy with which an average responding voltmeter indicates the rms value of a wave with harmonic content depends not only on the amplitude of the harmonic but also on the phase.

PEAK RESPONDING VOLTMETER 4.17

The basic difference between peak responding voltmeters and average responding voltmeters is the use of storage capacitors with the rectifying diode in the former case. The capacitor charges through the diode to the peak value of the applied voltage and the meter circuit then responds to the capacitor voltage.

The two most common types of peak responding voltmeters are given in Figs 4.26 (a) and (b).

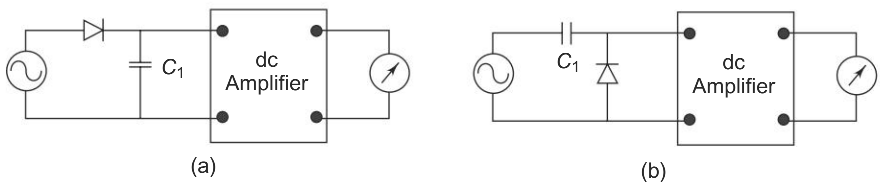


Fig. 4.26 Peak responding voltmeter

Figure 4.26 (a) shows a dc coupled peak voltmeter, in which the capacitor charges to the total peak voltage above ground reference. In this case the meter reading will be affected by the presence of dc with ac voltage.

In Fig. 4.26 (b), an ac coupled peak voltmeter circuit is shown. In both the circuits, the capacitor discharges very slowly through the high impedance input of the dc amplifier, so that a negligible small amount of current supplied by the circuit under test keeps the capacitor charged to the peak ac voltage. The dc amplifier is used in the peak responding meter to develop the necessary meter current.

The primary advantage of a peak responding voltmeter is that the rectifying diode and the storage capacitor may be taken out of the instrument and placed in the probe when no ac pre-amplification is required. The measured ac signal then travels no farther than the diode. The peak responding voltmeter is then able to measure frequencies of up to 100s of MHz with a minimum of circuit loading. The disadvantage of peak responding voltmeters is the error caused due to harmonic distortion in the input waveforms and limited sensitivity of the instrument because of imperfect diode characteristics.

TRUE RMS VOLTMETER 4.18

Complex waveform are most accurately measured with an rms voltmeter. This instrument produces a meter indication by sensing waveform heating power, which is proportional to the square of the rms value of the voltage. This heating power can be measured by amplifying and feeding it to a thermocouple, whose output voltages is then proportional to the E_{rms}.

However, thermocouples are non-linear devices. This difficulty can be overcome in some instruments by placing two thermocouples in the same thermal environment.

Figure 4.27 shows a block diagram of a true rms responding voltmeter.

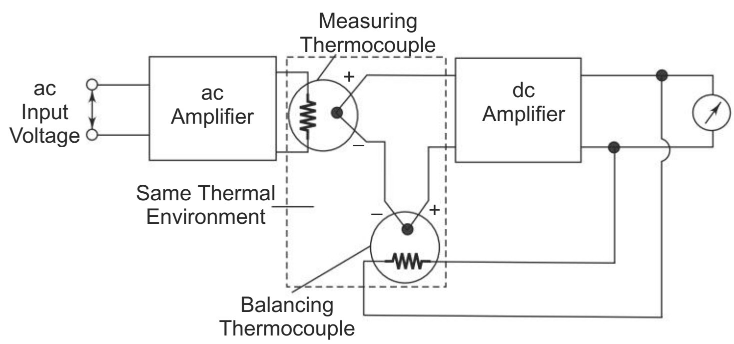


Fig. 4.27 True RMS voltmeter (Block diagram)

The effect of non-linear behaviour of the thermocouple in the input circuit (measuring thermocouple) is cancelled by similar non-linear effects of the thermocouple in the feedback circuit (balancing thermocouple). The two couples form part of a bridge in the input circuit of a dc amplifier.

The unknown ac voltage is amplified and applied to the heating element of the measuring thermocouple. The application of heat produces an output voltage that upsets the balance of the bridge.

The dc amplifier amplifies the unbalanced voltage; this voltage is fed back to the heating element of the balancing thermocouple, which heats the thermocouple, so that the bridge is balanced again, i.e. the outputs of both the thermocouples are the same. At this instant, the ac current in the input thermocouple is equal to the dc current in the heating element of the feedback thermocouple. This dc current is therefore directly proportional to the effective or rms value of the input voltage, and is indicated by the meter in the output circuit of the dc amplifier. If the peak amplitude of the ac signal does not exceed the dynamic range of the ac amplifier, the true rms value of the ac signal can be measured independently.

TRUE RMS METER 4.19

There exists a fundamental difference between the readings on a normal ac meter and on a true rms meter. The first uses a D'Arsonval movement with a full or half wave rectifier, and averages the values of the instantaneous rectified current.

The rms meter, however, averages the squares of the instantaneous current values (proportional, for example, to the instantaneous heating effect). The scale of the true rms meter is calibrated in terms of the square roots of the indicated current values. The resulting reading is therefore the square root of the average of the squared instantaneous input values, which is the rms value of the measured alternating current.

A true rms meter is always a combination of a normal mean value indicating meter and a squaring device whose output at any instant is proportional to the instantaneous squared input.

It can be shown that the ac component of the voltage developed across the common collector resistors of two transistors that are connected in parallel, and between the bases of which a small ac voltage is applied, is proportional to the square of the applied input voltage.

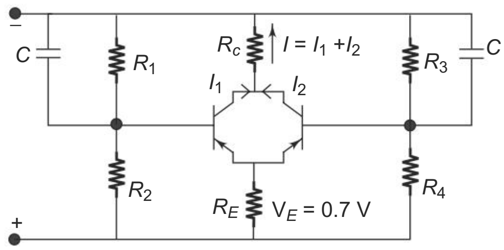


Fig. 4.28 Squaring device

The basic circuit of Fig. 4.28 employing two transistors is completed by a bridge arrangement in which the dc component is cancelled out. This bridge arrangement is given in Fig. 4.29.

One side of the bridge consists of two parallel connected transistors Q_2 and Q_3, and a common collector resistor R_{13}. The side of the bridge, employing P_1 for bias setting, is the basic squaring circuit. The other side of the bridge is made of transistor Q_4 (whose base is biased by means of potentiometer P_2 and collector resistance R_{16}.)

Potentiometer P_1,base bias balance of the squaring circuit, must be adjusted for symmetrical operation of transistors Q_2 and Q_3. To do this, the polarity of a small dc input voltage applied to terminals A and B (bases of Q_2 and Q_3) has to be reversed, and the reading of the output meter must be the same for both input polarities.

Potentiometer P_2 must be set so that for zero input signal (terminals A and B short-circuited), the bridge is balanced and the meter reads zero. The balance condition is reached if the voltage drop across the collector resistance R_{13} of $Q_2 - Q_3$, and collector resistance R_{16} of Q_4, are equal.

Transistor Q_1 is used to improve the temperature stability of the whole circuit, which is basically obtained by the emitter resistance R_{10}. Optimum temperature compensation is obtained if the voltage drop across the emitter resistance for no signal is 0.7 V for silicon transistor.

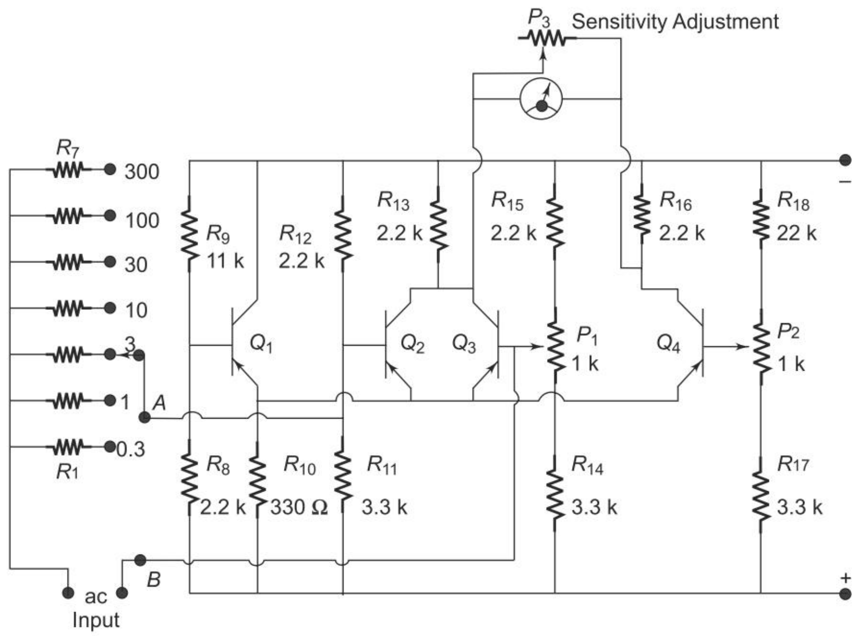


Fig. 4.29 True RMS meter

The low current through Q_2, Q_3, Q_4 requires a large emitter resistance value to fulfil the condition for compensation. Therefore, another transistor, Q_1 has been added to compensate for the temperature changes of Q_2 and Q_3.

The bias on this transistor has to be adjusted by selecting appropriate values of R_8 and R_9 so that the voltage drop across R_{10} in the balanced condition is 0.7 V for silicon transistor.

The input of the squaring devices (*AB*) is connected to a voltage divider that is calibrated in seven ranges, namely 0.3, 1, 3, 10, 30, 100, and 300 volts.

CONSIDERATIONS IN CHOOSING AN ANALOG VOLTMETER 4.20

In choosing an analog voltmeter the following factors are to be considered.

1. Input Impedance The input impedance or resistance of the voltmeter should be as high as possible. It should always be higher than the impedance of the circuit under measurement to avoid the loading effect, discussed in Section 4.6.

The shunt capacitance across the input terminals also determines the input impedance of the voltmeter. At higher frequencies the loading effect of the meter is noticeable, since the shunt capacitance reactance falls and the input shunt reduces the input impedance.

2. Voltage Ranges The voltage ranges on the meter scale may be in a 1–3–10 sequence with 10 db separation or a 1.5–5–15 sequence or in a single scale calibrated in decibels. In any case, the scale division should be compatible with the accuracy of the instrument.

3. Decibels For measurements covering a wide range of voltages, the use of the decibel scale can be very effective, e.g., in the frequency response curve of an amplifier, where the output voltage is measured as a function of the frequency of the applied input voltage.

4. Sensitivity v/s Bandwidth Noise consists of unwanted frequencies. Since noise is a function of the bandwidth, a voltmeter with a narrow bandwidth picks up less noise than a large bandwidth voltmeter.

In general, an instrument with a bandwidth of 10 Hz–10 MHz has a sensitivity of 1 mV. Some voltmeters whose bandwidth extends up to 5 MHz may have a sensitivity of 100 μV.

5. Battery Operation A voltmeter (VTVM) powered by an internal battery is essential for field work.

6. AC Current Measurements Current measurements can be made by a sensitive ac voltmeter and a series resistor.

To summarise, the general guidelines are as follows.

(i) For dc measurement, select the meter with the widest capability meeting the requirements of the circuit.

(ii) For ac measurements involving sine waves with less than 10% distortion, the average responding voltmeter is most sensitive and provides the best accuracy.

(iii) For high frequency measurement (> 10 MHz), the peak responding voltmeter with a diode probe input is best. Peak responding circuits are acceptable if inaccuracies caused by distortion in the input waveform are allowed (tolerated).

(iv) For measurements where it is important to find the effective power of waveforms that depart from the true sinusoidal form, the rms responding voltmeter is the appropriate choice.

OHMMETER (SERIES TYPE OHMMETER) 4.21

A D'Arsonval movement is connected in series with a resistance R_1 and a battery which is connected to a pair of terminals A and B, across which the unknown resistance is connected. This forms the basic type of series ohmmeter, as shown in Fig. 4.30 (a).

The current flowing through the movement then depends on the magnitude of the unknown resistance. Therefore, the meter deflection is directly proportional to the value of the unknown resistance.

Referring to Fig. 4.30 (a)

R_1 = current limiting resistance
R_2 = zero adjust resistance
V = battery
R_m = meter resistance
R_x = unknown resistance

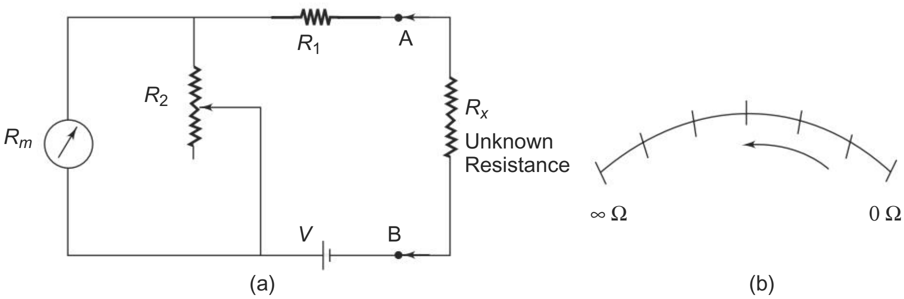


Fig. 4.30 (a) Series type ohmmeter (b) Dial of series ohmmeter

4.21.1 Calibration of the Series Type Ohmmeter

To mark the "0" reading on the scale, the terminals A and B are shorted, i.e. the unknown resistance $R_x = 0$, maximum current flows in the circuit and the shunt resistance R_2 is adjusted until the movement indicates full scale current (I_{fsd}). The position of the pointer on the scale is then marked "0" ohms.

Similarly, to mark the "∞" reading on the scale, terminals A and B are open, i.e. the unknown resistance $R_x = \infty$, no current flow in the circuit and there is no deflection of the pointer. The position of the pointer on the scale, is then marked as "∞" ohms.

By connecting different known values of the unknown resistance to terminals A and B, intermediate markings can be done on the scale. The accuracy of the instrument can be checked by measuring different values of standard resistance, i.e. the tolerance of the calibrated resistance, and noting the readings.

A major drawback in the series ohmmeter is the decrease in voltage of the internal battery with time and age. Due to this, the full scale deflection current drops and the meter does not read "0" when A and B are shorted. The variable shunt resistor R_2 across the movement is adjusted to counteract the drop in battery voltage, thereby bringing the pointer back to "0" ohms on the scale.

It is also possible to adjust the full scale deflection current without the shunt R_2 in the circuit, by varying the value of R_1 to compensate for the voltage drop. Since this affects the calibration of the scale, varying by R_2 is much better solution. The internal resistance of the coil R_m is very low compared to R_1. When R_2 is varied, the current through the movement is increased and the current through R_2 is reduced, thereby bringing the pointer to the full scale deflection position.

The series ohmmeter is a simple and popular design, and is used extensively for general service work.

Therefore, in a series ohmmeter the scale marking on the dial, has "0" on the right side, corresponding to full scale deflection current, and "∞" on the left side corresponding to no current flow, as given in Fig. 4.30 (b).

Values of R_1 and R_2 can be determined from the value of R_x which gives half the full scale deflection.

$$R_h = R_1 + R_2 \parallel R_m = R_1 + \frac{R_2 R_m}{R_2 + R_m}$$

where R_h = half of full scale deflection resistance.

The total resistance presented to the battery then equals $2R_h$ and the battery current needed to supply half scale deflection is $I_h = V/2\ R_h$.

To produce full scale current, the battery current must be doubled.

Therefore, the total current of the ckt, $I_t = V/R_h$

The shunt current through R_2 is given by $I_2 = I_t - I_{fsd}$

The voltage across shunt, V_{sh}, is equal to the voltage across the meter.

Therefore $$V_{sh} = V_m$$

$$I_2 R_2 = I_{fsd} R_m$$

Therefore $$R_2 = \frac{I_{fsd} R_m}{I_2}$$

But $$I_2 = I_t - I_{fsd}$$

$\therefore$ $$R_2 = \frac{I_{fsd} R_m}{I_t - I_{fsd}}$$

But $$I_t = \frac{V}{R_h}$$

Therefore $$R_2 = \frac{I_{fsd} R_m}{V/R_h - I_{fsd}}$$

Therefore $$R_2 = \frac{I_{fsd} R_m R_h}{V - I_{fsd} R_h} \tag{4.1}$$

As $$R_h = R_1 + \frac{R_2 R_m}{R_2 + R_m}$$

Therefore $$R_1 = R_h - \frac{R_2 R_m}{R_2 + R_m}$$

Hence $$R_1 = R_h - \frac{\dfrac{I_{fsd} R_m R_h}{V - I_{fsd} R_h} \times R_m}{\dfrac{I_{fsd} R_m R_h}{V - I_{fsd} R_h} + R_m}$$

Therefore $$R_1 = R_h - \frac{I_{fsd} R_m R_h}{V} \tag{4.2}$$

Hence, R_1 and R_2 can be determined.

Example 4.20 *A 100 Ω basic movement is to be used as an ohmmeter requiring full scale deflection of 1 mA and internal battery voltage of 3 V. A half scale deflection marking of 1 kΩ is required. Calculate*

(i) the value of R_1 and R_2 (ii) Maximum value of R_2 to compensate for a 3% drop in battery voltage

Solution Given $I_m = 1$ mA, $R_m = 100\ \Omega$, $R_h = 1\ \text{k}\Omega$, $V = 3$V

(i) To find R_1 and R_2

Step 1:

$$R_1 = R_h - \frac{I_{fsd} \times R_m \times R_h}{V} = 1\ \text{k}\Omega - \frac{1\ \text{mA} \times 100\ \times 1\ \text{k}\Omega}{3}$$

$$= 1\ \text{k}\Omega - \frac{100}{3} = 1000\ \Omega - 33.3\ \Omega$$

$$= 966.7\ \Omega$$

Step 2:

$$R_2 = \frac{I_{fsd} \times R_m \times R_h}{V - I_{fsd} \times R_h} = -\frac{1\ \text{mA} \times 100\ \times 1\ \text{k}\Omega}{3 - 1\ \text{mA} \times 1\ \text{k}\Omega} = \frac{100}{3-1} = \frac{100}{2} = 50\ \Omega$$

Step 3:

(ii) The internal battery voltage is 3 V. 3% of 3 V is .09 V.

Therefore, the battery voltage drop with 3% drop is 3 V – .09 V = 2.91V

$$R_2 = \frac{I_{fsd} \times R_m \times R_h}{V - I_{fsd} \times R_h} = -\frac{1\ \text{mA} \times 100\ \times 1\ \text{k}\Omega}{2.91 - 1\ \text{mA} \times 1\ \text{k}\Omega} = \frac{100}{2.91 - 1} = \frac{100}{1.91} = 52.36\ \Omega$$

$R_2 = 52.36\ \Omega$

Example 4.21 *A 100 Ω basic movement is to be used as an ohmmeter requiring a full scale deflection of 1 mA and internal battery voltage of 3 V. A half scale deflection marking of 2 k is desired. Calculate (i) value of R_1 and R_2, and (ii) The maximum value of R_2 to compensate for a 5% drop in battery voltage.*

Solution (i) Using the equations for R_1 and R_2 we have,

$$R_1 = R_h - \frac{I_{fsd} \times R_m \times R_h}{V} \quad \text{and} \quad R_2 = \frac{I_{fsd} \times R_m \times R_h}{V - (I_{fsd} \times R_h)}$$

Hence $$R_1 = 2\ \text{k} - \frac{1\ \text{mA} \times 100 \times 2\ \text{k}}{3} = 2\ \text{k} - \frac{200}{3} = 2\ \text{k} - 66.6$$

Therefore $$R_1 = 2000 - 66.6 = 1933.3\ \Omega$$

$$R_2 = \frac{1\ \text{mA} \times 100 \times 2\ \text{k}}{3 - 1\ \text{mA} \times 2\ \text{k}} = \frac{200}{1} = 200\ \Omega$$

(ii) The internal battery voltage is 3 V, therefore 5% of 3 V is 0.15 V. The battery voltage with 5% drop is 3 V – 0.15 V = 2.85 V.

$$R_2 = \frac{I_{fsd} \times R_h \times R_m}{V - I_{fsd} \times R_h} = \frac{1\text{ mA} \times 100 \times 2\text{k}}{2.85\text{ V} - 1\text{ mA} \times 2\text{k}} = \frac{200}{0.85} = 235.29\ \Omega$$

Example 4.22 *A 1 mA full scale deflection (fsd) current meter movement is to be used as an ohmmeter circuit. The meter movement has an internal resistance of 100 Ω and a 3 V battery will be used in the circuit. Mark off the meter face (dial) for reading resistance.*

Solution

Step 1: The value of R_s which will limit current to full scale deflection current can be calculated as

$$R_s = \frac{E}{I_m} - R_m = \frac{3}{1\text{ mA}} - 100\ \Omega = 3\text{ k}\Omega - 100\ \Omega = 2.9\text{ k}\Omega$$

Step 2: The value of R_x with a 20% deflection is

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) \qquad \text{where } P = \frac{I}{I_m} = 20\% = \frac{20}{100} = 0.2$$

Therefore,
$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m)$$

$$= \frac{(2.9\text{ k}\Omega + 0.1\text{ k}\Omega)}{0.2} - (2.9\text{ k}\Omega + 0.1\text{ k}\Omega)$$

$$R_x = \frac{3\text{ k}\Omega}{0.2} - 3\text{ k}\Omega = 15\text{ k}\Omega - 3\text{ k}\Omega = 12\text{ k}\Omega$$

Step 3: The value of R_x with a 40% deflection is

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) \qquad \text{where } P = \frac{I}{I_m} = 40\% = \frac{40}{100} = 0.4$$

Therefore,

$$R_x = \frac{(R_s + R_m)}{4} - (R_s + R_m) = \frac{(2.9\text{ k}\Omega + 0.1\text{ k}\Omega)}{0.4} - (2.9\text{ k}\Omega + 0.1\text{ k}\Omega)$$

$$R_x = \frac{3\text{ k}\Omega}{0.4} - 3\text{ k}\Omega = 7.5\text{ k}\Omega - 3\text{ k}\Omega = 4.5\text{ k}\Omega$$

Step 4: The value of R_x with a 50% deflection is

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) \qquad \text{where } P = \frac{I}{I_m} = 50\% = \frac{50}{100} = 0.5$$

Therefore,

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) = \frac{(2.9\text{ k}\Omega + 0.1\text{ k}\Omega)}{0.5} - (2.9\text{ k}\Omega + 0.1\text{ k}\Omega)$$

$$R_x = \frac{3\text{ k}\Omega}{0.5} - 3\text{ k}\Omega = 6\text{ k}\Omega - 3\text{ k}\Omega = 3\text{ k}\Omega$$

Step 5: The value of R_x with a 75% deflection is

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) \qquad \text{where } P = \frac{I}{I_m} = 75\% = \frac{75}{100} = 0.75$$

Therefore,

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) = \frac{(2.9\text{ k}\Omega + 0.1\text{ k}\Omega)}{0.75} - (2.9\text{ k}\Omega + 0.1\text{ k}\Omega)$$

$$R_x = \frac{3\text{ k}\Omega}{0.75} - 3\text{ k}\Omega = 4\text{ k}\Omega - 3\text{ k}\Omega = 1\text{ k}\Omega$$

Step 6: The value of R_x with a 90% deflection is

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) \qquad \text{where } P = \frac{I}{I_m} = 90\% = \frac{90}{100} = 0.90$$

Therefore,

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) = \frac{(2.9\text{ k}\Omega + 0.1\text{ k}\Omega)}{0.90} - (2.9\text{ k}\Omega + 0.1\text{ k}\Omega)$$

$$R_x = \frac{3\text{ k}\Omega}{0.90} - 3\text{ k}\Omega = 3.333\text{ k}\Omega - 3\text{ k}\Omega = 0.333\text{ k}\Omega$$

Step 7: The value of R_x with a 100% deflection is

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) \qquad \text{where } P = \frac{I}{I_m} = 100\% = \frac{100}{100} = 1$$

Therefore,

$$R_x = \frac{(R_s + R_m)}{P} - (R_s + R_m) = \frac{(2.9\text{ k}\Omega + 0.1\text{ k}\Omega)}{1} - (2.9\text{ k}\Omega + 0.1\text{ k}\Omega)$$

$$R_x = \frac{3\text{ k}\Omega}{1} - 3\text{ k}\Omega = 3\text{ k}\Omega - 3\text{ k}\Omega = 0$$

Example 4.23 *An ohmmeter is designed around a 1mA meter movement and a 3V battery. If the battery voltage decays to 2.8 V because of aging, calculate the resulting error at the midrange on the ohmmeter scale.*

Solution

Step 1: The total internal resistance of the ohmmeter is

$$R_t = E/I = 3/1 \text{ mA} = 3 \text{ k}\Omega$$

Step 2: The Ohmmeter scale should be marked as 3 kΩ at the midrange. An external resistance of 3 kΩ would cause the pointer to deflect to its midscale range.

Step 3: When the battery voltage decreases to 2.8 V, the ohmmeter is adjusted for full scale deflection by reducing R_s, the total internal resistance of the Ohmmeter.

Therefore, $R_t = E/I = 2.8/1 \text{ mA} = 2.8 \text{ k}\Omega$

Step 4: If a 2.8 kΩ resistor is now measured with the ohmmeter, we would expect less than midscale deflection , however the pointer will deflect to midscale which is marked as 3 kΩ. the aging of the battery has caused an incorrect reading.

Step 5: The % error of this reading is given by

$$\%\,\text{Error} = \frac{3\text{ k}\Omega - 2.8\text{ k}\Omega}{3\text{ k}\Omega} \times 100 = \frac{0.2\text{ k}\Omega}{3\text{ k}\Omega} \times 100 = \frac{20}{3} \times 100 = 6.66\%$$

Example 4.24 *Design a series type ohmmeter. The movement requires a 1 mA for full scale deflection and has an internal resistance of 100 Ω. The internal battery used has a voltage of 3 V. The desired value for half scale deflection is 2000 Ω. Calculate*

(a) the values of R_1 and R_2. (b) range of R_2 if the battery voltage varies from 2.8 V to 3.1 V (R_1 is the same as in (a).

Solution Value of R_1 and R_2 can be calculated as

Step 1

$$R_2 = \frac{I_m R_m R_h}{E - I_m R_h} = \frac{1\text{ mA} \times 100 \times 2\text{ k}\Omega}{3 - 1\text{ mA} \times 2\text{ k}\Omega} = \frac{200}{3-2} = 200\ \Omega$$

Step 2: The internal resistance of the ohmmeter is equal to half scale resistance. Therefore,

$$R_h = R_1 + \frac{R_2 R_m}{R_2 + R_m}$$

Therefore,

$$R_1 = R_h - \frac{R_2 R_m}{R_2 + R_m} = 2000\ \Omega - \frac{200 \times 100}{200 + 100}$$

$$R_1 = 2000\ \Omega - \frac{200 \times \cancel{100}}{\cancel{300}} = 2000\ \Omega - \frac{200}{3} = 2000 - 66.6 = 1933.4\ \Omega$$

$R_1 = 1933.4\ \Omega$ that is series resistance

Step 3: The value of resistance R_2 when battery is 2.7 V.

$$R_2 = \frac{I_m R_m R_h}{E - I_m R_h} = \frac{1\text{ mA} \times 100 \times 2\text{ k}\Omega}{2.7 - 1\text{ mA} \times 2\text{ k}\Omega}$$

$$R_2 = \frac{200}{2.7 - 2} = \frac{200}{0.7} = \frac{2000}{7} \approx 285\ \Omega$$

Step 4: The value of resistance R_2 when battery is 3.1 V

$$R_2 = \frac{1\text{ mA} \times 100 \times 2\text{ k}\Omega}{3.1 - 1\text{ mA} \times 2\text{ k}\Omega} = \frac{200}{3.1 - 2} = \frac{200}{1.1} = \frac{2000}{11} \approx 182\ \Omega$$

Multirange Ohmmeter The ohmmeter circuit shown in Fig. 4.30 (a) is only for a single range of resistance measurement. To measure resistance over a wide range of values, we need to extend the ohmmeter ranges. This type of ohmmeter is called a multirange ohmmeter, shown in Fig. 4.31.

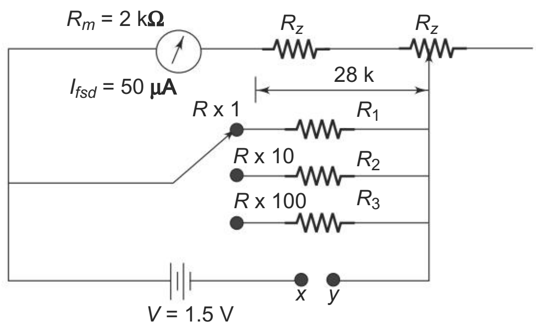


Fig. 4.31 Multirange ohmmeter

SHUNT TYPE OHMMETER 4.22

The shunt type ohmmeter given in Fig. 4.32 consists of a battery in series with an adjustable resistor R_1, and a D'Arsonval movement

The unknown resistance is connected in parallel with the meter, across the terminals A and B, hence the name shunt type ohmmeter.

In this circuit it is necessary to have an ON/OFF switch to disconnect the battery from the circuit when the instrument is not used.

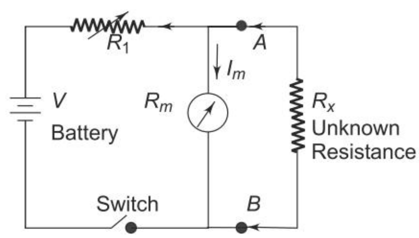


Fig. 4.32 Shunt type ohmmeter

4.22.1 Calibration of the Shunt Type Ohmmeter

To mark the "0" ohms reading on the scale, terminals A and B are shorted, i.e. the unknown resistance $R_x = 0$, and the current through the meter movement is zero, since it is bypassed by the short-circuit. This pointer position is marked as "0" ohms.

Similarly, to mark "∞" on the scale, the terminals A and B are opened, i.e. $R_x = \infty$, and full current flows through the meter movement; by appropriate selection of the value of R_1, the pointer can be made to read full scale deflection current. This position of the pointer is marked "∞" ohms. Intermediate marking can be done by connecting known values of standard resistors to the terminals A and B.

This ohmmeter therefore has a zero mark at the left side of the scale and an ∞ mark at the right side of the scale, corresponding to full scale deflection current as shown in Fig. 4.33.

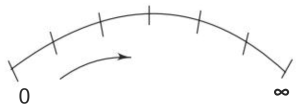


Fig. 4.33 Dial of shunt type ohmmeter

The shunt type ohmmeter is particularly suited to the measurement of low values of resistance. Hence it is used as a test instrument in the laboratory for special low resistance applications.

Example 4.25 *A shunt-type ohmmeter uses a 10 mA basic D'Arsonal movement with an internal resistance of 50 Ω. The battery voltage is 3 V. It is desired to modify the circuit by adding appropriate shunt resistance across the movement so that the instrument indicates 10 Ω at the midpoint scale. Calculate*

(a) the value of shunt resistance (b) value of current limiting resistance R_1.

Solution

Step 1: For half scale definition

$$I_h = 0.5 \text{ mA} \times I_m = 0.5 \text{ mA} \times 10 \text{ mA} = 5 \text{ mA}$$

Step 2: The voltage across the movement

$$V_m = I_m R_m = 5 \text{ mA} \times 50\ \Omega = 250 \text{ mV}$$

Step 3: From the circuit diagram of a shunt-type ohmmeter it is seen that

voltage across unknown resistance = Voltage across meter movement

Step 4: Therefore, current through the unknown resistance

$$I_x = \frac{V_m}{R_h} = \frac{250 \text{ mV}}{10\ \Omega} = 25 \text{ mA}$$

Step 5: Current through the shunt = $I_{sh} = I_x - I_m$

$$I_{sh} = 25 \text{ mA} - 5 \text{ mA} = 20 \text{ mA}$$

Step 6: Therefore, the value of shunt resistance

$$R_{sh} = \frac{250 \text{ mV}}{20 \text{ mA}} = 12.5\ \Omega$$

Step 7:

b) The total battery current

$$I_t = I_x + I_m + I_{sh} = 25 \text{ mA} + 5 \text{ mA} + 20 \text{ mA} = 50 \text{ mA}$$

Step 8: Voltage drop across the limiting resistor

$$= 3 - 250 \text{ mV}$$
$$= 2.75 \text{ V}$$

Step 9: Therefore

$$R_1 = \frac{2.75}{50 \text{ mA}} = \frac{275}{50 \text{ mA}} \times \frac{1}{100} = \frac{275}{5} = 55\ \Omega$$

Example 4.26 *(a) Determine the current through the meter I_m when a 20 Ω resistor is connected across the terminals 'x' and 'y' is measured on $R \times 1$ range.*

(b) Show that this same current flows through the meter movement when a 200 Ω resistor is measured on the $R \times 100$ range.

(c) When a 2 kΩ resistor is measured on the $R \times 100$.

Solution From Fig. 4.31

Step 1:

(a) When the ohmmeter is set on the $R \times 1$ range circuit as shown in Fig Ex4.26 (a).

The voltage across the parallel combination is calculated as

$$V = 3 \times \frac{10\ \Omega}{10\ \Omega + 20\ \Omega} = \frac{30\ \Omega}{30\ \Omega} = 1 \text{ V}$$

Step 2: The current through the meter is calculated as

$$I_m = \frac{1 \text{ V}}{30 \text{ K}} = \frac{1 \text{ mA}}{30} = \frac{100\cancel{0}\ \mu\text{A}}{3\cancel{0}} = 33.3\ \mu\text{A}$$

(b) When the ohmmeter is set on the $R \times 10$ range, circuit as shown in Ex4.26(b).

Step 3: The voltage across the parallel combination is calculated as

$$V = 3 \times \frac{100\ \Omega}{100\ \Omega + 200\ \Omega} = \frac{300\ \Omega}{300\ \Omega} = 1 \text{ V}$$

Step 4: Therefore,

$$I_m = \frac{1 \text{ V}}{30 \text{ k}\Omega} = 33.3\ \mu\text{A}$$

(c) When the ohmmeter is set on $R \times 100$, the circuit is shown in Fig. Ex4.26(c).

Step 5: The voltage across the parallel combination is calculated as

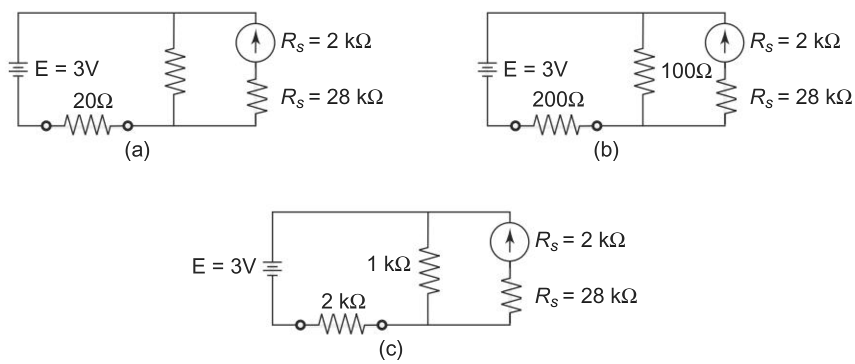


Fig. Ex4.26

$$V = 3 \times \frac{1000\ \Omega}{1000\ \Omega + 2000\ \Omega} = \frac{3000\ \Omega}{3000\ \Omega} = 1\ \text{V}$$

Therefore, $$I_m = \frac{1\ \text{V}}{30\ \text{k}\Omega} = 33.3\ \mu\text{A}$$

CALIBRATION OF DC INSTRUMENT 4.23

The process of calibration involves the comparison of a given instrument with a standard instrument, to determine its accuracy. A dc voltmeter may be calibrated with a standard, or by comparison with a potentiometer. The circuit in Fig. 4.34 is used to calibrate a dc voltmeter; where a test voltmeter reading V is compared to the voltage drop across R. The voltage drop across R is accurately measured with the help of a standard meter. A rheostat, shown in Fig. 4.34, is used to limit the current.

A voltmeter tested with this method can be calibrated with an accuracy of ±0.01%.

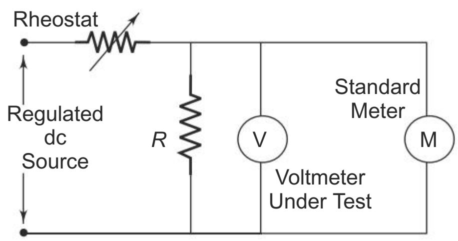


Fig. 4.34 Calibration of voltmeter

CALIBRATION OF OHMMETER 4.24

An ohmmeter is generally considered to be an instrument of moderate accuracy and low precision. A rough calibration may be done by measuring a standard resistance and noting the readings on the ohmmeter. Doing this for several points on the ohmmeter scale and on several ranges allows one to obtain an indication of the accuracy of the instrument.

MULTIMETER 4.25

A multimeter is basically a PMMC meter. To measure dc current the meter acts as an ammeter with a low series resistance.

Range changing is accomplished by shunts in such a way that the current passing through the meter does not exceed the maximum rated value.

A multimeter consists of an ammeter, voltmeter and ohmmeter combined, with a function switch to connect the appropriate circuit to the D'Arsonval movement.

Figure 4.35 shows a meter consisting of a dc milliammeter, a dc voltmeter, an ac voltmeter, a microammeter, and an ohmmeter.

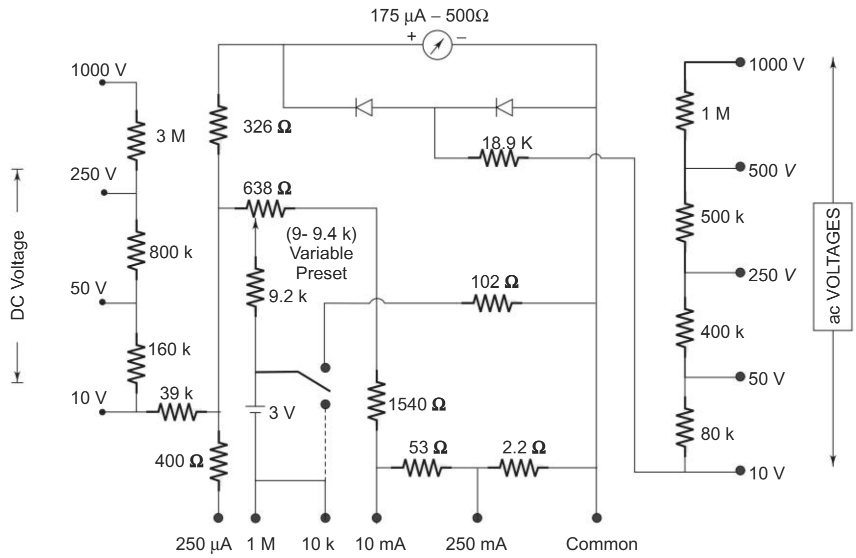


Fig. 4.35 Diagram of a multimeter

Microammeter Figure 4.36 shows a circuit of a multimeter used as a microammeter.

DC Ammeter Figure 4.37 shows a multimeter used as a dc ammeter.

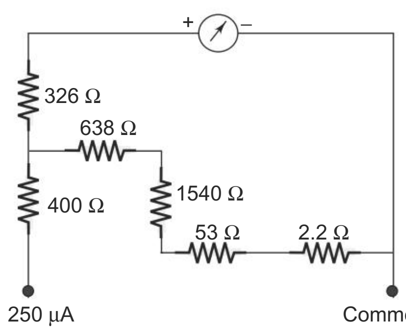


Fig. 4.36 Microammeter section of a multimeter

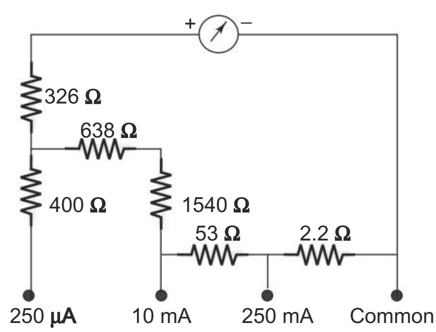


Fig. 4.37 dc ammeter section of a multimeter

DC Voltmeter Figure 4.38 shows the dc voltmeter section of a multimeter.

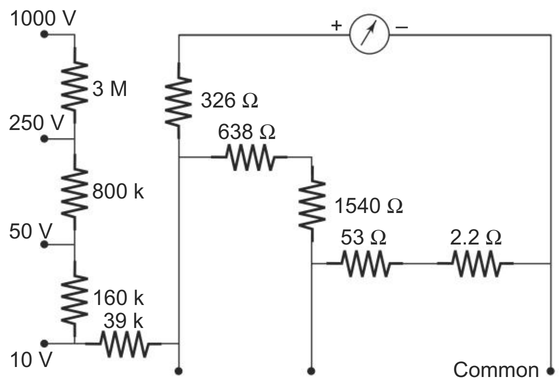


Fig. 4.38 DC voltmeter section of a multimeter

AC Voltmeter Figure 4.39 shows the ac voltmeter section of a multimeter. To measure ac voltage, the output ac voltage is rectified by a half wave rectifier before the current passes through the meter. Across the meter, the other diode serves as protection. The diode conducts when a reverse voltage appears across the diodes, so that current bypasses the meter in the reverse direction.

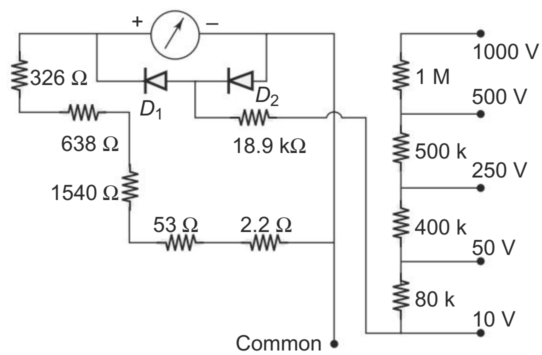


Fig. 4.39 AC voltmeter section of a multimeter

Ohmmeter Referring to Fig. 4.40 which shows the ohmmeter section of a multimeter, in the 10 k range the 102 Ω resistance is connected in parallel with the total circuit resistance and in the 1 MΩ range the 102 Ω resistance is totally disconnected from the circuit.

Therefore, on the 1 M range the half scale deflection is 10 k. Since on the 10 k range, the 102 Ω resistance is connected across the total resistance, therefore, in this range, the half scale deflection is 100 Ω. The measurement of resistance is done by applying a small voltage installed within the meter. For the 1 M range, the internal resistance is 10 kΩ, i.e. value at midscale, as shown in Fig. 4.41. And for the 10 k range, the internal resistance is 100 Ω, i.e. value at mid-scale as shown in Fig. 4.42.

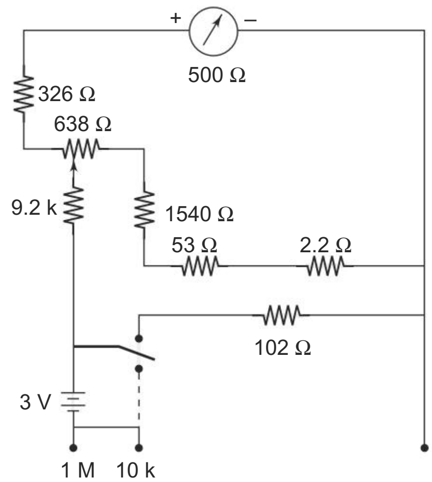


Fig. 4.40 Ohmmeter section of a multimeter

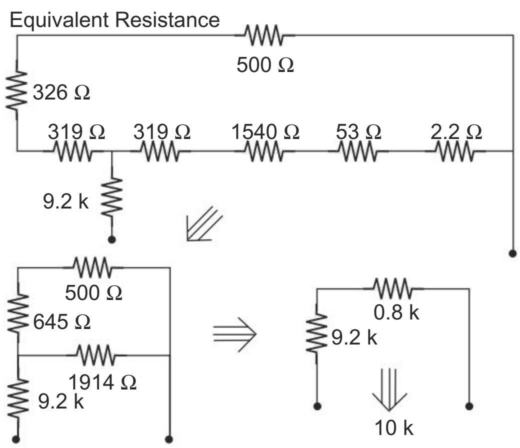


Fig. 4.41 Equivalent resistance on 1 M Ω range

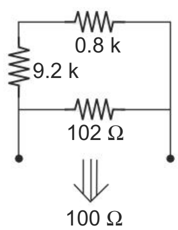


Fig. 4.42 Half scale deflection is 100 Ω on 10k range

The range of an ohmmeter can be changed by connecting the switch to a suitable shunt resistance. By using different values of shunt resistance, different ranges can be obtained.

By increasing the battery voltage and using a suitable shunt, the maximum values which the ohmmeter reads can be changed.

MULTIMETER OPERATING INSTRUCTIONS 4.26

The combination volt-ohm-milliammeter is a basic tool in any electronic laboratory. The proper use of this instrument increases its accuracy and life. The following precautions should be observed.

1. To prevent meter overloading and possible damage when checking voltage or current, start with the highest range of the instrument and move down the range successively.
2. For higher accuracy, the range selected should be such that the deflection falls in the upper half on the meter scale.
3. For maximum accuracy and minimum loading, choose a voltmeter range such that the total voltmeter resistance (ohms per volt × full scale voltage) is at least 100 times the resistance of the circuit under test.
4. Make all resistance readings in the uncrowded portion on the meter scale, whenever possible.
5. Take extra precautions when checking high voltages and checking current in high voltage circuits.
6. Verify the circuit polarity before making a test, particularly when measuring dc current or voltages.
7. When checking resistance in circuits, be sure power to the circuit is switched off, otherwise the voltage across the resistance may damage the meter.
8. Renew ohmmeter batteries frequently to insure accuracy of the resistance scale.
9. Recalibrate the instrument at frequent intervals.
10. Protect the instrument from dust, moisture, fumes and heat.

Review Questions

1. Explain how a PMMC can be used as a basic voltmeter.
2. Explain with a diagram the operation of a multirange voltmeter. State the limitations of a multirange voltmeter.
3. Explain the operation of an Ayrton shunt.
4. Compare a multirange voltmeter with the Ayrton shunt voltmeter.
5. State the drawbacks of using an Aryton shunt.
6. Why is an Ayrton shunt called a universal shunt.
7. Define sensitivity of voltmeters. What is the significance of sensitivity in voltmeters?
8. State the effects of using a voltmeter of low sensitivity.
9. Explain the above with examples of loading effect.
10. Explain with a diagram the working of a Transistor Voltmeter (TVM). What are the drawbacks of a transistor voltmeter?
11. Explain why an FET is used at the input stage of a transistor voltmeter.
12. Explain why a transistor voltmeter cannot be used for measurement in the μV range.
13. Describe with a diagram how voltage in μV range is measured.
14. Explain the principle of operation of a chopper. Explain the use of a chopper in microvoltmeter.
15. Describe with a diagram the operation of a chopper type microvoltmeter.
16. Describe the operation of an electronic voltmeter using an IC OPAMP.
17. Explain the functions of a diode used in an electronic voltmeter. Explain the function of offset used in an electronic voltmeter.
18. Explain with diagram the operation of a dc differential voltmeter.

19. Explain how a PMMC can be used as an ac voltmeter.
20. State why silicon diodes are preferred as rectifiers in ac voltmeter.
21. Explain the operation of a full wave rectifier type ac voltmeter.
22. Explain with a diagram the operation of a half wave rectifier type ac voltmeter.
23. Why is a PMMC movement shunted by a resistor when used as an ac voltmeter?
24. Compare sensitivity of an ac voltmeter with that of the dc voltmeter.
25. Compare sensitivity of an ac voltmeter using FWR with that of a HWR.
26. Explain with a diagram how a multirange ac voltmeter can be constructed using a PMMC.
27. Explain with a diagram the operations of an average responding voltmeter. Where is it used?
28. Explain with a diagram the operations of a peak responding voltmeter. Where is it used?
29. Compare average responding voltmeters with peak responding voltmeters.
30. Explain the operating principle of a true RMS voltmeter.
31. Explain with diagram the operation of true RMS voltmeter.
32. Compare a true RMS voltmeter with an ac voltmeter.
33. What is the need of using a squaring device in a true RMS meter? Explain with a diagram the operation of a squaring device.
34. Compare a true RMS meter with an average responding meter.
35. Explain with a diagram the operation of a series type ohmmeter. Explain with a diagram how an ohmmeter is calibrated.
36. Why is there a '0' mark on the right-hand side for a series type ohmmeter? How can you estimate the internal resistance of a series type ohmmeter from its dial?
37. Explain with a diagram the operation of a shunt type ohmmeter.
38. Compare series type and shunt type ohmmeters.
39. How can you distinguish between a series type ohmmeter and a shunt type ohmmeter from dial calibration?
40. Define sensitivity of a multimeter.
41. Explain with the help of a diagram, the working of a simple multimeter.
42. Explain with the help of a diagram the various sections of a multimeter.
43. Draw a practical multimeter.
44. Explain with a diagram how a multimeter can be used as a voltmeter and ammeter.
45. Explain with a diagram how a multimeter can be used for measuring resistance.

Multiple Choice Questions

1. The instrument required to measure voltage is
   (a) ohmmeter (b) ammeter
   (c) voltmeter (d) wattmeter
2. A D'Arsonval movement is
   (a) taut band
   (b) PMMC
   (c) electrodynamometer
   (d) moving iron type
3. To select the range, a multirange voltmeter uses
   (a) double pole double throw switch
   (b) make before break type switch
   (c) single pole double throw switch
   d) simple switch
4. The sensitivity of a voltmeter is defined as
   (a) Ω / V (b) V / Ω
   (c) I / Ω (d) Ω / I

5. Loading effect in a voltmeter can be avoided by
   (a) using an accurate and precise instrument
   (b) using a low sensitivity voltmeter
   (c) using a high sensitivity voltmeter
   (d) using high voltage range
6. The input stage of TVM consists of
   (a) UJT stage (b) FET stage
   (c) BJT stage (d) SCR stage
7. TVM is used to measure
   (a) dc mV (b) dc μV
   (c) ac μV (d) ac mV
8. Chopper type voltmeter is used to measure
   (a) dc μV, (b) dc mV,
   (c) ac μV, (d) ac mV.
9. A diode used as rectifier in ac voltmeter should have
   (a) high forward current and low reverse currents
   (b) high forward current and high reverse current
   (c) low forward current and high reverse current
   (d) low forward current and low reverse current
10. The ac voltmeter using PMMC measures
    (a) true RMS voltage
    (b) peak voltage
    (c) average voltage
    (d) instantaneous voltage
11. To move the operating point of rectifier used in an ac voltmeter in the linear region, the meter is shunted by
    (a) capacitor (b) diode
    (c) inductor (d) resistor
12. A true RMS voltmeter measures
    (a) average value
    (b) instantaneous value
    (c) RMS value
    (d) peak value
13. The ohms per volt rating on ac ranges as compared to the same rating on dc ranges is
    (a) less
    (b) more
    (c) equal
    (d) none of the above
14. In a 20 kΩ / V sensitivity multimeter, the input resistance for measuring ac voltage in 10 V fsd is
    (a) 200 kΩ (b) 20 kΩ
    (c) 10 kΩ (d) 2 kΩ
15. Ac measurement is achieved by connecting a/an in series with a PMMC.
    (a) resistor (b) diode
    (c) inductor (d) capacitor
16. The internal resistance of an ohmmeter can be estimated from
    (a) 0 deflection
    (b) full scale deflection
    (c) half scale deflection
    (d) quarter deflection

Practice Problems

1. What series resistance must be used to extend the 0–200 V range of a 20000 Ω/V meter to a 2000 V? What must be the power of this resistor?
2. Calculate the value of a series resistor used to extend the 0–100 V range of a 20 kΩ / V Voltmeter to a 500 V range. Calculate the power of the resistor.
3. A basic D'Arsonval movement with a full scale deflection of 50 μA and having an internal resistance of 500 Ω is available. It is to be converted into a 0–1 V, 0–5 V, 0–20 V, 0–100 V multirange voltmeter using individual multipliers for each range. Calculate the values of the individual resistor.
4. A basic D'Arsonval movement with a full scale deflection of 50 μA and an internal resistance of 1800 Ω is available. It is to be converted into a 0–1 V, 0–5 V, 0–25 V and 0–225 V multirange voltmeter using individual

multipliers for each range. Calculate the values of the individual resistors.

5. Convert a basic D'Arsonval movement with an internal resistance of 100 Ω and a full scale deflection of 1 mA into a multirange dc voltmeter with voltage ranges of 0–1 V, 0–10 V, 0–50 V .
6. A meter movement has an internal resistance of 100 Ω and requires 1 mA dc full scale deflection. Shunting resistor R_{sh} placed across the movement has a value of 100 Ω. Diodes $D1$ and D_2 have an average forward resistance of 400 Ω and are assumed to have infinite reverse resistance in the reverse direction. For 10 V ac range, calculate (i) the value of the multiplier, (ii) the voltmeter sensitivity on ac range. (Rs = 1800, S = 225 Ω/V). Refer to Fig. 4.43.

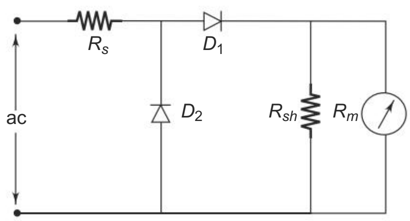


Fig. 4.43

7. The circuit diagram of Fig. 4.44 shows a full wave rectifier ac voltmeter. The meter movement has an internal resistance of 250 Ω and required 1 mA for full scale deflection. The diodes each have a forward resistance of 50 Ω and infinite reverse resistance.

Calculate:
(i) the series resistance required for full scale meter deflection when 25 V rms is applied to the meter terminals.
(ii) the ohms per volt rating of this ac voltmeter.

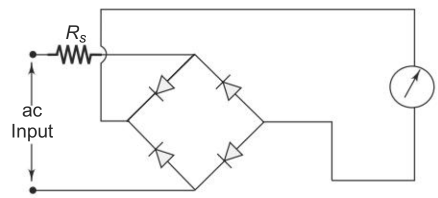


Fig. 4.44

7. A series ohmmeter uses a 50 Ω basic movement requiring a full scale deflection of 1 mA. The internal battery voltage is 3 V. The desired scale marking for half scale deflection is 2000 Ω.
Calculate
(i) values of R_1 and R_2
(ii) maximum value of R_2 to compensate for a 10% drop in battery.
8. A series type ohmmeter is designed to operate with a 6 V battery. The meter movement has an internal resistance of 2 kΩ and requires a current of 100 μA for full scale deflection. The value of R_1 is 49 k.
(i) Assuming the battery voltage has fallen to 5.9 V, calculate the value of R_2 required to "0" the meter.
(ii) Under the condition mentioned in (i), an unknown resistance is connected to the meter, causing a 60% deflection. Calculate the value of the unknown resistance.

Further Reading

1. John. H. Fasal, *Simplified Electronics Measurements*, Hayden Book Co., 1971.
2. *Handbook of Electronic Measurements*, vols. I & II, Polytechnic Institute of Brooklyn, 1956.
3. Larry D. Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley and Sons, 1987.
4. W.D. Copper and A.D. Helfrick, *Electronic Instrumentation and Measurement Techniques*, 3rd Edition, 1985. Prentice-Hall of India.

Chapter 5

Digital Voltmeters

INTRODUCTION 5.1

Digital voltmeters (DVMs) are measuring instruments that convert analog voltage signals into a digital or numeric readout. This digital readout can be displayed on the front panel and also used as an electrical digital output signal.

Any DVM is capable of measuring analog dc voltages. However, with appropriate signal conditioners preceding the input of the DVM, quantities such as ac voltages, ohms, dc and ac current, temperature, and pressure can be measured. The common element in all these signal conditioners is the dc voltage, which is proportional to the level of the unknown quantity being measured. This dc output is then measured by the DVM.

DVMs have various features such as speed, automation operation and programability. There are several varieties of DVM which differ in the following ways:

1. Number of digits
2. Number of measurements
3. Accuracy
4. Speed of reading
5. Digital output of several types.

The DVM displays ac and dc voltages as discrete numbers, rather than as a pointer on a continuous scale as in an analog voltmeter. A numerical readout is advantageous because it reduces human error, eliminates parallax error, increases reading speed and often provides output in digital form suitable for further processing and recording. With the development of IC modules, the size, power requirements and cost of DVMs have been reduced, so that DVMs compete with analog voltmeters in portability and size. Their outstanding qualities are their operating and performance characteristics, as detailed below.

1. Input range from + 1.000 V to + 1000 V with automatic range selection and overload indication
2. Absolute accuracy as high as ±0.005% of the reading
3. Resolution 1 part in million (1 μV reading can be read or measured on 1 V range)
4. Input resistance typically 10 MΩ, input capacitance 40 pF
5. Calibration internally from stabilised reference sources, independent of measuring circuit

6. Output in BCD form, for print output and further digital processing. Optional features may include additional circuitry to measure current, ohms and voltage ratio.

RAMP TECHNIQUE 5.2

The operating principle is to measure the time that a linear ramp takes to change the input level to the ground level, or vice-versa. This time period is measured with an electronic time-interval counter and the count is displayed as a number of digits on an indicating tube or display. The operating principle and block diagram of a ramp type DVM are shown in Figs 5.1 and 5.2.

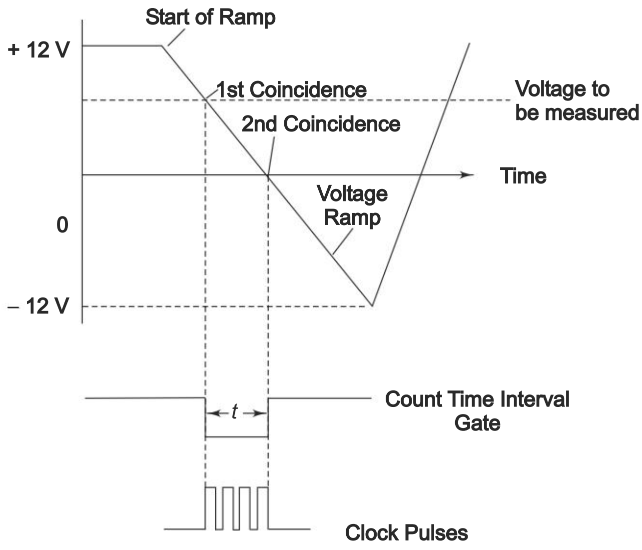


Fig. 5.1 Voltage to time conversion

The ramp may be positive or negative; in this case a negative ramp has been selected.

At the start of the measurement a ramp voltage is initiated (counter is reset to 0 and sampled rate multivibrator gives a pulse which initiates the ramp generator). The ramp voltage is continuously compared with the voltage that is being measured. At the instant these two voltage become equal, a coincidence circuit generates a pulse which opens a gate, i.e. the input comparator generates a start pulse. The ramp continues until the second comparator circuit senses that the ramp has reached zero value. The ground comparator compares the ramp with ground. When the ramp voltage equals zero or reaches ground potential, the ground comparator generates a stop pulse. The output pulse from this comparator closes the gate. The time duration of the gate opening is proportional to the input voltage value.

In the time interval between the start and stop pulses, the gate opens and the oscillator circuit drives the counter. The magnitude of the count indicates the magnitude of the input voltage, which is displayed by the readout. Therefore, the voltage is converted into time and the time count represents the magnitude of the voltage. The sample rate multivibrator determines the rate of cycle of measurement. A typical value is 5 measuring cycles per second, with an accuracy of ± 0.005% of the reading. The sample rate circuit provides an initiating pulse for the ramp generator to start its next ramp voltage. At the same time a reset pulse is generated, which resets the counter to the zero state.

Any DVM has a fundamental cycle sequence which involves sampling, displaying and reset sequences.

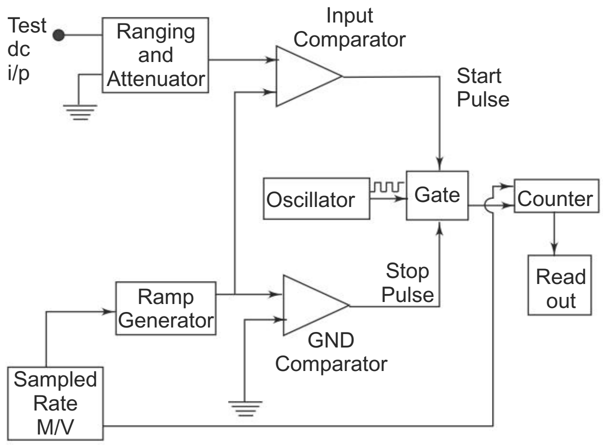


Fig. 5.2 Block diagram of ramp type DVM

Advantages and Disadvantages The ramp technique circuit is easy to design and its cost is low. Also, the output pulse can be transmitted over long feeder lines. However, the single ramp requires excellent characteristics regarding linearity of the ramp and time measurement. Large errors are possible when noise is superimposed on the input signal. Input filters are usually required with this type of converter.

DUAL SLOPE INTEGRATING TYPE DVM (VOLTAGE TO TIME CONVERSION) 5.3

In ramp techniques, superimposed noise can cause large errors. In the dual ramp technique, noise is averaged out by the positive and negative ramps using the process of integration.

Principle of Dual Slope Type DVM As illustrated in Fig. 5.3, the input voltage 'e_i' is integrated, with the slope of the integrator output proportional to the test input voltage. After a fixed time, equal to t_1, the input voltage is disconnected and the integrator input is connected to a negative voltage $-e_r$. The integrator output will have a negative slope which is constant and proportional to the magnitude of the input voltage. The block diagram is given in Fig. 5.4.

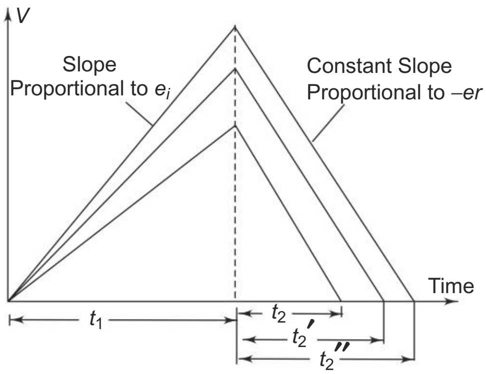


Fig. 5.3 Basic principle of dual slope type DVM

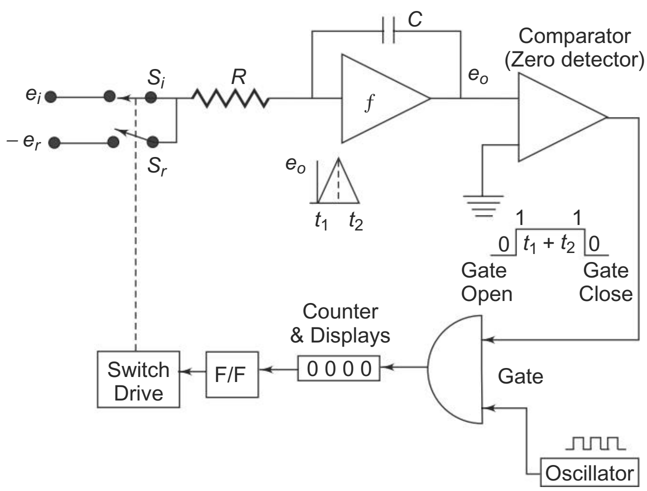


Fig. 5.4 Block diagram of a dual slope type DVM

At the start a pulse resets the counter and the F/F output to logic level '0'. S_i is closed and S_r is open. The capacitor begins to charge. As soon as the integrator output exceeds zero, the comparator output voltage changes state, which opens the gate so that the oscillator clock pulses are fed to the counter. (When the ramp voltage starts, the comparator goes to state 1, the gate opens and clock pulse drives the counter.) When the counter reaches maximum count, i.e. the counter is made to run for a time 't_1' in this case 9999, on the next clock pulse all digits go to 0000 and the counter activates the F/F to logic level '1'. This activates the switch drive, e_i is disconnected and $-e_r$ is connected to the integrator. The integrator output will have a negative slope which is constant, i.e. integrator output now decreases linearly to 0 volts. Comparator output state changes again and locks the gate. The discharge time t_2 is now proportional to the input voltage. The counter indicates the count during time t_2. When the negative slope of the integrator reaches zero, the comparator switches to state 0 and the gate closes, i.e. the capacitor C is now discharged with a constant slope. As soon as the comparator input (zero detector) finds that e_o is zero, the counter is stopped. The pulses counted by the counter thus have a direct relation with the input voltage.

During charging

$$e_o = -\frac{1}{RC}\int_0^t e_i\,dt = -\frac{e_i\,t_1}{RC} \tag{5.1}$$

During discharging

$$e_o = \frac{1}{RC}\int_0^{t_2} -e_r dt = -\frac{e_r\,t_2}{RC} \tag{5.2}$$

Subtracting Eqs 5.2 from 5.1 we have

$$e_o - e_o = \frac{-e_r\,t_2}{RC} - \left(\frac{-e_i\,t_1}{RC}\right)$$

$$0 = \frac{-e_r\,t_2}{RC} - \left(\frac{-e_i\,t_1}{RC}\right)$$

$\Rightarrow$
$$\frac{e_r\,t_2}{RC} = \frac{e_i\,t_1}{RC}$$

$\therefore$
$$e_i = e_r\frac{t_2}{t_1} \tag{5.3}$$

If the oscillator period equals T and the digital counter indicates n_1 and n_2 counts respectively,

$\therefore$
$$e_i = \frac{n_2\,T}{n_1\,T}e_r \quad \text{i.e.} \quad e_i = \frac{n_2}{n_1}e_r$$

Now, n_1 and e_r are constants. Let $K_1 = \dfrac{e_r}{n_1}$. Then $e_i = K_1\,n_2$ (5.4)

From Eq. 5.3 it is evident that the accuracy of the measured voltage is independent of the integrator time constant. The times t_1 and t_2 are measured by the count of the clock given by the numbers n_1 and n_2 respectively. The clock oscillator period equals T and if n_1 and e_r are constants, then Eq. 5.4 indicates that the accuracy of the method is also independent of the oscillator frequency.

The dual slope technique has excellent noise rejection because noise and superimposed ac are averaged out in the process of integration. The speed and accuracy are readily varied according to specific requirements; also an accuracy of ± 0.05% in 100 ms is available.

INTEGRATING TYPE DVM (VOLTAGE TO FREQUENCY CONVERSION) 5.4

The principle of operation of an integrating type DVM is illustrated in Fig. 5.5.

A constant input voltage is integrated and the slope of the output ramp is proportional to the input voltage. When the output reaches a certain value, it is discharged to 0 and another cycle begins. The frequency of the output waveform is propor-tional to the input voltage. The block diagram is illustrated in Fig. 5.6.

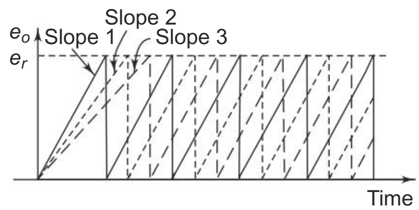


Fig. 5.5 Voltage to frequency conversion

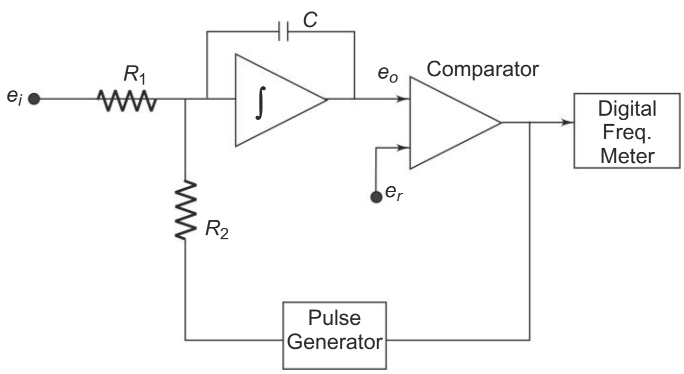


Fig. 5.6 Block diagram of an integrating type DVM

The input voltage produces a charging current, e_i/R_1, that charges the capacitor 'C' to the reference voltage e_r. When e_r is reached, the comparator changes state, so as to trigger the precision pulse generator. The pulse generator produces a pulse of precision charge content that rapidly discharges the capacitor. The rate of charging and discharging produces a signal frequency that is directly proportional to e_i.

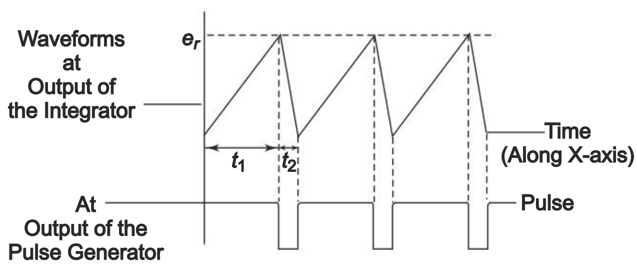


Fig. 5.7

The voltage-frequency conversion can be considered to be a dual slope method, as shown in Fig. 5.7.

Referring to Eq. 5.3 we have

$$e_i = \frac{e_r\, t_2}{t_1}$$

But in this case e_r and t_2 are constants.

Let $$K_2 = e_r\, t_2$$

$\therefore$ $$e_i = K_2\left(\frac{1}{t_1}\right) = K_2(f_0)$$

The output frequency is proportional to the input voltage e_i. This DVM has the disadvantage that it requires excellent characteristics in linearity of the ramp. The ac noise and supply noise are averaged out.

Example 5.1 *An integrator contains a 100 kΩ and 1 μF capacitor. If the voltage applied to the integrator input is 1 V, what voltage will be present at the output of the integrator after 1 s.*

Solution Using the equation

$$e_o = \frac{e_i \times t_1}{RC} = \frac{1 \times 1\ \text{s}}{100\ \text{k} \times 1\ \mu\text{F}} = \frac{1}{0.1} = 10\ \text{V}$$

Example 5.2 *Now if a reference voltage is applied to the integrator of the above example at time t_1 is 5 V in amplitude, what is the time interval of t_2?*

Solution Using the equation

$$\frac{e_i \times t_1}{RC} = \frac{e_r \times t_2}{RC}$$

Therefore, $t_2 = \dfrac{e_i}{e_r} \times t_1$; $\quad t_2 = \dfrac{1 \times 1}{5} = 0.2\ \text{s}$

Example 5.3 *An integrator consists of a 100 kΩ and 2 μF capacitor. If the applied voltage is 2V, what will be the output of the integrator after 2 seconds?*

Solution Given, $R = 100\ \text{k}\Omega$, $C = 2\ \mu\text{F}$, $e_1 = 2$ V and $t_1 = 2$ s

Using the equation,

$$e_0 = \frac{e_1 \times t_1}{R \times C} = \frac{2\ \text{V} \times 2\ \text{s}}{100\ \text{k} \times 2\ \mu\text{F}} = \frac{4}{200 \times 10^3 \times 10^{-6}} = \frac{4 \times 10^3}{200} = \frac{4000}{200} = 20\ \text{V}$$

Example 5.4 *Now if a reference voltage of 10 V is applied to the integrator of the above example (Ex 5.3) at time t_1, what is the time interval of t_2?*

Solution Given reference voltage 10 V.

$$\frac{e_1 \times t_1}{R \times C} = \frac{e_2 \times t_2}{R \times C}$$

therefore, $$t_2 = \frac{e_1 \times t_1}{e_2} = \frac{2 \times 2\ \text{s}}{10} = 0.4\ \text{s}$$

MOST COMMONLY USED PRINCIPLES OF ADC (ANALOG TO DIGITAL CONVERSION) 5.5

5.5.1 Direct Compensation

The input signal is compared with an internally generated voltage which is increased in steps starting from zero. The number of steps needed to reach the full compensation is counted. A simple compensation type is the staircase ramp.

The Staircase Ramp The basic principle is that the input signal V_i is compared with an internal staircase voltage, V_c, generated by a series circuit consisting of a pulse generator (clock), a counter counting the pulses and a digital to analog converter, converting the counter output into a dc signal. As soon as V_c is equal to V_i, the input comparator closes a gate between the clock and the counter, the counter stops and its output is shown on the display. The basic block diagram is shown in Fig. 5.8.

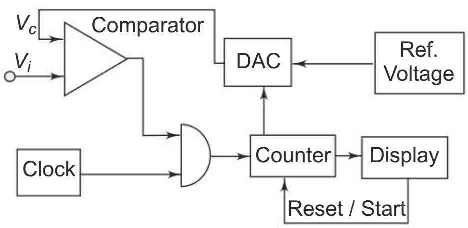


Fig. 5.8 Block diagram of a staircase ramp type

Operation of the Circuit The clock generates pulses continuously. At the start of a measurement, the counter is reset to 0 at time t_1 so that the output of the digital to analog converter (DAC) is also 0. If V_i is not equal to zero, the input comparator applies an output voltage that opens the gate so that clock pulses are passed on to the counter through the gate. The counter starts counting and the DAC starts to produce an output voltage increasing by one small step at each count of the counter. The result is a staircase voltage applied to the second input of the comparator, as shown in Fig. 5.9.

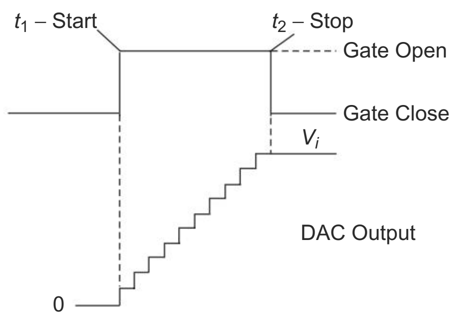


Fig. 5.9 Staircase waveform

This process continues until the staircase voltage is equal to or slightly greater than the input voltage V_i. At that instant t_2, the output voltage of the input comparator changes state or polarity, so that the gate closes and the counter is stopped.

The display unit shows the result of the count. As each count corresponds to a constant dc step in the DAC output voltage, the number of counts is directly proportional to V_c and hence to V_i. By appropriate choice of reference voltage, the step height of the staircase voltage can be determined. For example, each count can represent 1 mV and direct reading of the input voltage in volts can be realised by placing a decimal point in front of the 10 decade.

The advantages of a staircase type DVM are as follows:

1. Input impedance of the DAC is high when the compensation is reached.
2. The accuracy depends only on the stability and accuracy of the voltage and DAC. The clock has no effect on the accuracy.

The disadvantages are the following:

1. The system measures the instantaneous value of the input signal at the moment compensation is reached. This means the reading is rather unstable, i.e. the input signal is not a pure dc voltage.
2. Until the full compensation is reached, the input impedance is low, which can influence the accuracy.

SUCCESSIVE APPROXIMATIONS 5.6

The successive approximation principle can be easily understood using a simple example; the determination of the weight of an object. By using a balance and placing the object on one side and an approximate weight on the other side, the weight of the object is determined.

If the weight placed is more than the unknown weight, the weight is removed and another weight of smaller value is placed and again the measurement is performed. Now if it is found that the weight placed is less than that of the object, another weight of smaller value is added to the weight already present, and the measurement is performed. If it is found to be greater than the unknown weight the added weight is removed and another weight of smaller value is added. In this manner by adding and removing the appropriate weight, the weight of the unknown object is determined. The successive approximation DVM works on the same principle. Its basic block diagram is shown in Fig. 5.10. When the start pulse signal activates the control circuit, the successive approximation register (SAR) is cleared. The output of the SAR is 00000000. V_{out} of the D/A converter is 0. Now, if $V_{in} > V_{out}$ the comparator output is positive. During the first clock pulse, the control circuit sets the D_7 to 1, and V_{out} jumps to the half reference voltage. The SAR output is 10000000. If V_{out} is greater than V_{in}, the comparator output is negative and the control circuit resets D_7. However, if V_{in} is greater than V_{out}, the comparator output is positive and the control circuits keeps D_7 set. Similarly the rest of the bits beginning from D_7 to D_0 are set and tested. Therefore, the measurement is completed in 8 clock pulses.

Table 5.1

$V_{in} = 1\ V$	*Operation*	D_7	D_6	D_5	D_4	D_3	D_2	D_1	D_0	*Compare*	*Output*	*Voltage*
00110011	D_7 Set	1	0	0	0	0	0	0	0	$V_{in} < V_{out}$	D_7 Reset	2.5
”	D_6 Set	0	1	0	0	0	0	0	0	$V_{in} < V_{out}$	D_6 Reset	1.25
”	D_5 Set	0	0	1	0	0	0	0	0	$V_{in} > V_{out}$	D_5 Set	0.625
”	D_4 Set	0	0	1	1	0	0	0	0	$V_{in} > V_{out}$	D_4 Set	0.9375
”	D_3 Set	0	0	1	1	1	0	0	0	$V_{in} < V_{out}$	D_3 Reset	0.9375
”	D_2 Set	0	0	1	1	0	1	0	0	$V_{in} < V_{out}$	D_2 Reset	0.9375
”	D_1 Set	0	0	1	1	0	0	1	0	$V_{in} > V_{out}$	D_1 Set	0.97725
”	D_0 Set	0	0	1	1	0	0	1	1	$V_{in} > V_{out}$	D_0 Set	0.99785

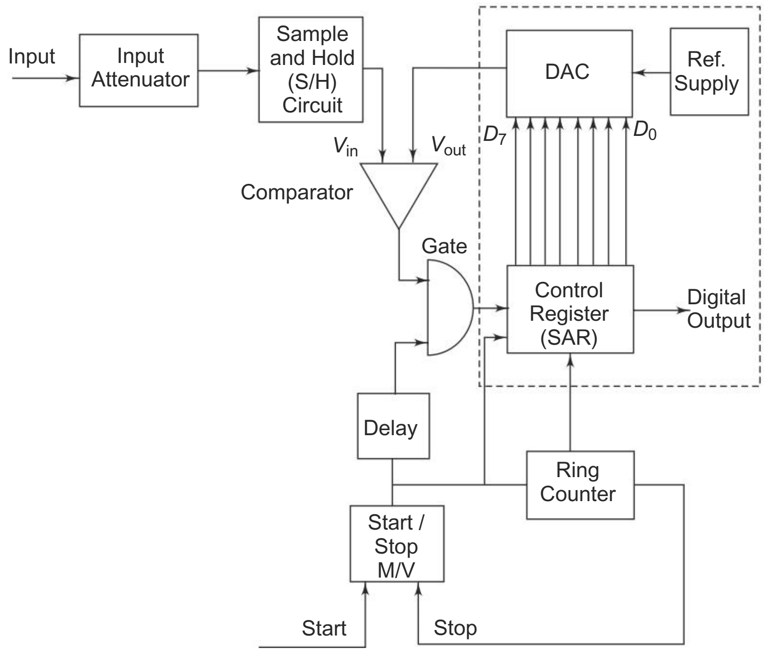


Fig. 5.10 Successive approximation DVM

At the beginning of the measurement cycle, a start pulse is applied to the start-stop multivibrator. This sets a 1 in the MSB of the control register and a 0 in all bits (assuming an 8-bit control) its reading would be 10000000. This initial setting of the register causes the output of the D/A converter to be half the reference voltage, i.e. 1/2 V. This converter output is compared to the unknown input by the comparator. If the input voltage is greater than the converter reference voltage, the comparator output produces an output that causes the control register to retain the 1 setting in its MSB and the converter continues to supply its reference output voltage of 1/2 V_{ref}.

The ring counter then advances one count, shifting a 1 in the second MSB of the control register and its reading becomes 11000000. This causes the D/A converter to increase its reference output by 1 increment to 1/4 V, i.e. 1/2 V + 1/4 V, and again it is compared with the unknown input. If in this case the total reference voltage exceeds the unknown voltage, the comparator produces an output that causes the control register to reset its second MSB to 0. The converter output then returns to its previous value of 1/2 V and awaits another input from the SAR. When the ring counter advances by 1, the third MSB is set to 1 and the converter output rises by the next increment of 1/2 V + 1/8 V. The measurement cycle thus proceeds through a series of successive approximations. Finally, when the ring counter reaches its final count, the measurement cycle stops and the digital output of the control register represents the final approximation of the unknown input voltage.

Example *Suppose the converter can measure a maximum of 5 V, i.e. 5 V corresponds to the maximum count of 11111111. If the test voltage V_{in} = 1 V the following steps will take place in the measurement. (Refer to Table 5.1 and Fig. 5.11.)*

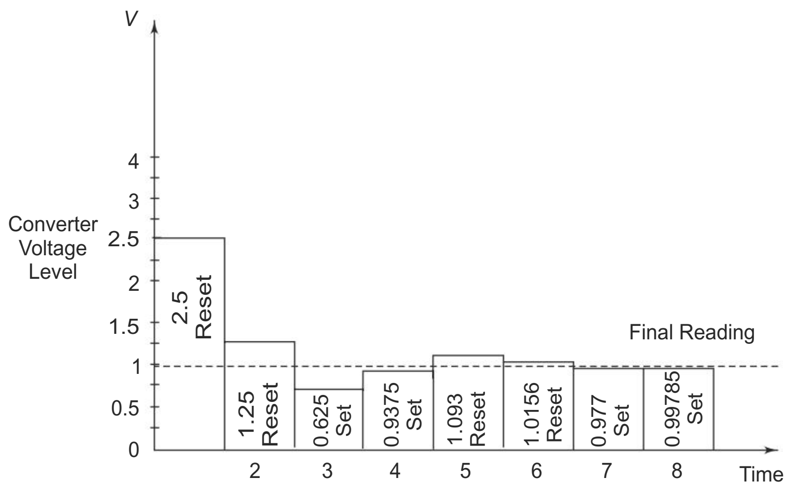


Fig. 5.11 Various output levels for each bit (8-Bit shows the voltage level very nearly equal to 1 V)

Therefore, V_{in} nearly equals V_{out}, i.e. V_{in} = 1 V and V_{out} = 0.99785. The main advantage of this method is speed. At best it takes *n* clock pulses to produce an *n* bit result. Even if the set, test, set or reset operation takes more than 1 clock pulse, the SAR method is still considerably faster than the counter method. However the control circuit is more complex in design and cost is enhanced. This digital voltmeter is capable of 1000 readings per second.

With input voltages greater than dc, the input level changes during digitisation and decisions made during conversion are not consistent. To avoid this error, a sample and hold circuit is used and placed in the input directly following the input attenuator and amplifier.

In its simplest form, the sample and hold (S/H) circuit can be represented by a switch and a capacitor, as shown in Fig. 5.12.

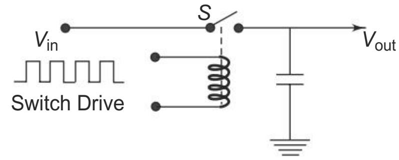


Fig. 5.12 Simple sample hold circuit

In the Sample mode, the switch is closed and the capacitor charges to the instantaneous value of the input voltage.

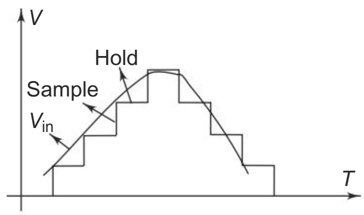


Fig. 5.13 Output waveform of a sample and hold circuit

In the Hold mode, the switch is opened and the capacitor holds the voltage that it had at the instant the switch was opened. If the switch drive is synchronized with the ring counter pulse, the actual measurement and conversion takes place when the S/H circuit is in the Hold mode. The output waveform of a sample and hold circuit is shown in Fig. 5.13.

An actual sample and hold circuit is shown in Fig. 5.14. The sample pulse operates switches 1 and 3. The hold pulse operates switches 2 and 4. The sample-hold pulses are complementary.

In the sample mode the hold capacitor is charged up by the Opamp. In the hold mode, the capacitor is switched into the feedback loop, while input resistors R_1 and R_f are switched to ground. Opamps are used to increase the available driving current into the capacitor or to isolate the capacitor from an external load on the output.

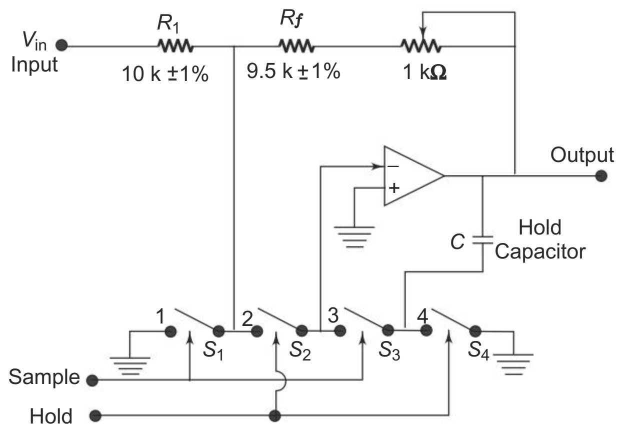


Fig. 5.14 Practical sample and hold circuit

The S/H circuit is basically an Opamp that charges the capacitor during the Sample mode and retains the charge during the Hold mode.

CONTINUOUS BALANCE DVM OR SERVO BALANCING POTENTIOMETER TYPE DVM 5.7

The basic block diagram of a servo balancing potentiometer type DVM is shown in Fig. 5.15.

The input voltage is applied to one side of a mechanical chopper comparator, the other side being connected to the variable arm of a precision potentiometer. The output of the chopper comparator, which is driven by the line voltage at the line frequency rate, is a square wave signal whose amplitude is a function of the difference in voltages connected to the opposite side of the chopper. The square wave signal is amplified and fed to a power amplifier, and the amplified square wave difference signal drives the arm of the potentiometer in the direction needed to make the difference voltage zero. The servo-motor also drives a mechanical readout, which is an indication of the magnitude of the input voltage.

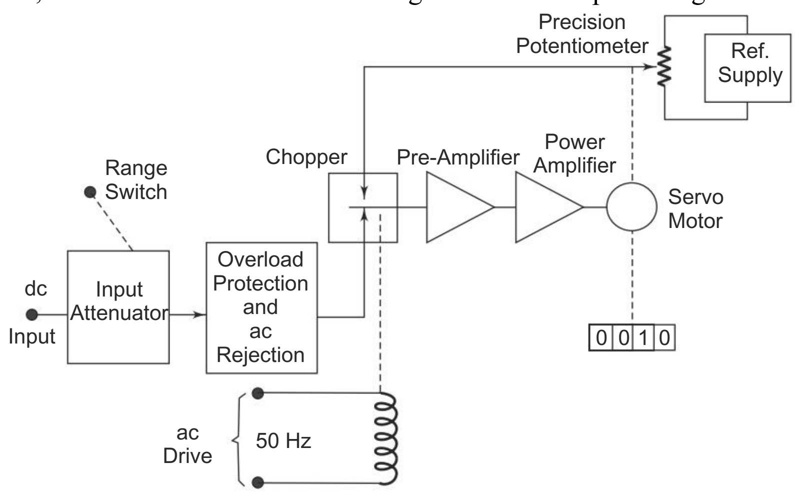


Fig. 5.15 Block diagram of a servo balancing potentiometer type DVM

This DVM uses the principle of balancing, instead of sampling, because of mechanical movement. The average reading time is 2 s.

3½-DIGIT 5.8

The number of digit positions used in a digital meter determines the resolution. Hence a 3 digit display on a DVM for a 0 –1 V range will indicate values from 0 – 999 mV with a smallest increment of 1 mV.

Normally, a fourth digit capable of indicating 0 or 1 (hence called a Half Digit) is placed to the left. This permits the digital meter to read values above 999 up to 1999, to give overlap between ranges for convenience, a process called over-ranging. This type of display is called a 3½ digit display, shown in Fig. 5.16.

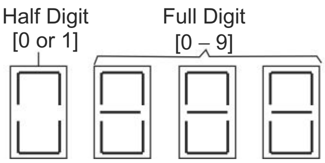


Fig. 5.16 3½-Digit display

RESOLUTION AND SENSITIVITY OF DIGITAL METERS 5.9

Resolution If n = number of full digits, then resolution (R) is $1/10^n$.

The resolution of a DVM is determined by the number of full or active digits used,

If $n = 3$, $$R = \frac{1}{10^n} = \frac{1}{10^3} = 0.001 \text{ or } 0.1\,\%$$

Sensitivity of Digital Meters Sensitivity is the smallest change in input which a digital meter is able to detect. Hence, it is the full scale value of the lowest voltage range multiplied by the meter's resolution.

$$\text{Sensitivity } S = (fs)_{\min} \times R$$

where $(f_S)_{\min}$ = lowest full scale of the meter
R = resolution expressed as decimal

Example 5.5 *What is the resolution of a 3½ digit display on 1 V and 10 V ranges?*

Solution Number of full digits is 3. Therefore, resolution is $1/10^n$ where $n = 3$. Resolution $R = 1/10^3 = 1/1000 = 0.001$

Hence the meter cannot distinguish between values that differ from each other by less than 0.001 of full scale.

For full scale range reading of 1 V, the resolution is $1 \times 0.001 = 0.001$ V.

For full scale reading of 10 V range, the resolution is $10 \text{ V} \times 0.001 = 0.01$ V.

Hence on 10 V scale, the meter cannot distinguish between readings that differ by less than 0.01 V.

Example 5.6 *A 4½ digit voltmeter is used for voltage measurements.*

(i) Find its resolution

(ii) How would 12.98 V be displayed on a 10 V range?

(iii) How would 0.6973 be displayed on 1 V and 10 V ranges.

Solution

(i) Resolution = $1/10^n = 1/10^4 = 0.0001$
where the number of full digits is $n = 4$

(ii) There are 5 digit places in 4½ digits, therefore 12.98 would be displayed as 12.980.
Resolution on 1 V range is $1 \text{ V} \times 0.0001 = 0.0001$
Any reading up to the 4th decimal can be displayed.
Hence 0.6973 will be displayed as 0.6973.

(iii) Resolution on 10 V range = $10 \text{ V} \times 0.0001 = 0.001$ V
Hence decimals up to the 3rd decimal place can be displayed.
Therefore on a 10 V range, the reading will be 0.697 instead of 0.6973.

GENERAL SPECIFICATIONS OF A DVM 5.10

Display	: 3-1/2 digits, LCD
Unit Annunciation	: mV, V, mA, Ω, kΩ, MΩ, buzzer, B(low battery)
	: MANU (Manual), ac and —▷|— (diode test)
Max. Indication	: 1999 or – 1999
Over-range indication	: only (1) or (– 1) displayed at the MSB position.
Polarity	: AUTO negative polarity indication.
Zero adjustment	: Automatic
Functions	: DC volts, AC volts, DC amps, AC amps, Ohms, continuity test, diode test.
Ranging	: Selectable automatic or manual
Automatic	: Instrument automatically selects maximum range for measurement and display. Auto ranging operates on all functions except for dc or ac current.
Manual	: Switch selection as desired
Sampling Rate	: 2 sample/s, nominal
Low Battery	: B mark on LCD readout
Temperature	: Operating 0°C – 40°C, < 80% RH (Relative humidity)
	: Storage – 20°C – 60°C, < 70% RH
Power	: Two AA size 1.5 V batteries. Life 2000 hours typically with zinc-carbon.
Standard accessories	: Probe red-black, safety fuse 250 – 0.2 A
Size	: 160 (L) × 80 (B) × 30 (H)
Weight	: 250 g without batteries.
Input impedance	: 11 MΩ – 1000 MΩ
Accuracy	: ± 0.5% – 0.7% or ± 5 digit for dc
	: 1.0% reading or ± 5 digit for ac at 40 – 500 kHz

MICROPROCESSOR-BASED RAMP TYPE DVM 5.11

A basic block diagram of a microprocessor-based Ramp type DVM and its operating waveform is shown in Fig. 5.17 (a) and (b) respectively. Depending on the command fed to the control input of the multiplexer by the microprocessor, input 1 of the comparator can be consecutively connected to the input 1, 2 or 3 of the multiplexer.

The multiplexer has three inputs -- input 1 is connected to ground potential, input 2 is the unknown input, and input 3 is the reference voltage input. The comparator has two inputs -- input 1 accepts the output signal from the multiplexer, and input 2 accepts the ramp voltage from the ramp generator.

The microprocessor remains suspended in the resting state until it receives a command to start conversion. During the resting period, it regularly sends reset signals to the ramp generator. Each time the ramp generator is reset, its capacitor discharges. It produces a ramp, i.e. a sawtooth voltage whose duration, T_r and amplitude, V_m remain constant. The time duration between the consecutive pulses is sufficiently large enough for the capacitor to get discharged.

Whenever a conversion command arrives at the microprocessor at a time t_1, the multiplexer first connects input 1 of the comparator to its input 1 (i.e. ground potential) and brings the former to ground potential.

The microprocessor pauses until another sawtooth pulse begins. When input 2 voltage, arriving from the ramp generator becomes equal to input 1 of the comparator, the comparator sends a signal to the microprocessor, that ramp voltage is zero. The microprocessor measures this time interval Δt_1 (shown in Fig. 5.17 (b)), by counting the number of clock pulses supplied by the clock generator during this time interval. Let the count during this time be N_1, which is then stored by the microprocessor. A command from the microprocessor now causes the comparator input 1 to be connected to input 2 of the multiplexer. This connects the unknown voltage, V_x to the input 1 of the comparator. At an instant, when the ramp voltage equals the unknown voltage, the comparator sends a signal to the microprocessor that measure the time interval Δt_2 (Fig. 5.17 (b)). The count N_2, during this time interval is also stored.

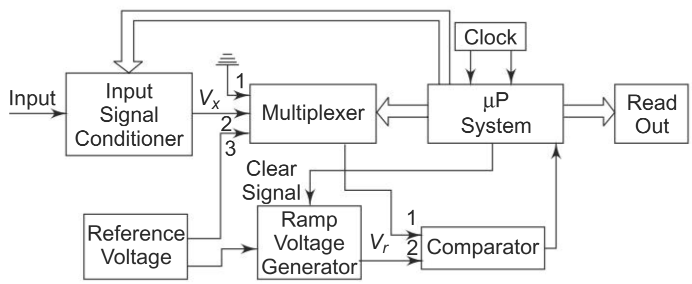


Fig. 5.17 (a) Basic block diagram of a microprocessor-based ramp type DVM

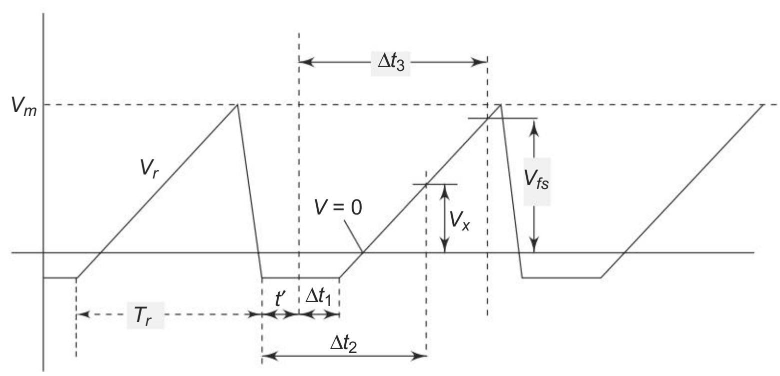


Fig. 5.17 (b) Operating waveform of a rp-based ramp type DVM

Now, the next command from the microprocessor causes the comparator input 1 to be connected to the input 3 of the multiplexer, which is the reference voltage (full scale voltage). The value of the reference voltage sets the upper limit of measurement, that is, full scale value. At the instant, when the ramp voltage equals the reference voltage, a pulse is sent to the microprocessor from the comparator output to measure this time interval, Δt_3 (Fig. 5.17 (b)). The count, N_3 during this time interval is also stored.

The microprocessor then computes the unknown voltage V_x by the equation

$$V_x = C \cdot \frac{(N_2 - N_1)}{(N_3 - N_1)}$$

where C is the coefficient dependent on the characteristics of the instrument and the units selected to express the result.

In this method of measurement, the zero drift has practically no effect on the result, because of the variation of slope of the ramp.

Hence from the Fig. 5.17 (b),

$$\frac{(\Delta t_2 - \Delta t_1)}{(\Delta t_3 - \Delta t_1)} = \frac{V_x}{V_{fs}}$$

$$\therefore \qquad V_x = V_{fs} \cdot \frac{(\Delta t_2 - \Delta t_1)}{(\Delta t_3 - \Delta t_1)}$$

Since the clock pulse repetition frequency f_c and full scale voltage V_{fs} are maintained at a very high level of stability and clock pulses allowed to fall within all the time intervals come from a common source, the above equation may be rewritten as

$$V_x = C \cdot \frac{(N_2 - N_1)}{(N_3 - N_1)}$$

where N_1, N_2, N_3 are the counts representing respectively, the zero drift, the unknown voltage, and the full scale voltage.

Advantages

1. Its scale size remains constant due to zero drift correction and maximum count.
2. The accuracy of the instrument is not affected by the time and temperature instabilities of the circuit element values.
3. There is a good repeatability in switching instants in the presence of noise and interference. This is because the ramp approaches the point at which the comparator operates always the same side and always the same rate.

Disadvantages

Noise and interference cannot be suppressed.

Review Questions

1. State the advantages of a DVM over an analog meter.
2. How are DVMs classified?
3. State the operating and performance characteristics of a digital voltmeter.
4. State distinguished features of digital instruments as compared to analog instruments.
5. Explain the operating principle of a ramp type DVM.
6. Describe with a diagram the operation of ramp type DVM. State limitations of a ramp type DVM and how it is overcome.
7. Explain with the help of diagram the working principle of dual slope type DVM.
8. Describe with a diagram the operation of a dual slope type DVM.
9. State the advantages of a dual slope DVM over a ramp type DVM.
10. Explain the operating principle of Voltage to frequency type DVM.
11. Describe with the help of a block diagram the operation of integrating type digital voltmeter.
12. State the principle of a staircase ramp type DVM.
13. Describe with a diagram the operation of a staircase ramp type DVM. State the advantages of a staircase ramp type DVM.
14. Explain with a diagram, the basic principle of a successive approximation type DVM.
15. Describe with a diagram the operation of a SAR type DVM.
16. State features of a SAR type DVM over other DVMs. What are the advantages of a SAR type DVM over other DVMs?
17. State the principle of operation of a continuous balance DVM.
18. Describe with a diagram the operation of a servo balancing potentiometer type DVM.
19. Define sensitivity of a digital meter. State significance of a half digit.
20. Define the term overrange and half digit. Define resolution of digital meter.
21. What is a sample hold circuit? State its significance. What is the advantage of using a sample hold circuit?
22. State some of the general features of a DVM.
23. State the basic principle of a microprocessor based DVM.
24. Describe with a block diagram the operation of a microprocessor based DVM. State the advantages of a microprocessor based DVM.

Multiple Choice Questions

1. Measurement in a ramp type DVM is performed during the
   (a) negative slope
   (b) positive slope
   (c) both of the slope
   (d) none of the above
2. Measurement by dual slope DVM is performed during
   (a) rising slope
   (b) falling slope
   (c) rising and falling slope
   (d) none of the slope
3. The principle of voltage to time conversion is used in
   (a) dual slope type DVM
   (b) successive approximation type DVM
   (c) integrating type DVM
   (d) none of the above
4. SAR type DVM uses the principle of
   (a) voltage to time conversion
   (b) voltage to frequency conversion
   (c) voltage to binary conversion
   (d) voltage to current conversion

5. Dual slope operates on the principle of
   (a) voltage to time conversion
   (b) voltage to frequency conversion
   (c) frequency to voltage conversion
   (d) voltage to current conversion

Practice Problems

1. The lowest range on a 4-1/2 digit multimeter is 10 mV full scale. Determine the sensitivity of the meter.
2. The lowest range on a 3-1/2 digit multimeter is 10 mV full scale. Determine the sensitivity of the meter.
3. A 3-1/2 digit DVM is used for measuring voltage .Determine the resolution. How would a voltage of 14.42 be displayed on 10 V range and 100 V range.
4. A 3-1/2 digit DVM has an accuracy of ±0.5% of reading ±1 digit. Determine the possible error in V, when the instrument is reading 5 V in 10 V range
5. A 3-1/2 digit voltmeter is used for voltage measurement.
   (i) Find its resolution.
   (ii) How would 11.52 V be displayed on the 10 V range?
   (iii) How would .5234 V be displayed on 1 V and 10 V ranges?
6. Determine the binary equivalent of 2.567 V for a SAR type DVM having 10 bits output and a reference voltage of +5 V.
7. Determine the binary equivalent of 8.735 V for a SAR type DVM having 10 bits output and a reference voltage of +12 V.

Further Reading

1. A.J. Bouwens, *Digital Instrumentation*, McGraw-Hill, 1986.
2. K.J. Dean, *Digital Instruments*, Chapman and Hall, 1965.
3. E.O. Doebelin, *Measurements Systems: Applications and Design*, 4th Edition, McGraw-Hill, 1990.

Chapter

Digital Instruments 6

INTRODUCTION 6.1

Digital instruments are rapidly replacing their analog counterparts. The parameters of interest in a laboratory environment are (i) voltage (ii) current (iii) power (iv) frequency, and (v) logic.

We shall consider digital systems which measure the above parameters. To enable digital systems to recognise information, inputs which are analog in nature must be converted to digital form. Hence any digital instrument would invariably consist of an analog to digital converter in its input stage. The basic building block of a digital instrument is shown in Fig. 6.1.

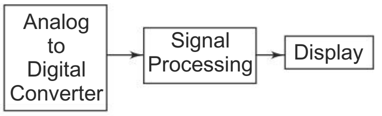


Fig. 6.1 Building block of a digital instrument

The display block may be analog or digital in nature. If an analog readout is desired, it becomes necessary to include a stage involving digital to analog conversion.

Digital systems may consist of the following components.

1. Resistors
2. Capacitors
3. Transistors
4. Linear ICs
5. Digital ICs
6. Display devices
7. Analog to digital converters
8. Digital to analog converters

The digital form of measurement can be used to display the measured quantity numerically instead of a deflection, as in conventional analog meters. Data in digital form facilitates various operations that are normally required in signal processing. An increase in the availability and type of computer facilities and a decrease in the cost of various modules required for digital systems is accelerating the development of digital instrumentation for measurement and signal processing.

DIGITAL MULTIMETERS 6.2

Analog meters require no power supply, they give a better visual indication of changes and suffer less from electric noise and isolation problems. These meters are simple and inexpensive.

Digital meters, on the other hand, offer high accuracy, have a high input impedance and are smaller in size. They gives an unambigious reading at greater viewing distances. The output available is electrical (for interfacing with external equipment), in addition to a visual readout.

The three major classes of digital meters are panel meters, bench type meters and system meters.

All digital meters employ some kind of analog to digital (A/D) converters (often dual slope integrating type) and have a visible readout display at the converter output.

Panel meters are usually placed at one location (and perhaps even a fixed range), while bench meters and system meters are often multimeters, i.e. they can read ac and dc voltage currents and resistances over several ranges.

The basic circuit shown in Fig. 6.2 (a) is always a dc voltmeter. Current is converted to voltage by passing it through a precision low shunt resistance while alternating current is converted into dc by employing rectifiers and filters. For resistance measurement, the meter includes a precision low current source that is applied across the unknown resistance; again this gives a dc voltage which is digitised and readout as ohms.

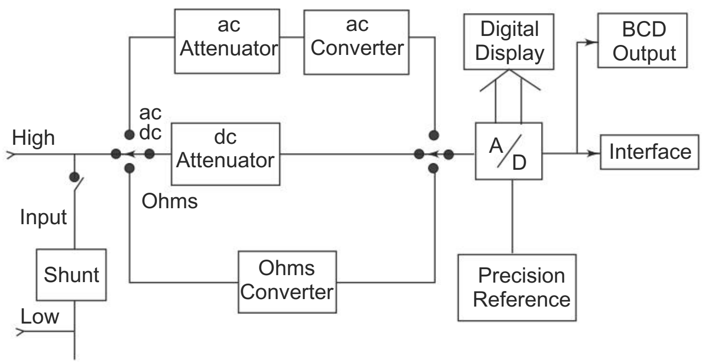


Fig. 6.2 (a) Digital multimeter

Bench meters are intended mainly for stand alone operation and visual operation reading, while system meters provide at least an electrical binary coded decimal output (in parallel with the usual display), and perhaps sophisticated interconnection and control capabilities, or even microprocessor based computing power.

A basic digital multimeter (DMM) is made up of several A/D converters, circuitry for counting and an attenuation circuit. A basic block diagram of a DMM is shown in Fig. 6.2 (b). The current to voltage converter shown in the block diagram of Fig. 6.2 (b) can be implemented with the circuit shown in Fig. 6.2 (c).

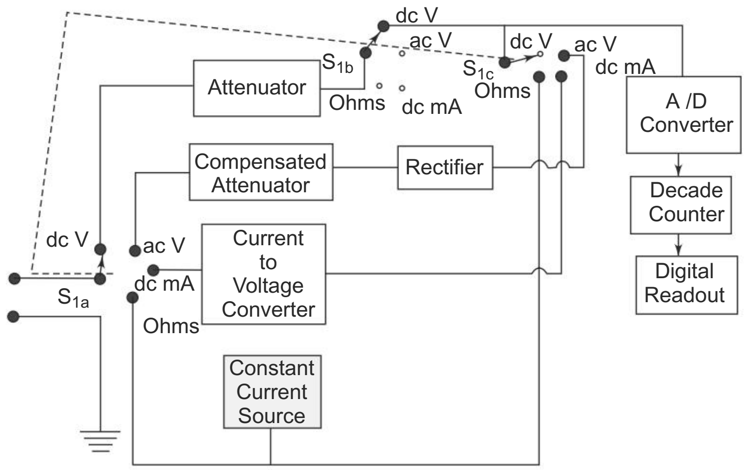


Fig. 6.2 (b) Block diagram of a basic digital multimeter

The current to be measured is applied to the summing junction ($\Sigma\ i$) at the input of the opamp. Since the current at the input of the amplifier is close to zero because of the very high input impedance of the amplifier, the current I_R is very nearly equal to I_i, the current I_R causes a voltage drop which is proportional to the current, to be developed across the resistors. This voltage drop is the input to the A/D converter, thereby providing a reading that is proportional to the unknown current.

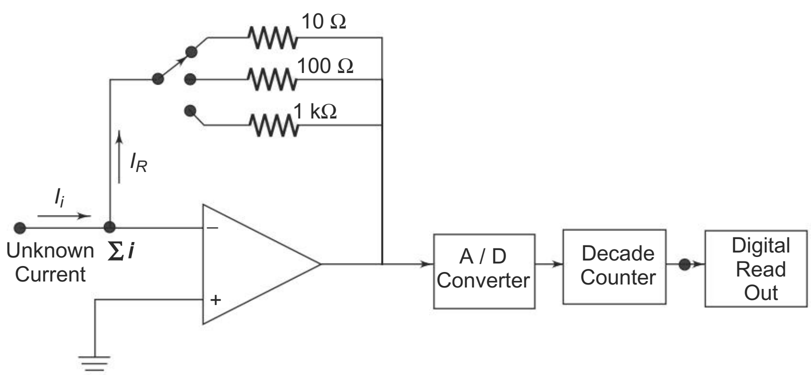


Fig. 6.2 (c) Current to voltage converter

Resistance is measured by passing a known current, from a constant current source, through an unknown resistance. The voltage drop across the resistor is applied to the A/D converter, thereby producing an indication of the value of the unknown resistance.

6.2.1 Digital Panel Meters (DPM)

Digital panel meters are available in a very wide variety of special purpose functions. They have a readout range from the basic 3 digit (999 counts, accuracy of ± 0.1% of reading, ± 1 count) to high precision 4¾ digit ones (± 39,999 counts, accuracy ± 0.005% of reading ± 1 count). Units are available to accept inputs such as dc voltage (from microvolts range to ± 20 volts) ac voltage (for true rms measurement), line voltage, strain gauge bridges (meter provides bridge excitation), RTDs (meter provides sensor excitation), thermocouples of many types (meter provides cold junction compensation and linearisation) and frequency inputs, such as pulse tachometers.

Figure 6.3 shows some details of a high precision unit with an input resistance of $10^9\ \Omega$, ± 0.00250% resolution (10 μV), and ± 0.005% of reading ± 1 count accuracy, which uses a dual slope A/D conversion with automatic zero. The sampling rate is 2.5 per second when it is free running and a maximum of 10 per second when it is externally triggered.

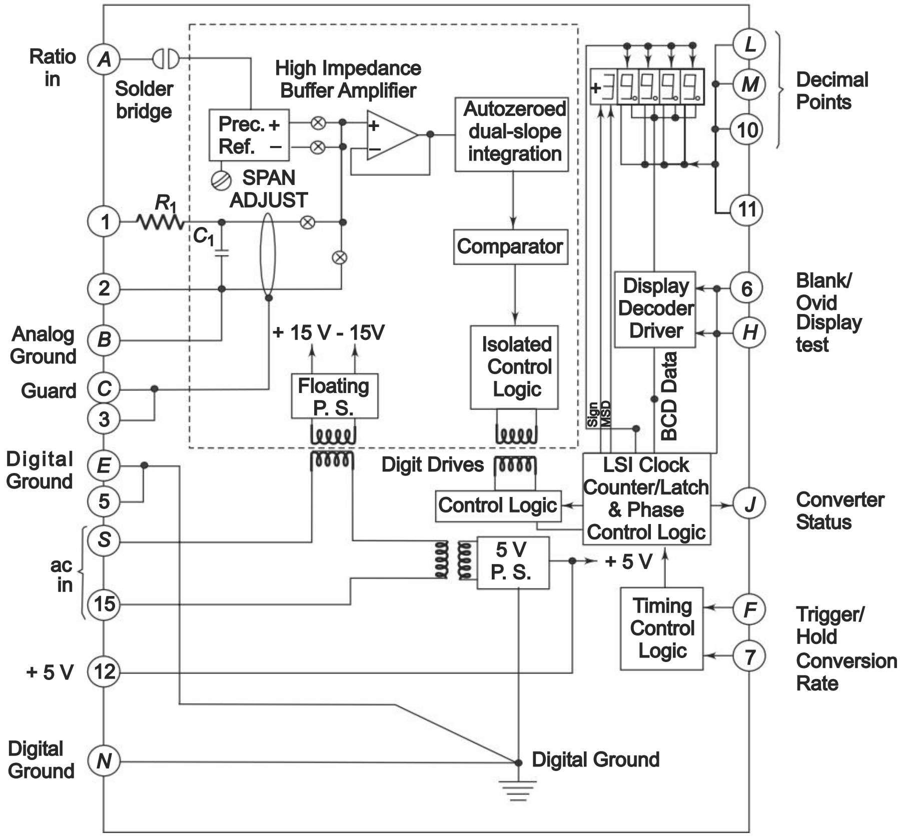


Fig. 6.3 High precision digital panel meter

These meters can be obtained with a tri-state binary coded decimal output. Tri-state outputs provide a high impedance (disconnected) state, in addition to the usual digital high and low. This facilitates interconnection with the micro-computer data buses, since any number of devices can be serviced by a single bus, one at a time, disconnecting all except the two that are communicating with each other.

6.2.2 Bench Type Meters

Bench type meters range from inexpensive hand held units with a 3½ digit readout and 0.5% accuracy, to 5½ digit (200,000 count) devices with 1 μV resolution. Digital nanovoltmeters are designed to measure extremely low voltages and they provide resolution down to about 10 nV (comparable analog meters go to about 1 nV).

Digital picoammeters measure very small currents and can resolve about 1 pA (analog instruments go to 3 femto (10^{-15}) amperes). When extremely high input impedance is required for current, voltage, resistance or charge measurements, an electrometer type of instrument is employed.

Digital electrometer can resolve 10^{-17} A, 10 μV and 1 femto charge and measure resistance as high as 200 T Ω (Tera = 10^{12}). The input impedance can be as high as 10,000 T Ω.

6.2.3 System Type Meters

System type DVMs or DMMs are designed to provide the basic A/D conversion function in data systems assembled by interfacing various peripheral devices with DVM capabilities and their cost vary widely.

A microprocessor is used to provide several mathematical functions in addition to managing the meter operations. A modified dual slope A/D converter is used with selectable integration times, ranging from 0.01 to 100 power lines cycle. At maximum speed (330 readings per second) accuracy is ± 0.1%, while 0.57 readings per second gives a 6½ digit resolution and 0.001% accuracy. Ac and dc voltages and resistance modes are available. The mathematical functions include the following.

1. Null
2. First reading is subtracted from each successive reading and the difference is displayed. (The first reading can be manually entered from the key-board.)
3. The function STAT accumulates reading and calculates mean and variance, (STAT-Statistics).
4. With dBm (R), the user enters the resistance and then all readings are displayed as power dissipated in R in decibel units (referred to 1 mV).
5. With THMS°F (voltmeter in ohms range), the temperature of a thermistor probe is displayed in degrees Fahrenheit or Centigrade (THMS - Temperature of Thermistor)

6. The function $(X - Z)/Y$ provides offsetting and scaling with user entered Z and Y constants (where X is the reading).
7. The function $100 \times (X - Y)/Y$ determines the percentage deviation, and $20 \log X/Y$ displays X in decibels relative to the value of Y. An internal memory (RAM) can be used to store the results of measurements and programs for taking the measurements.

DIGITAL FREQUENCY METER 6.3

Principle of Operation The signal waveform is converted to trigger pulses and applied continuously to an AND gate, as shown in Fig. 6.4. A pulse of 1 s is applied to the other terminal, and the number of pulses counted during this period indicates the frequency.

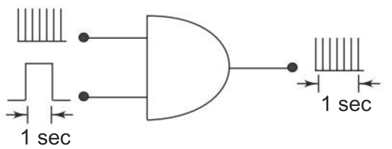


Fig. 6.4 Principle of digital frequency measurement

The signal whose frequency is to be measured is converted into a train of pulses, one pulse for each cycle of the signal. The number of pulses occurring in a definite interval of time is then counted by an electronic counter. Since each pulse represents the cycle of the unknown signal, the number of counts is a direct indication of the frequency of the signal (unknown). Since electronic counters have a high speed of operation, high frequency signals can be measured.

6.3.1 Basic Circuit of a Digital Frequency Meter

The block diagram of a basic circuit of a digital frequency meter is shown in Fig. 6.5.

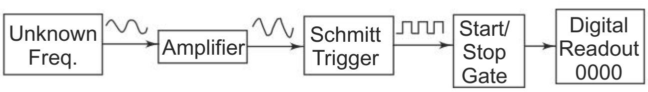


Fig. 6.5 Basic circuit of a digital frequency meter

The signal may be amplified before being applied to the Schmitt trigger. The Schmitt trigger converts the input signal into a square wave with fast rise and fall times, which is then differentiated and clipped. As a result, the output from the Schmitt trigger is a train of pulses, one pulse for each cycle of the signal.

The output pulses from the Schmitt trigger are fed to a START/STOP gate. When this gate is enabled, the input pulses pass through this gate and are fed directly to the electronic counter, which counts the number of pulses.

When this gate is disabled, the counter stops counting the incoming pulses. The counter displays the number of pulses that have passed through it in the time interval between start and stop. If this interval is known, the unknown frequency can be measured.

6.3.2 Basic Circuit for Frequency Measurement

The basic circuit for frequency measurement is as shown in Fig. 6.6. The output of the unknown frequency is applied to a Schmitt trigger, producing positive pulses at the output. These pulses are called the counter signals and are present at point *A* of the main gate. Positive pulses from the time base selector are present at point *B* of the START gate and at point *B* of the STOP gate.

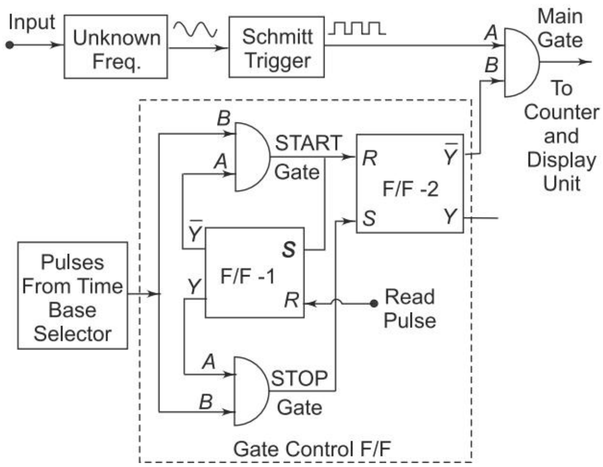


Fig. 6.6 Basic circuit for measurement of frequency showing gate control F/F

Initially the Flip-Flop (F/F–1) is at its logic 1 state. The resulting voltage from output *Y* is applied to point *A* of the STOP gate and enables this gate. The logic 0 stage at the output $\bar{Y}$ of the F/F–1 is applied to the input A of the START gate and disables the gate.

As the STOP gate is enabled, the positive pulses from the time base pass through the STOP gate to the Set (S) input of the F/F–2 thereby setting F/F–2 to the 1 state and keeping it there.

The resulting 0 output level from $\bar{Y}$ of F/F–2 is applied to terminal *B* of the main gate. Hence no pulses from the unknown frequency source can pass through the main gate.

In order to start the operation, a positive pulse is applied to (read input) reset input of F/F–1, thereby causing its state to change. Hence $\bar{Y} = 1$, $Y = 0$, and as a result the STOP gate is disabled and the START gate enabled. This same read pulse is simultaneously applied to the reset input of all decade counters, so that they are reset to 0 and the counting can start.

When the next pulse from the time base arrives, it is able to pass through the START gate to reset F/F–2, therefore, the F/F–2 output changes state from 0 to 1, hence $\bar{Y}$ changes from 0 to 1. This resulting positive voltage from $\bar{Y}$ called the gating signal, is applied to input *B* of the main gate thereby enabling the gate.

Now the pulses from the unknown frequency source pass through the main gate to the counter and the counter starts counting. This same pulse from the

START gate is applied to the set input of F/F–1, changing its state from 0 to 1. This disables the START gate and enables the STOP gate. However, till the main gate is enabled, pulses from the unknown frequency continue to pass through the main gate to the counter.

The next pulse from the time base selector passes through the enabled STOP gate to the set input terminal of F/F–2, changing its output back to 1 and $\bar{Y} = 0$. Therefore the main gate is disabled, disconnecting the unknown frequency signal from the counter. The counter counts the number of pulses occurring between two successive pulses from the time base selector. If the time interval between this two successive pulses from the time base selector is 1 second, then the number of pulses counted within this interval is the frequency of the unknown frequency source, in Hertz.

The assembly consisting of two F/Fs and two gates is called a gate control F/F. The block diagram of a digital frequency meter is shown in Fig. 6.7.

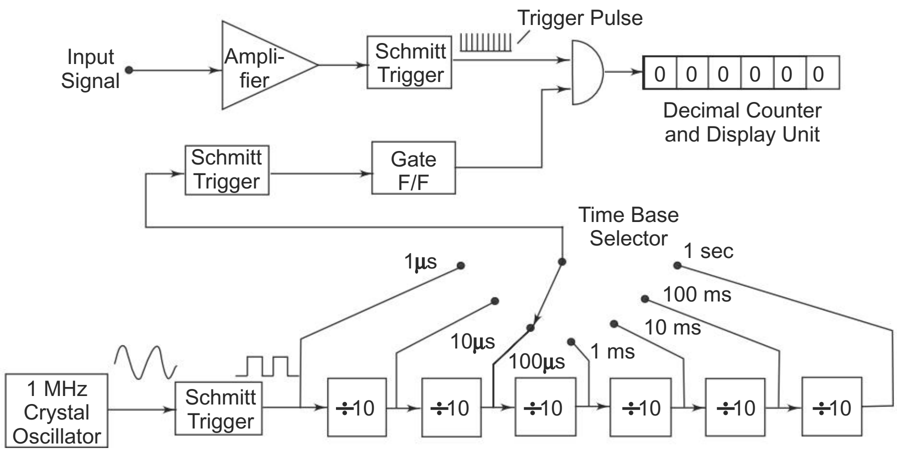


Fig. 6.7 Block diagram of a digital frequency meter

The input signal is amplified and converted to a square wave by a Schmitt trigger circuit. In this diagram, the square wave is differentiated and clipped to produce a train of pulses, each pulse separated by the period of the input signal. The time base selector output is obtained from an oscillator and is similarly converted into positive pulses.

The first pulse activates the gate control F/F. This gate control F/F provides an enable signal to the AND gate. The trigger pulses of the input signal are allowed to pass through the gate for a selected time period and counted. The second pulse from the decade frequency divider changes the state of the control F/F and removes the enable signal from the AND gate, thereby closing it. The decimal counter and display unit output corresponds to the number of input pulses received during a precise time interval; hence the counter display corresponds to the frequency.

6.3.3 High Frequency Measurement (Extending the Frequency Range)

The direct count range of digital frequency meter (DFM) extends from dc to a few 100 MHz. The limitations arises because of the counters used along with the DFM. The counters cannot count at the speed demanded by high frequency measurement.

This range of a few 100 MHz covers only a small portion of the frequency spectrum. Therefore, techniques other than direct counting have been used to extend the range of digital frequency meters to above 40 GHz. The input frequency is reduced before it is applied to a digital counter. This is done by special techniques. Some of the techniques used are as follows.

1. Prescaling The high frequency signal by the use of high speed is divided by the integral numbers such as 2, 4, 6, 8 etc. divider circuits, to get it within the frequency range of DFM (for example synchronous counters).

2. Heterodyne Converter The high frequency signal is reduced in frequency to a range within that of the meter, by using heterodyne techniques.

3. Transfer Oscillator A harmonic or tunable LF continuous wave oscillator is zero beat (mixed to produce zero frequency) with the unknown high frequency signal. The LF oscillator frequency is measured and multiplied by an integer which is equal to the ratio of the two frequencies, in order to determine the value of the unknown HF.

4. Automatic Divider The high frequency signal is reduced by some factor, such as 100:1, using automatically tuned circuits which generates an output frequency equal to 1/100th or 1/1000th of the input frequency.

DIGITAL MEASUREMENT OF TIME 6.4

Principle of Operation The beginning of the time period is the start pulse originating from input 1, and the end of the time period is the stop pulse coming from input 2.

The oscillator runs continuously, but the oscillator pulses reach the output only during the period when the control F/F is in the 1 state. The number of output pulses counted is a measure of the time period.

6.4.1 Time Base Selector

It is clear that in order to know the value of frequency of the input signal, the time interval between the start and stop of the gate must be accurately known. This is called time base.

The time base consist of a fixed frequency crystal oscillator, called a clock oscillator, which has to be very accurate. In order to ensure its accuracy, the crystal is enclosed in a constant temperature oven. The output of this constant frequency oscillator is fed to a Schmitt trigger, which converts the input sine wave to an output consisting of a train of pulses at a rate equal to the frequency of the clock oscillator. The train of pulses then passes through a series of frequency divider decade assemblies connected in cascade. Each decade divider consists of a decade counter and divides the frequency by ten. Outputs are taken from each decade frequency divider by means of a selector switch; any output may be selected.

The circuit of Fig. 6.8 consists of a clock oscillator having a 1 MHz frequency. The output of the Schmitt trigger is 10^6 pulses per second and this point corresponds to a time of 1 microsecond. Hence by using a 6 decade frequency divider, a time base with a range of 1 μs – 10 μs – 100 μs – 1 ms – 10 ms – 100 ms – 1 s can be selected using a selector switch.

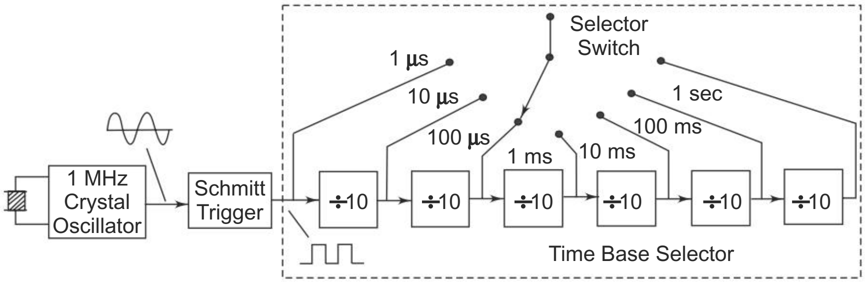


Fig. 6.8 Time base selector

6.4.2 Measurement of Time (Period Measurement)

In some cases it is necessary to measure the time period rather than the frequency. This is especially true in the measurement of frequency in the low frequency range. To obtain good accuracy at low frequency, we should take measurements of the period, rather than make direct frequency measurements. The circuit used for measuring frequency (Fig. 6.7) can be used for the measurement of time period if the counted signal and gating signal are interchanged.

Figure 6.9 shows the circuit for measurement of time period. The gating signal is derived from the unknown input signal, which now controls the enabling and disabling of the main gate. The number of pulses which occur during one period of the unknown signal are counted and displayed by the decade counting assemblies. The only disadvantage is that for measuring the frequency in the low frequency range, the operator has to calculate the frequency from the time by using the equation $f = 1/T$.

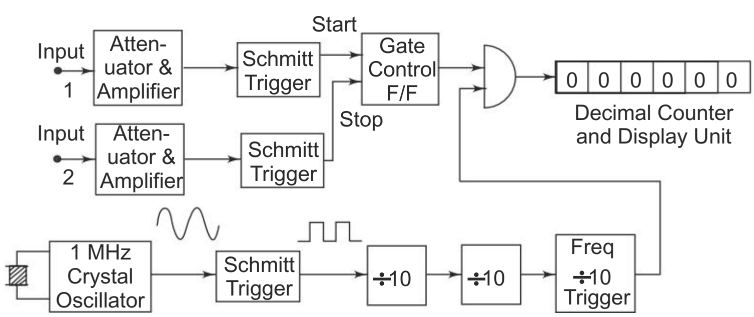


Fig. 6.9 Basic block diagram of time measurement

For example, when measuring the period of a 60 Hz frequency, the electronic counter might display 16.6673 ms, hence the frequency is

$$f = \frac{1}{T} = \frac{1}{16.6673 \times 10^{-3}} = 59.9977 \text{ Hz.}$$

The accuracy of the period measurement and hence of frequency can be greatly increased by using the multiple period average mode of operation. In this mode, the main gate is enabled for more than one period of the unknown signal. This is obtained by passing the unknown signal through one or more decade divider assemblies (DDAs) so that the period is extended by a factor of 10,000 or more.

Hence the digital display shows more digital of information, thus increasing accuracy. However, the decimal point location and measurement units are usually changed each time an additional decade divider is added, so that the display is always in terms of the period of one cycle of the input signal, even though the measurements may have lasted for 10,100 or more cycles.

Figure 6.10 show the multiple average mode of operation. In this circuit, five more decade dividing assemblies are added so that the gate is now enabled for a much longer interval of time than it was with single DDA.

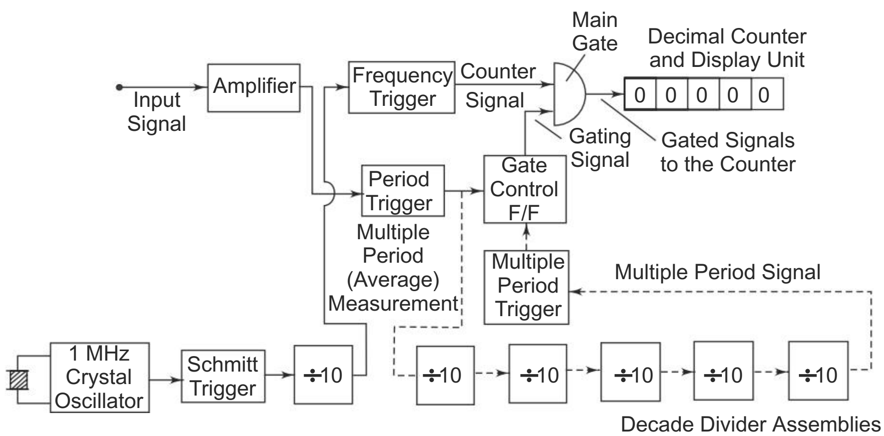


Fig. 6.10 Block diagram of a single and multiple period (average) measurement

6.4.3 Ratio and Multiple Ratio Measurement

The ratio measurement involves the measurement of the ratio of two frequencies. The measurement in effect is a period measurement. A low frequency is used as gating signal while the high frequency is the counted signal. Hence the low frequency takes the place of the time base. The block diagram for the ratio measurements and multiple ratio is shown in Fig. 6.11.

The number of cycles of high frequency signal f_1 which occur during the period of lower frequency signal f_2 are counted and displayed by the decimal counter and display unit. In multiple ratio measurements the period of low frequency signal is extended by a factor of 10,100, etc. by using DDAs.

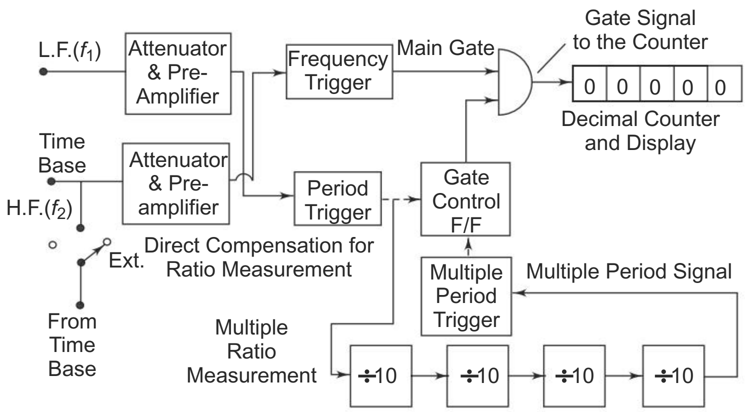


Fig. 6.11 Block diagram for ratio and multiple ratio measurement

UNIVERSAL COUNTER 6.5

All measurements of time period and frequency by various circuits can be assembled together to form one complete block, called a Universal Counter Timer.

The universal counter uses logic gates which are selected and controlled by a single front panel switch, known as the function switch. A simplified block diagram is shown in Fig. 6.12.

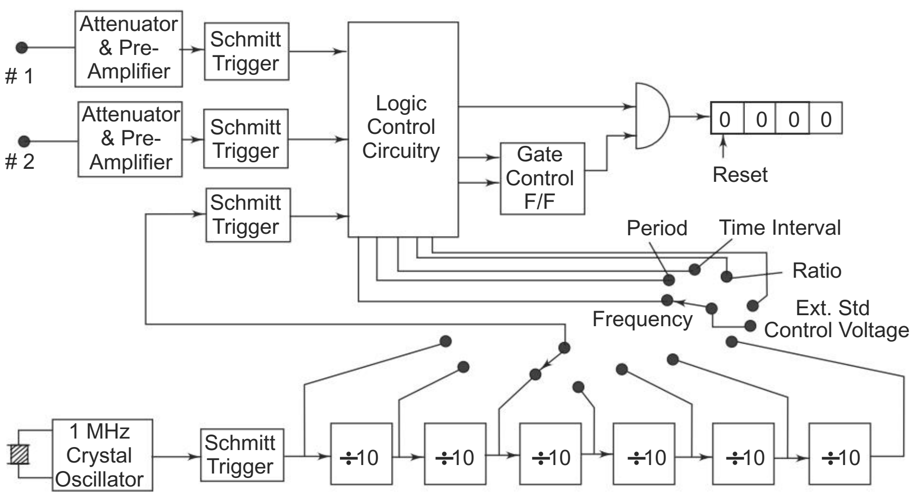


Fig. 6.12 Block diagram of universal counter-timer

With the function switch in the frequency mode, a control voltage is applied to the specific logic gate circuitry. Hence, the input signal is connected to the counted signal channel of the main gate.

The selected output from the time base dividers is simultaneously gated to the control F/F, which enables or disables the main gate. Both control paths are latched internally to allow them to operate only in proper sequence.

When the function switch is on the period mode, the control voltage is connected to proper gates of the logic circuitry, which connects the time base signals to the counted signal channel of the main gate. At the same time the logic circuitry connects the input to the gate control for enabling or disabling the main gate. The other function switches, such as time interval ratio and external standards perform similar functions. The exact details of switching and control procedures vary from instrument to instrument.

DECADE COUNTER 6.6

A decade counter is a circuit of flip-flops (F/Fs) in cascade, which counts in the base 10 (decimal number system). This means that there is a sequence of ten distinct counts in increasing order. Three F/Fs used in cascade progress through 8 distinct states (binary numbers from 000 to 111), while 4 F/Fs in cascade progress through 16 distinct states (binary numbers 0000 to 1111). Hence to get a count of 10, a minimum of 4 F/Fs are required (because 8 distinct states are less while 16 are too many for a decade counter). This problem can be overcome by using 4 F/Fs in cascade and resetting the output of each F/F to 0 after the desired 10 counts. Figure 6.13 (a) shows a decade counter using 4 negative edged triggered F/Fs in cascade.

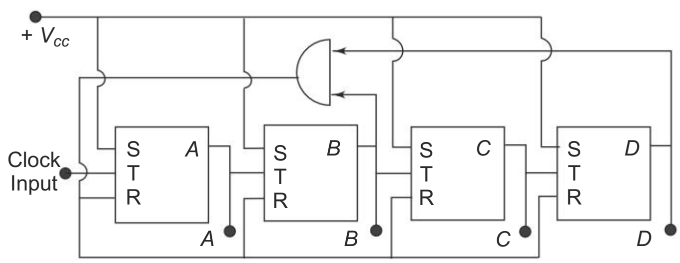


Fig. 6.13 (a) Decade counter

The outputs of the F/F B and D are high (equal to binary 1) after 10 pulses have been applied to the counter. Therefore, the output signal of the decade counter is 1010. This output has to be reset on the very next pulse which is done by the use of an AND gate that resets all F/F's to 0, when the outputs of B and D are 1. The waveform shown in Fig. 6.13 (b) shows the pulse train applied to the trigger input (clock) of the decade counter (shown in Fig. 6.13(a)) and the output waveform of each F/F.

At the beginning all the F/Fs are reset to 0000. The clock pulse is applied to the trigger input T of the F/F. Since this is a negative edged triggered F/F, at the negative edge or falling edge of the trigger input the F/F A will toggle, and hence the output of F/F A changes to level 1; all other F/Fs undergo no change. The outputs from the F/Fs will be 0001. At the next clock pulse the F/F A will toggle back to 0, and the output of F/F A falls from 1 to 0 and is applied to the T input of the next F/F B, toggling it. The output of the F/F B changes to 1 and the output of the decade counter goes to 0010.

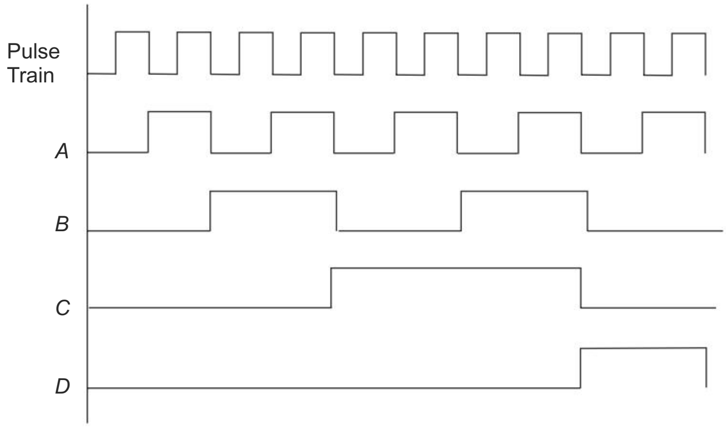


Fig. 6.13 (b) Waveforms of a decade counter

Similarily, as the clock progresses, the F/F toggles in a straight sequence up to 10 (binary 1010). On the very next pulse the AND gate is enabled, which makes the reset input of F/F B high. As soon as the clock pulse arrives, all F/Fs are reset to 0 and the output of the decade counter goes back to 0000. Therefore, by using the AND gate the counter is reset at the tenth pulse.

If some indicator device, such as a lamp or LED has to be driven, the signal must be encoded in the decimal system. This can be done by the use of AND gates. The output of the F/Fs is applied to the AND gates. The AND gates inputs are set to a unique set of conditions that occur only once during the ten trigger pulses. Therefore, each of the 10 lights is ON for only one particular pulse, which allows us to determine at a glance how many pulses have been counted. An F/F divides its input frequency by two. It can be seen from the waveform of Fig. 6.13(b) that the F/F acts as a frequency divider.

Large scale integration (LSI) has made it possible to incorporate the entire decade counter divider circuit with binary to decimal encoding in one or more IC's.

ELECTRONIC COUNTER 6.7

The decade counter can be easily incorporated in a commercial test instrument called an electronic counter. A decade counter, by itself, behaves as a totaliser by totalling the pulses applied to it during the time interval that a gate pulse is present. Typical modes of operation are totalising, frequency, period, ratio, time interval and averaging.

6.7.1 Totalising

In the totalising mode, as shown in Fig. 6.14, the input pulses are counted (totalised) by the decade counter as long as the switch is closed. If the

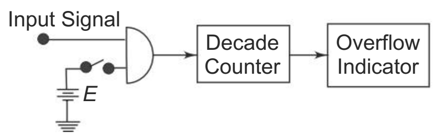


Fig. 6.14 Block diagram of the totalising mode of an electronic counter

count pulse exceeds the capacity of the decade counter, the overflow indicator is activated and the counter starts counting again.

6.7.2 Frequency Mode

If the time interval in which the pulses are being totalised is accurately controlled, the counter operates in the frequency mode. Accurate control of the time interval is achieved by applying a rectangular pulse of known duration to the AND gate, as shown in Fig. 6.15, in place of the dc voltage source. This technique is referred to as gating the counter. A block diagram of an electronic counter operating in the frequency mode is shown in Fig. 6.15. The frequency of the input signal is computed as

$$f = \frac{N}{t}$$

where f = frequency of the input signal
N = pulse counted
t = duration of the gate pulse

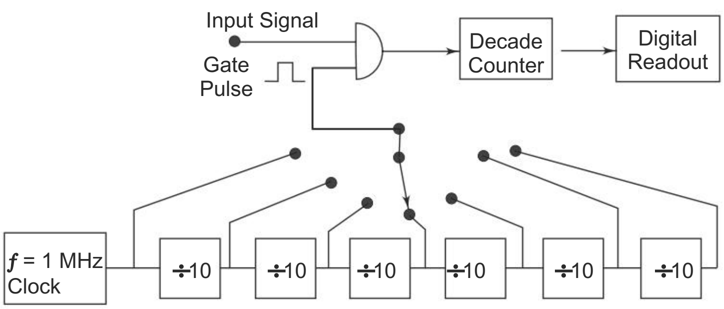


Fig. 6.15 Block diagram of electronic counter frequency mode

6.7.3 Ratio Mode

The ratio mode of operation simply displays the numerical value of the ratio of the frequencies of the two signals.

The low frequency signal is used in place of the clock to provide a gate pulse. The number of cycles of the high frequency signal, which are stored in the decade counter during the presence of an externally generated gate pulse, is read directly as a ratio of the frequency. A basic circuit for the ratio mode of operation is shown in Fig. 6.16.

6.7.4 Period Mode

In some applications, it is desirable to measure the period of the signal rather than its frequency. Since the period is the reciprocal of the frequency, it can easily be measured by

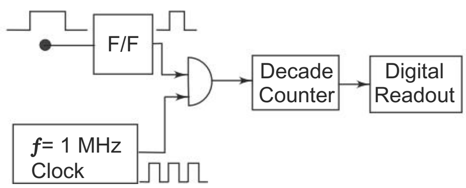


Fig. 6.16 Block diagram of electronic counter in period mode

using the input signal as a gating pulse and counting the clock pulses, as shown in Fig. 6.16.

The period of the input signal is determined from the number of pulses of known frequency or known time duration which are counted by the counter during one cycle of the input signal. The period is computed as

$$T = \frac{N}{f}$$

where N = pulse counted
f = frequency of the clock

6.7.5 Time Interval Mode

The time interval mode of operation measures the time elapsed between two events. The measurement can be done using the circuit of Fig. 6.17.

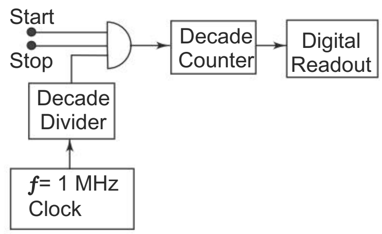


Fig. 6.17 Block diagram of electronic counter in time interval mode

The gate is controlled by two independent inputs, the start input, which enables the gate, and the stop input which disables it. During the time interval between the start and stop signal, clock pulses accumulate in the register, thus providing an indication of the time interval between the start and stop of the event.

Electronic counters find many applications in research and development laboratories, in standard laboratories, on service benches and in everyday operations of many electronics installations.

Counters are used in communication to measure the carrier frequency, in a digital system to measure the clock frequency, and so on.

DIGITAL MEASUREMENT OF FREQUENCY (MAINS) 6.8

The conventional method of measuring the frequency of an electrical signal consists of counting the number of cycles of the input electrical signal during a specified gate interval. The length of the gate interval decides the resolution of the measurement. The shorter the gate interval, the lesser is the resolution. Now, for frequencies of the order of kHz and above, it is possible to get a resolution of 0.1% better with a nominal gate time of 1 (sec). But for low frequencies, in order to obtain a resolution of even 0.5%, the gate time has to be considerably larger. For example, consider the case when the input electrical signal frequency is around 50 Hz. In order to obtain a resolution of 0.1 Hz, the gate interval has to be 10 seconds and in order to obtain a resolution of 0.01 Hz, the gate interval has to be 100 s. These gate periods of 10 s and 100 s are too long and in many cases it is desirable to obtain an indication of the frequency in far less time. Hence, direct or ordinary frequency counters are at a great disadvantage when it comes to low frequency measurements.

For the mains frequency monitor, the frequency range of interest is rather narrow, (50 ± 5%) Hz. The technique employed in the measurement of mains frequency, yields only a parabolic calibration curve. But within the narrow frequency range, which in this case is (50 ± 5%) Hz, the calibration is conveniently flat. Hence, the error due to the non-linear calibration is less than 0.2% at a frequency deviation as large as 5% from the centre frequency, which is 50 Hz. The error is 0.02% at a frequency deviation as large as 2% from the centre frequency and as the frequency approaches the centre value of 50 Hz, the error approaches zero.

The principle that backs up this technique is as follows.

Consider the relationship between the frequency and time period of the input electrical signal, $f = 1/T$. For an ac power line, the desired frequency is 50 Hz; let us denote it as F_{50}. If we denote the corresponding period by T_{50}, then T_{50} = 1/50 s = 20 ms. Now, let us further denote any frequency x as F_x and the corresponding period as T_x. Now, we introduce a new time scale for measuring the period of the electrical input signal whose frequency measurement is desired. Let us formulate the relationship between the MKS scale and the new scale as 20 ms = 50 ku (ku = kilounit), where ku is the unit of the new time scale. Therefore, 1 s = 50 × (1/20) × 1000 = 2500 ku.

Let us determine the periods corresponding to various frequencies from 45 Hz to 55 Hz in terms of ku.

$$T_{45} = 1/45 \text{ s} = 2500/45 \text{ ku} = 55.55 \text{ ku}$$
$$T_{46} = 1/46 \text{ s} = 2500/46 \text{ ku} = 54.35 \text{ ku}$$
$$T_{47} = 1/47 \text{ s} = 2500/47 \text{ ku} = 53.19 \text{ ku}$$
$$T_{48} = 1/48 \text{ s} = 2500/48 \text{ ku} = 52.08 \text{ ku}$$
$$T_{49} = 1/49 \text{ s} = 2500/49 \text{ ku} = 51.02 \text{ ku}$$
$$T_{50} = 1/50 \text{ s} = 2500/50 \text{ ku} = 50.00 \text{ ku}$$
$$T_{51} = 1/51 \text{ s} = 2500/51 \text{ ku} = 49.02 \text{ ku}$$
$$T_{52} = 1/52 \text{ s} = 2500/52 \text{ ku} = 48.08 \text{ ku}$$
$$T_{53} = 1/53 \text{ s} = 2500/53 \text{ ku} = 47.16 \text{ ku}$$
$$T_{54} = 1/54 \text{ s} = 2500/54 \text{ ku} = 46.29 \text{ ku}$$
$$T_{55} = 1/55 \text{ s} = 2500/55 \text{ ku} = 44.44 \text{ ku}$$

Now within this narrow frequency range of 45 – 55 Hz, we can form an empirical relation between the frequency and period of signals, as $F_x = 100 - T_x$, where F_x is the frequency in Hz and T_x is the period in ku.

In order to determine the frequency of the input electrical signal, the period of the input signal has to be measured in terms of ku and then subtracted from 100. If an indication of the frequency F_x correct to 1 Hz (resolution is 1 Hz), is sufficient, then it is enough if the period T_x of the input signal is measured correct to 1 ku. In order to obtain 4 digit indication, the period T_x of the input signal has to be measured correct to 0.01 ku. To determine the period of the input electrical (power line) signal in terms of 0.01 ku, we need a reference signal whose period is 0.01 ku. This signal is obtained from a stable crystal controlled reference oscillator. The number of cycles of this reference signal during 1 cycle of the input signal gives the period T_x in terms of 0.07 ku. Rather than employing separate circuitry to determine the period and then subtracting it from 100, an UP/DOWN counter is used in the down counted mode. Thus the processes of counting the period and doing the subtraction are done simultaneously by the same circuitry.

The schematic diagram shown in Fig. 6.18, is the circuitry of the digital mains frequency measurement, consisting of the input wave shaper, reference clock generator, sequence control logic unit, counter display and intermediate latch circuitry.

Let us analyse the circuit, assuming the input signal is exactly 50 Hz. When the reference clock frequency is 1 MHz, there will be 20,000 pulses in the 20 ms period. These 20,000 pulses are divided by the other two flip-flops by a factor of 4, to get 5000 pulses in the measuring period of 20 ms. Now these pulses are counted down in the 10000 modulo counter and are displayed.

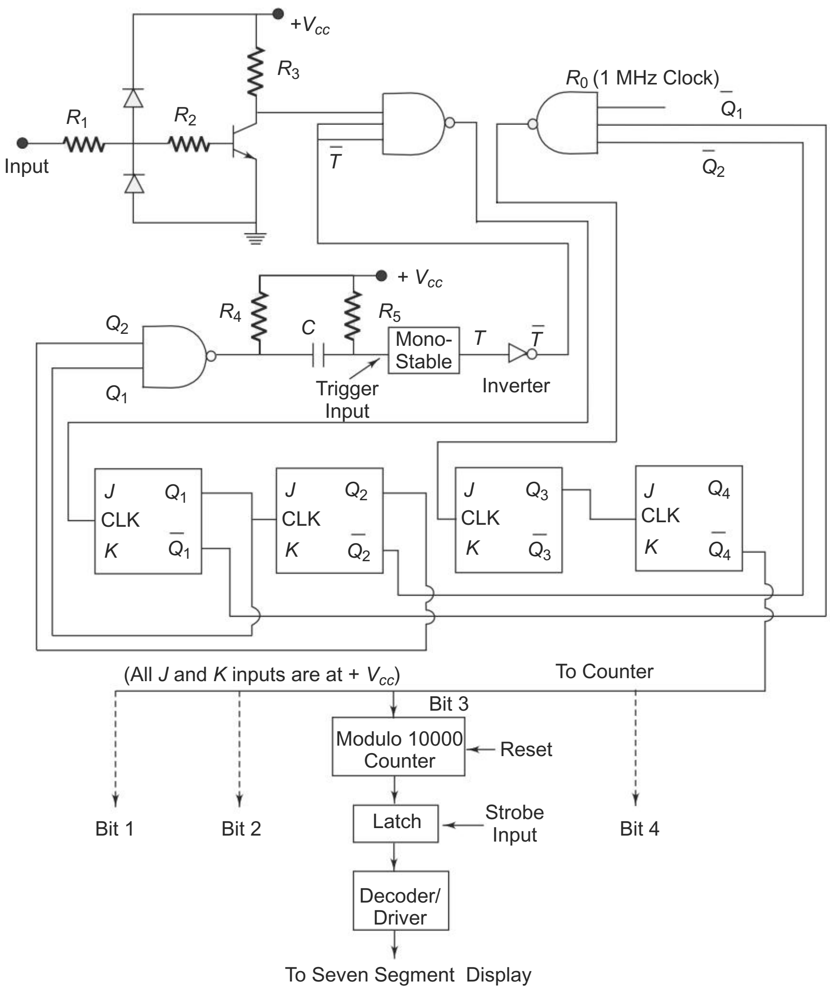


Fig. 6.18 Digital measurement of mains frequency

Let us consider the case, when the input signal frequency is 48 Hz. The period i.e. 1/48 s = 20.83 ms. Within this period, the number of pulses will be 5208. (For 20 ms the count is 5000, hence for 20.83 ms the count is

$$\frac{5000 \times 20.83 \text{ ms}}{20 \text{ ms}} = 5208$$

Therefore, 20.83 ms the count is 5208). Now, these pulses are counted down and the display reading is 10000 – 5208 = 4792; 48 Hz is displayed as 47.92 Hz.

DIGITAL TACHOMETER 6.9

The technique employed in measuring the speed of a rotating shaft is similar to the technique used in a conventional frequency counter, except that the selection of the gate period is in accordance with the rpm calibration.

Let us assume, that the rpm of a rotating shaft is R. Let P be the number of pulses produced by the pick up for one revolution of the shaft. Therefore, in one minute the number of pulses from the pick up is $R \times P$. Then, the frequency of the signal from the pick up is $(R \times P)/60$. Now, if the gate period is G s the pulses counted are $(R \times P \times G)/60$. In order to get the direct reading in rpm, the number of pulses to be counted by the counter is R. So we select the gate period as $60/P$, and the counter counts

$$\frac{(R \times P \times 60)}{60 \times P} = R \text{ pulses}$$

and we can read the rpm of the rotating shaft directly. So, the relation between the gate period and the number of pulses produced by the pickup is $G = 60/P$. If we fix the gate period as one second (G = 1 s), then the revolution pickup must be capable of producing 60 pulses per revolution.

Figure 6.19 shows a schematic diagram of a digital tachometer.

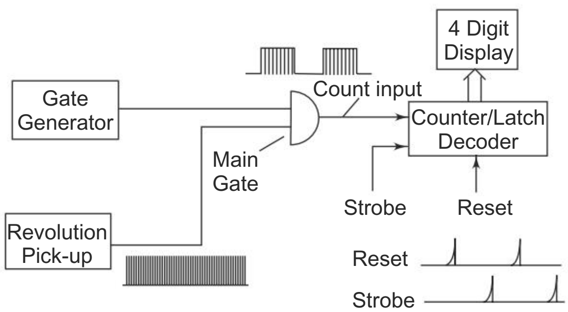


Fig. 6.19 Basic block diagram of a digital tachometer

DIGITAL pH METER 6.10

The measurement of hydrogen ion activity (pH) in a solution can be accomplished with the help of a pH meter. For those unfamiliar with the terminology, a very brief review is included.

pH is a quantative measure of acidity. If the pH is less than 7, the solution is acidic (the lower the pH, the greater the acidity). A neutral solution has a pH of 7 and alkaline (basic) solutions have a pH greater than 7.

The pH unit is defined as

$$\text{pH} = -\log\,(\text{concentration of } H^+)$$

where H^+ is the hydrogen or hydronium ion. (Analog pH meters are discussed in Chapter 10.)

A digital pH meter differs from an ordinary pH meter, in this the meter is replaced by an analog to digital converter (ADC) and a digital display. A frequently used ADC for this application is the dual slope converter. A basic block diagram of a digital pH meter is shown in Fig. 6.20.

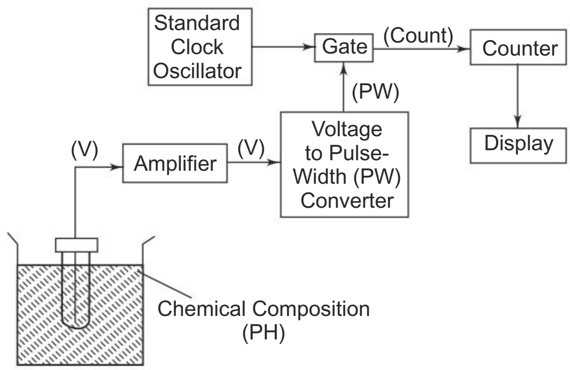


Fig. 6.20 Digital pH meter

The dual slope circuit produces a pulse which has a duration proportional to the input signal voltage, that is, a *T* pulse width signal. The pulse width is converted to a digital signal using the pulse to turn an oscillator On or Off, generating a count digital signal. The count signal is in turn counted or converted to a parallel digital signal for display by the counter.

AUTOMATION IN DIGITAL INSTRUMENTS 6.11

One of the advantages of digital multimeters is their ease of operation. The reading is easy to take and does not lend itself to errors of interpretation. Moreover, the number of ranges is limited because the ranges move in steps of 10 (instead of the $\sqrt{10}$ steps used for analog instruments). Demand from users for simple forms of computation signalling and control, and advances in digital circuitry have led to further development, in which more and more automatic functions have been incorporated in digital voltmeters. Nearly all instruments today have automatic polarity display and automatic decimal point positioning, while many have automatic ranging and zeroing too.

This automation includes automatic polarity indication, automatic ranging, and automatic zeroing.

1. Automatic Polarity Indication The polarity indication is generally obtained from the information in the ADC. For integrating ADCs, only the polarity of the integrated signal is of importance. The polarity should thus be measured at the very end of the integration period (see Fig. 6.21). As the length of the integration period is determined by counting a number of clock pulses, it is logical to use the last count or some of the last counts to start the polarity measurement. The output of the integrator is then used to set the polarity flip-flop, the output of which is stored in memory until the next measurement is made.

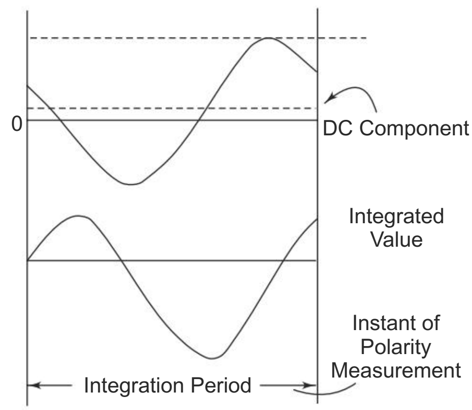


Fig. 6.21 Polarity of integrated signal has to be measured at very end of integration period

2. Automatic Ranging The object of automatic ranging is to get a reading with optimum resolution under all circumstances (e.g. 170 mV should be displayed as 170.0 and not as 0.170). Let us take the example of a 3½ digit display, i.e. one with a maximum reading of 1999. This maximum means that any higher value must be reduced by a factor of 10 before it can be displayed (e.g. 201 mV as 0201). On the other hand, any value below 0200 can be displayed with one decade more resolution (e.g. 195 mV as 195.0). In other words, if the display does not reach a value of 0200, the instrument should automatically be switched to a more sensitive range, and if a value of higher than 1999 is offered, the next less sensitive range must be selected.

Generally the lower limit is taken lower than 0200 (example 0180). Otherwise, a voltage exhibiting slight fluctuations around 2000 would be displayed successively as 1999.9, 0200 and 0201, which would be confusing. By introducing an overlap in the ranges (see Fig. 6.22), we ensure that all values are displayed in the same range (in the above example, as 0199, 0200 or 0201). Values around 0180 also give a stable display e.g. 1798, 1800 and 1807).

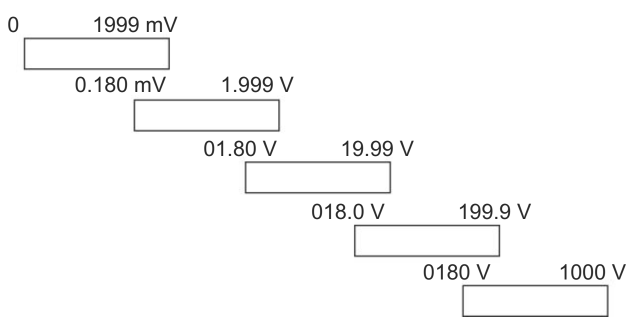


Fig. 6.22 Example of overlapping ranges in automatic ranging instruments

The design of an automatic ranging system is indicated in the block diagram in Fig. 6.23.

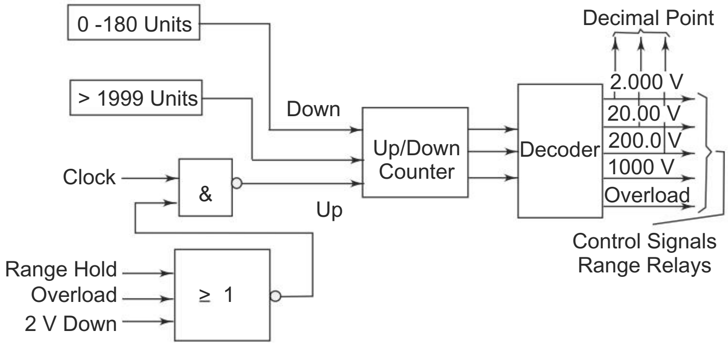


Fig. 6.23 Block diagram of automatic ranging system

The information contained in the counter of the ADC yields a control pulse for down ranging when the count is less than 180 and one for up ranging when the count exceeds 1999 units. The Up/Down counter of the automatic ranging circuit reacts to this information at the moment that a clock pulse (a pulse at the end of the measuring period, also used to transfer new data to memory), is applied, and the new information is used to set the range relays via the decoder. At the same time the decimal point in the display is adapted to the new range, when more than range step has to be made, several measuring periods are needed to reach the final result. Clock pulses, and so automatic ranging, can be inhibited, for example, by a manual range hold command, by a signal that exceeds the maximum range (only for up counts), and course by reaching the most sensitive range, but then only for down counts.

3. Automatic Zeroing Each user of a voltmeter expects the instrument to indicate zero when the input is short-circuited. In a digital voltmeter with a maximum reading of 1999, a zero error of 0.05% of full scale deflection is sufficient to give a reading of 0001. For this reason, and in the interests of optimum accuracy with low valued readings, a zero adjustment is necessary. To increase the ease of operation, many instruments now contain an automatic zeroing circuit.

In a system used in several multimeters, the zero error is measured just before the real measurement and stored as an analog signal. A simplified circuit diagram of a circuit that can be used for this purpose is given in Fig. 6.24, for a dual slope ADC.

Before the real measurement is made, switches S_3, S_4 and S_5 are closed, say for 50 ms, thus grounding the input, giving the integrator a short RC time, and connecting the output of the comparator to capacitor C. This capacitor is now charged by the offset voltages to the amplifier, the integrator and the comparator. When switches S_3, S_4 and S_5 are opened again to start the real measurement, the total offset voltage of the circuit (equal to zero error) is stored in this capacitor, and the real input voltage is measured correctly.

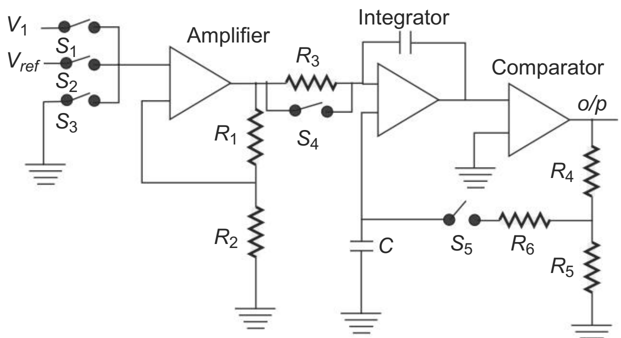


Fig. 6.24 Simplified circuit diagram of automatic zeroing circuit that can be used with dual slope ADC

6.11.1 Fully Automatic Digital Instrument

A multimeter with automatic polarity indication, automatic zero correction and automatic ranging (of course coupled with automatic decimal point indication) only needs a signal applied to its input, and a command as to what quantity (V_{dc}, V_{ac}, I or R) to measure; it does all the rest itself.

The digital part of a typical instrument is organised so as to produce a display or a digital output signal, as shown in Fig. 6.25. Before a measurement can begin, the functions of the instrument must be set, that is, we must select the quantity to be measured (e.g. voltage), the ranging mode (automatic or manual), and the start mode (internal or with an external trigger signal). This can be done by the front panel controls, or via a remote control input. In both cases, the signals are fed to the function control unit, while the information on ranging is passed to the range control unit.

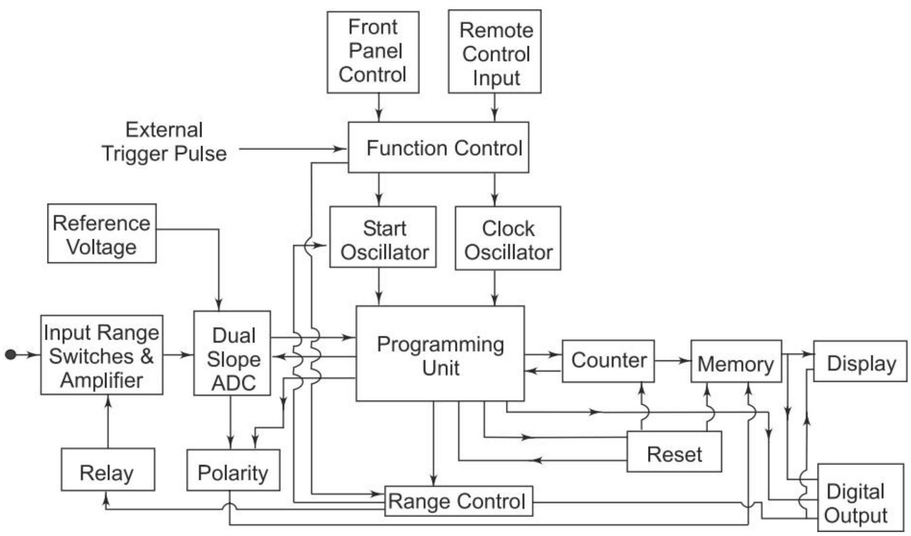


Fig. 6.25 Block diagram of an automatic instrument

Let us assume that in the instrument in question, the ADC is of the dual slope integration type, and that a choice can be made between the combinations given in Table 6.1.

Table 6.1

Measurements	*Integration Time*	*Clock Frequency*
4	100 ms	200 kHz
20	20 ms	1 MHz
200	2 ms	1 MHz

It will be clear that as the number of measurements per second increases, the integration time must be reduced and that it is useful to increase the clock frequency at the same time to maintain good resolution. To select the desired combination, information on the number of measurements per second must be fed to the start oscillator and clock oscillator. The latter constantly supplies clock pulses to the programming unit. The former is free running when the DVM is set for internal start, but it waits for an external trigger signal when the DVM is set for external control. Let us now follow (see Fig. 6.25) the various steps involved in the performance of a measurement, for the case that the instrument is set for automatic ranging and external triggering.

An incoming trigger pulse causes the start oscillator to deliver a pulse to the programming unit, and a measurement is started. The programming unit starts both the counter and the ADC. The ADC is connected to the input. The counter counts the clock pulses to determine the integration time and sends two signals back to the programming unit, one just before the end of the integration period, and the other at the end of this period. The first signal is used by the programming unit to activate the polarity detector, which determines the polarity of the integrated signal, while the second serves to switch the ADC input from the reference signal. At the same time, the counter is reset to zero and starts counting the down integration time of the ADC until it is stopped by the zero-detector signal of the ADC. At that moment, the programming unit compares the counter reading with the automatic ranging limits, and passes an up or down signal to the range control, if necessary. This unit switches the input range switches via a relay and triggers the start oscillator for a new measurement in a more sensitive or less sensitive range. In the meantime, the programming unit will also have reset the counter. This process continues until a measurement which is within the automatic ranging limits has been made. The programming unit then transfers the new data from the counter to the memory, together with the polarity information so as to make them available to the display unit and the digital output. Finally, the programming unit delivers a transfer pulse to the digital output, to warn an instrument connected to this output (e.g. a printer) that new data has been made available.

DIGITAL PHASE METER 6.12

The simplest technique to measure the phase difference between two signals employs two flip-flops. The signals to be fed must be of the same frequency. First, the signals must be shaped to a square waveform without any change in their phase positions, by the use of a zero crossing detector. The process of measuring the phase difference can be illustrated by the schematic diagram shown in Fig. 6.26.

The block diagram consists of two pairs of preamplifier's, zero crossing detectors, J-K F/Fs, and a single control gate. Two signals having phases P_o and P_x respectively are applied as inputs to the preamplifier and attenuation circuit. The frequency of the two inputs is the same but their phases are different.

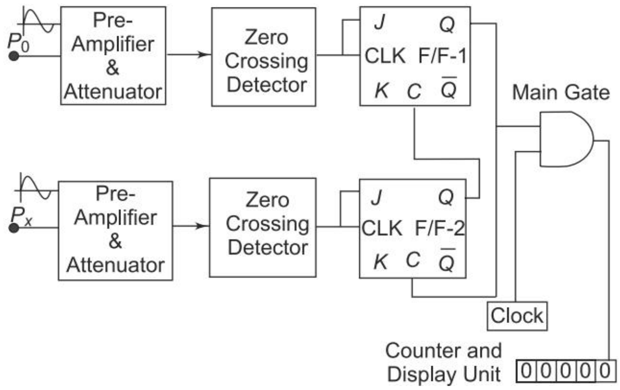


Fig. 6.26 Digital phase meter

As the P_o input signal increases in the positive half cycle, the zero crossing detector changes its state when the input crosses zero (0) giving a high (1) level at the output. This causes the J–K F/F–1 to be set (1), that is, the output (Q) of F/F–1 goes high. This high output from the F/F–1 enables the AND gate, and pulses from the clock are fed directly to the counter. The counter starts counting these pulses. Also this high output level of F/F–1 is applied to the clear input of F/F–2 which makes the output of the F/F–2 go to zero (0).

Now as the input P_x which has a phase difference with respect to P_o, crosses zero (0) in the positive half cycle, the zero detector is activated, causing its output to go high (1). This high input in turn toggles the J–K F/F–2, making its output go high. This output (Q) of F/F–2 is connected to the clear input of F/F–1 forcing the F/F–1 to reset. Hence the output of F/F–1 goes to zero (0). The AND gate is thus disabled, and the counter stops counting.

The number of pulses counted while enabling and disabling the AND gate is in direct proportion to the phase difference, hence the display unit gives a direct readout of the phase difference between the two inputs having the same frequency f.

If the input signal frequency is *f*, then the clock frequency must be 360 times the input frequency for accurate measurements.

DIGITAL CAPACITANCE METER 6.13

Since the capacitance is linearly proportional to the time constant, when a capacitor is charged by a constant current source and discharged through a fixed resistance, we can use a 555 timer along with some digital test equipment to measure capacitances.

One obvious way is to measure the time period of the oscillations. By choosing the right size of charging resistance, we can get a reading directly in microfarads or *nanofarads*. Unlike many capacitance measuring schemes, this one easily handles electrolytics up to the tens of thousands of microfarads.

A better way is to measure only the capacitor discharge time, as shown in Fig. 6.27. With this method, any leakage in the capacitor under test will make the capacitor appear smaller in value than it actually is, and is an effective indicator of how the test capacitor will behave in most timing and bypass circuits.

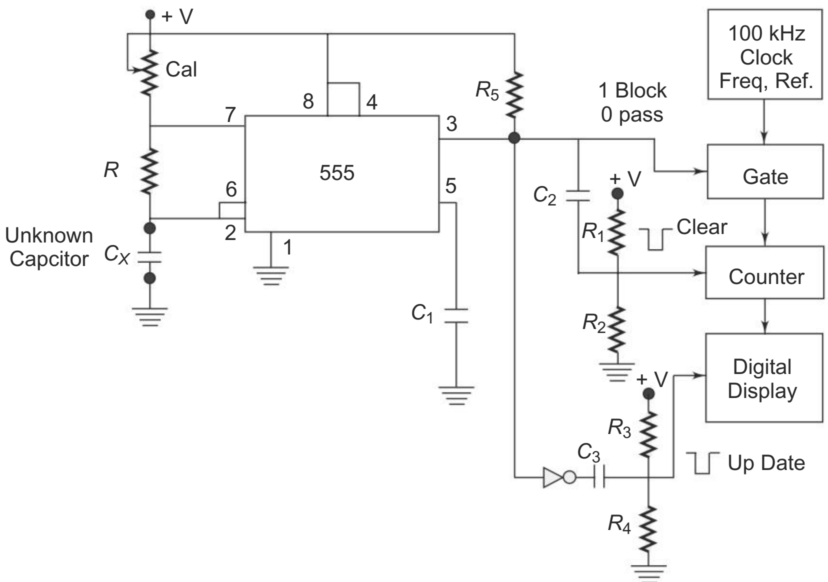


Fig. 6.27 Block diagram of a basic digital capacitance meter

In this circuit, the 555 timer is used as an astable multivibrator. At the peak of the charging curve, a digital counter is reset and a clock of 100 kHz pulses is turned on and routed to the counter. When the discharge portion of the cycle is completed, the display is updated and the value of the capacitor is readout. By selecting the proper reference frequency and charging currents, one can obtain a direct digital display of the value of the capacitance.

Be sure to properly shield the leads and keep them short for low capacity measurements, since the 50 Hz hum can cause some slight instability.

MICROPROCESSOR-BASED INSTRUMENTS 6.14

Digital instruments are designed around digital logic circuits without memory.

The use of microprocessors as an integral part of measuring instruments has given rise to a whole new class of instruments, called intelligent instruments.

Figure 6.28 shows a block diagram of a microprocessor based impedance measuring instrument. The operation makes interface with the instrument via the IEEE 488 bus to allow control by, or to make the measurement available to, a large external computer system. The timing clock signal and the ac test signal are provided by frequency division of the oscillator signal.

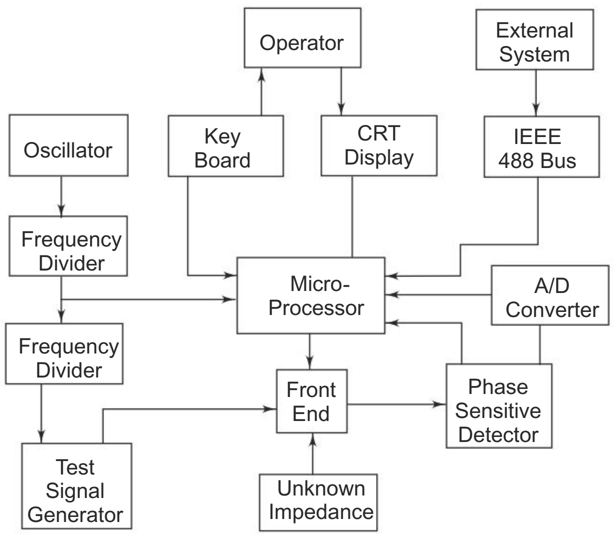


Fig. 6.28 Block diagram of a μp (microprocessor) based instrument

The front end circuit applies the test signal to the unknown impedance and an standard impedance provides an output signal, proportional to the voltage across each, to the phase sensitive detector. Signal transfer is controlled by the microprocessor. The phase sensitive detector, which is also controlled by the microprocessor, converts the ac inputs of the impedance in vector form to a dc output. The A/D converter provides the digital data,which is used by the microprocessor to compute the value of the unknown impedance. This is then displayed on the CRT or sent as output to the IEEE 488 bus.

THE IEEE 488 BUS 6.15

The purpose of IEEE 488 bus is to provide digital interfacing between programmable instruments. There are many instrumentation systems in which interactive instruments, under the command of a central controller, provide superior error-free results when compared with conventional manually operated systems.

Problems such as impedance mismatch, obtaining cables with proper connectors and logic level compatibility are also eliminated by designing the system around a bus-compatible instrument.

The basic structure of an IEEE 488 bus showing interfacing between interactive instruments is given in Fig. 6.29.

Every device in the system must be able to perform at least one of the roles, namely talker, listener or controller. A talker can send data to other devices via the bus. Some devices, such as programmable instruments, can both listen and talk.

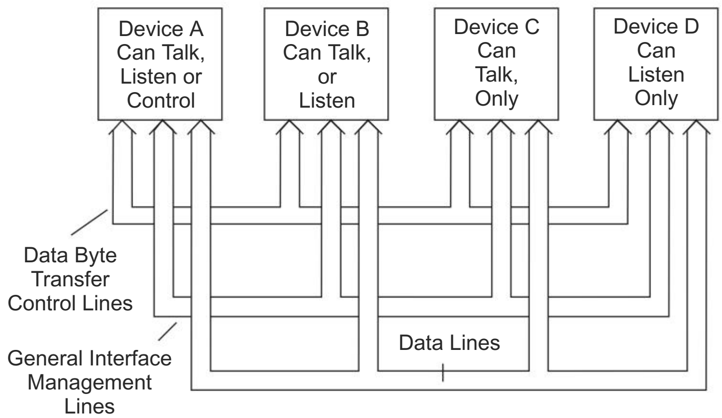


Fig. 6.29 Block Diagram of devices interfaced with IEEE 488 bus

In the listen mode it may receive an instruction to make a particular measurement and in the talk mode it may send its measurand. A controller manages the operation of the bus system. It controls data gathering and transfer by designating which devices talk or listen as well as controlling specific actions within other devices.

Review Questions

1. State the advantages of digital instruments over analog instruments.
2. Explain the operation of a basic digital multimeter.
3. Describe with a diagram the working of a digital multimeter.
4. Explain how current can be measured by digital multimeter. Explain how resistance can be measured by a DMM.
5. Describe the operation of current to voltage converter.
6. Explain with a diagram the working of a digital panel meter.
7. Stae the difference between DMM and DPM.
8. State different types of digital meters.
9. Explain the working principle of a digital frequency meter.

10. Explain with the help of block diagram the operation of a DFM.
11. State the function and explain the operation of a gate control flip-flop.
12. State the function and explain the operation of a time base selector.
13. Explain with a diagram the basic principle of operation of digital time measurement.
14. Describe with the help of a block diagram the operation of digital time measurement.
15. Explain with a diagram the operation of period measurement.
16. What is a universal counter ? How can it be used to measure the following:
    (i) Frequency (ii) Time
    (iii) Period (iv) Ratio
17. Describe with the help of a block diagram the operation of a universal counter-timer.
18. What is an electronic counter ? How can it be used to measure the following:
    (i) Totalising (ii) Frequency
    (iii) Period mode (iv) Ratio mode
    (v) Time interval mode,
19. Explain with a diagram the operation of a digital measurement of mains frequency.
20. On what principle does a digital tachometer operate? Explain with a diagram the working of a digital tachometer.
21. Explain with a diagram the working of a digital pH meter. How is pH measured?
22. On what principle does a digital capacitance meter operate? Describe with a diagram the operation of a digital capacitance meter.
23. How is automation in digital instruments obtained? Why are the ranges in digital instruments overlapped?
24. How can measurements of impedance be obtained using microprocessors?

Multiple Choice Questions

1. A time base selector is used to select
   (a) frequency (b) time
   (c) amplitude (d) voltage
2. A time base selector basically consists of
   (a) LC oscillator
   (b) RC oscillator
   (c) crystal oscillator
   (d) Wien bridge oscillator
3. Schmitt trigger used in digital measurement time converts input to
   (a) square wave
   (b) sine wave
   (c) pulses
   (d) sawtooth wave.
4. A frequency meter is used to measure
   (a) frequency
   (b) ratio
   (c) time interval
   (d) phase
5. Frequency dividers used in the frequency counter divide the frequency by
   (a) 10 (b) 2
   (c) 100 (d) 20

Further Reading

1. Malvino and Leach, *Digital Principles*, McGraw-Hill, New York.
2. Larry D. Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley & Sons, 1987.
3. Malmstadt, Enke and Others, *Instrumentation for Scientist*.
4. E.O. Doebelin, *Measurement Systems: Applications and Design*, 4th Edition, 1990, McGraw-Hill.

Chapter 7

Oscilloscope

INTRODUCTION 7.1

The Cathode Ray Oscilloscope (CRO) is probably the most versatile tool for the development of electronic circuits and systems.

The CRO allows the amplitude of electrical signals, whether they are voltage, current, or power, to be displayed as a function of time.

The CRO depends on the movement of an electron beam, which is bombarded (impinged) on a screen coated with a fluorescent material, to produce a visible spot. If the electron beam is deflected on both the conventional axes, i.e. *X*-axis and *Y*-axis, a two-dimensional display is produced.

The beam is deflected at a constant rate relative of time along the *X*-axis and is deflected along the *Y*-axis in response to an stimulus, such as a voltage. This produces a time-dependent variation of the input voltage.

The oscilloscope is basically an electron beam voltmeter. The heart of the oscilloscope is the Cathode Ray Tube (CRT) which makes the applied signal visible by the deflection of a thin beam of electrons. Since the electron has practically no weight, and hence no inertia, therefore the beam of electrons can be moved to follow waveforms varying at a rate of millions of times/second. Thus, the electron beam faithfully follows rapid variations in signal voltage and traces a visible path on the CRT screen. In this way, rapid variations, pulsations or transients are reproduced and the operator can observe the waveform as well as measure amplitude at any instant of time.

Since it is completely electronic in nature, the oscilloscope can reproduce HF waves which are too fast for electro mechanical devices to follow. Thus, the oscilloscope has simplified many tests and measurements. It can also be used in any field where a parameter can be converted into a proportional voltage for observation, e.g. meteorology, biology, and medicine.

The oscilloscope is thus a kind of voltmeter which uses beam instead of a pointer, and kind of recorder which uses an electron beam instead of a pen.

BASIC PRINCIPLE 7.2

7.2.1 Electron Beam

To understand the principle of an oscilloscope, let us consider a torch which is focussed on a piece of cardboard (held perpendicular to the torch). The light beam will make a bright spot where it strikes the cardboard or screen. Hold the torch still, the spot remains still, move the torch, the spot also moves. If the movement is slow, the eye can follow the movement, buf if it is too fast for the eye to follow, persistence of vision causes the eye to see the pattern traced by the spot. Hence when we wave the torch from side to side, a horizontal line is traced; we can similarly have a vertical line or a circle. Hence, if the torch is moved in any manner at a very rapid rate, light would be traced, just like drawing or writing.

A similar action takes place in the CRT of an oscilloscope. The torch is replaced by an electron gun, the light beam by a narrow electron beam, and the cardboard by the external flat end of a glass tube, which is chemically coated to form a fluorescent screen. Here the electron gun generates the beam which moves down the tube and strikes the screen. The screen glows at the point of collision, producing a bright spot.

When the beam is deflected by means of an electric or magnetic field, the spot moves accordingly and traces out a pattern.

The electron gun assembly consists of the indirectly heated cathode with its heater, the control grid, and the first and second anodes.

The control grid in the CRT is cylindrical, with a small aperture in line with the cathode. The electrons emitted from the cathode emerge from this aperture as a slightly divergent beam. The negative bias voltage applied to the grid, controls the beam current. The intensity (or brightness) of the phosphorescent spot depends on the beam current. Hence this control grid bias knob is called or labelled as intensity.

The diverging beam of electrons is converged and focussed on the screen by two accelerating anodes, which form an electronic lens. Further ahead of the grid cylinder is another narrow cylinder, the first anode. It is kept highly positive with respect to the cathode. The second anode is a wider cylinder following the first. Both the cylinders have narrow apertures in line with the electron beam. The second anode is operated at a still higher positive potential and does most of the accleration of the beam. The combination of the first anode cylinder and the wider second anode cylinder produces an electric field that focuses the electron beam on the screen, as a lens converges a diverging beam of light.

The electronic lens action is controlled by the focus control. If this control is turned to either side of its correct focussing position, the spot on the screen becomes larger and blurred. Bringing it back to its correct position brightens and concentrates the spot. With this proper focus, the small spot can be deflected to produce sharp narrow lines that trace the pattern on the CRT screen.

The electron beam may be deflected transversely by means of an electric field (electrostatic deflection) or a magnetic field (electromagnetic deflection).

Most oscilloscopes use electrostatic deflection, since it permits high frequency operation and requires negligible power. Electromagnetic deflection is most common in TV picture tubes.

Electrons are negatively charged particles, they are attracted by a positive charge or field and repelled by a negative charge. Since the electron beam is a stream of electrons, a positive field will divert it in one direction and a negative field in the opposite direction. To move the beam in this way in the CRT, deflecting plates are mounted inside the tube and suitable deflecting voltages are applied to them.

These plates are arranged in two pairs; H_1 and H_2 for deflecting the beam horizontally, and V_1 and V_2 for deflecting it vertically. Leads are taken out for external connections. The beam passes down the tube between the four plates, as shown in Fig. 7.1.

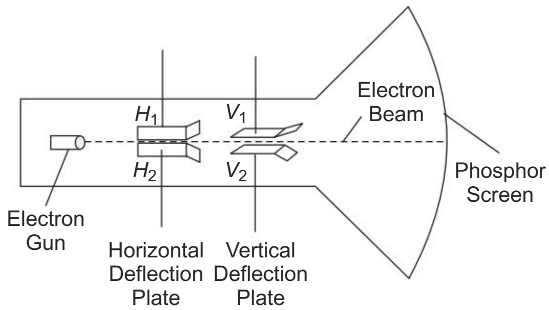


Fig. 7.1 Basic diagram of a CRT

When the plates are at zero voltage the beam is midway between them and the spot is in the centre of the screen. When H_1 is made positive with respect to the cathode (and all other plates are at zero voltage), it attracts the beam and the spot moves horizontally to the left. When H_2 is made positive, it attracts the beam and the spot moves horizontally to the right. Similarly when V_1 is made positive, the spot moves vertically upwards and when V_2 is made positive it moves vertically downwards. In each of these deflections, the displacement of the beam, and therefore, the distance travelled by the spot, is proportional to the voltage applied at the plates.

Figure 7.2 shows the various positions of the electron beam for different voltages applied to the two pairs of plates. If a negative voltage is applied to any plate, the beam will be repelled rather than attracted and the deflection will be in the opposite direction. For example, if V_1 is made negative, the beam will be deflected vertically downward.

As mentioned before, when a spot moves too rapidly for the eye to follow it traces a line. The same happens when a rapidly pulsating or ac voltage is applied to the deflecting plates, the beam is moved back and forth so rapidly that the spot traces a line. When a positive pulsating voltage is applied to H_1 (or negative pulsating to H_2) the spot traces a horizontal line from the centre to the left. Similarly when a positive pulsating voltage is applied to H_2 (or negative to H_1), the spot traces a horizontal line from the centre to the right. Similarly, when the pulsating voltage is applied to V_1, we get a vertical line from the centre upwards and when applied to V_2, we get a vertical line from the centre downwards.

Now, when an alternating voltage is applied to H_1 or H_2, the spot moves from the centre to one side, back to the centre and on to the other side, back again and so on, tracing a line that passes through the centre of the screen (because of the attraction and repulsion of the beam by the positive and negative ac half cycles). Hence, a horizontal line is traced when an ac voltage is applied to either horizontal plates. Similarly, a vertical line is traced when an ac voltage is applied to the vertical plates.

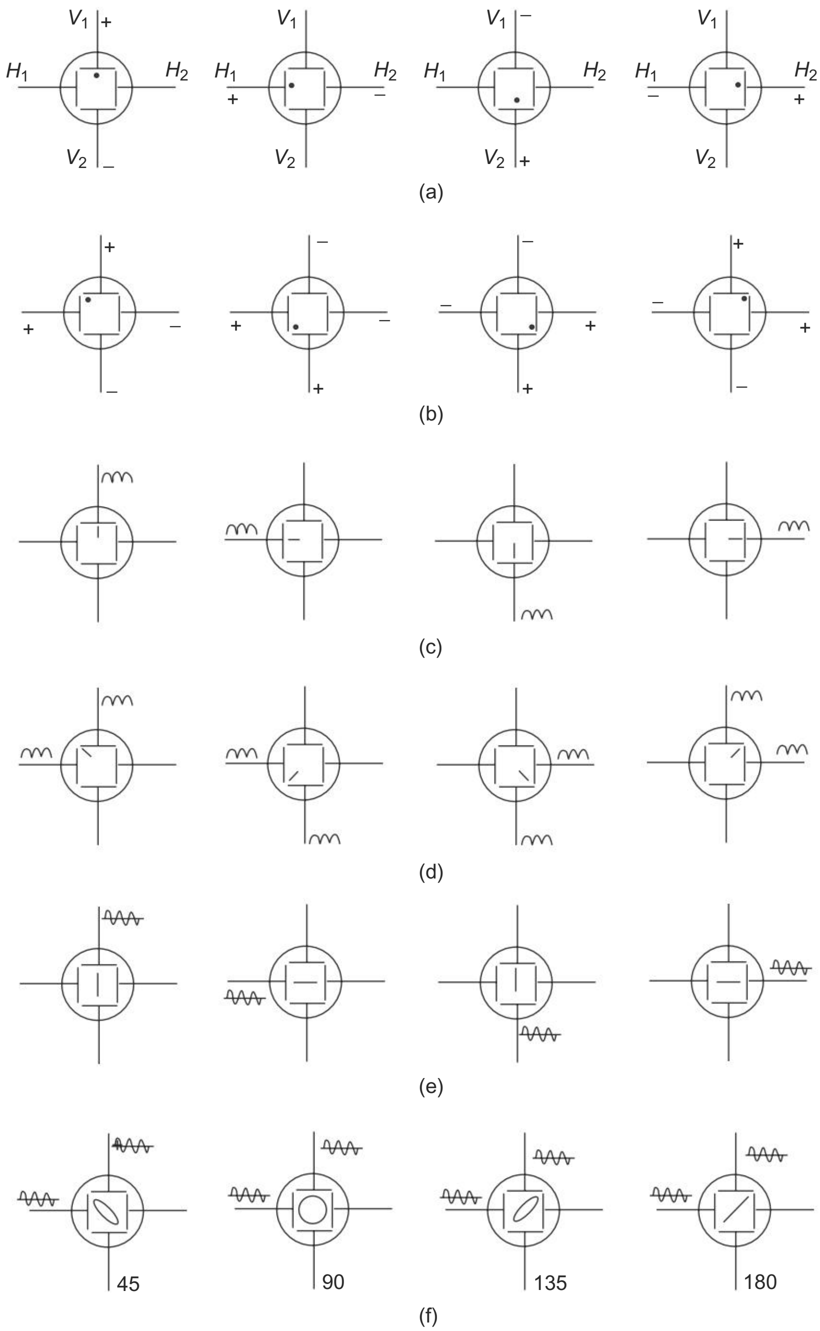


Fig. 7.2 (a) Applying dc voltage to vertical and horizontal plates (b) Applying dc voltage to both vertical and horizontal plates (c) Applying pulsating dc to vertical or horizontal plates (d) Applying pulsating dc to both vertical and horizontal (e) Applying a sine wave to vertical or horizontal (f) Applying phase-shifted sine waves to vertical and horizontal plates

Now, let us see what happens to the beam when voltage is applied simultaneously to both vertical and horizontal plates.

When voltage is applied to the vertical and horizontal plates simultaneously, the deflection of the beam is proportional to the resultant of the two voltages and the position of the beam is inbetween the horizontal and vertical axis of the screen.

Suppose a steady voltage is applied to one horizontal and one vertical plate. When these two deflection voltages are equal, the position of the spot is 45°. The angle is greater than 45° (spot close to V-axis) when the vertical voltage is greater than the horizontal, and less than 45° (spot close to the H-axis) when the horizontal voltage is greater than the vertical voltage. When the two voltages are reversed in polarity, the deflection is in the opposite direction.

If instead of a steady voltage a pulsating positive voltage is applied to the same plates as before, a tilt of 45° is obtained from the horizontal if the two voltages are equal and in phase. The tilt is greater than 45° if the vertical voltage is greater than the horizontal voltage, and less than 45° if the horizontal voltage is greater than the vertical voltage. When a negative voltage is applied to both plates, the trace extends in the opposite direction. When an alternating voltage in phase is applied to the plates, the tilt of the trace is 45° from the horizontal when the two voltages are equal, it traces a straight line at an angle of 45 degrees. Again the tilt is greater than 45° if the vertical voltage is greater, and less than 45° if the horizontal voltage is greater.

The ac trace has equal length from the centre of screen to either tip, when the ac is symmetrical. When it is asymmetrical, the shorter part corresponds to the lower voltage half cycle.

A single trace is obtained only when the phase angles are 0°, 180°, or 360°. At other phase angles a double line trace is obtained at equal voltages, the pattern becomes an ellipse with a right tilt for angle between 0–90°, a circle at 90° and an ellipse with a left tilt between 90–180°. Again, a left tilt between 180–270°, a circle at 270°, an ellipse with lift tilt between 270–360°.

CRT FEATURES 7.3

Electrostatic CRTs are available in a number of types and sizes to suit individual requirements. The important features of these tubes are as follows.

1. Size Size refers to the screen diameter. CRTs for oscilloscopes are available in sizes of 1, 2, 3, 5, and 7 inches. 3 inches is most common for portable instruments.

For example a CRT having a number 5GP1. The first number 5 indicates that it is a 5 inch tube.

Both round and rectangular CRTs are found in scopes today. The vertical viewing size is 8 cm and horizontal is 10 cm.

2. Phosphor The screen is coated with a fluorescent material called phosphor. This material determines the colour and persistence of the trace, both of which are indicated by the phosphor.

The trace colours in electrostatic CRTs for oscilloscopes are blue, green and blue green. White is used in TVs, and blue-white, orange, and yellow are used for radar.

Persistence is expressed as short, medium and long. This refers to the length of time the trace remains on the screen after the signal has ended.

The phosphor of the oscilloscope is designated as follows.

P1 — Green medium
P2 — Blue green medium
P5 — Blue very short
P11 — Blue short

These designations are combined in the tube type number. Hence 5GP1 is a 5 inch tube with a medium persistence green trace.

Medium persistence traces are mostly used for general purpose applications.

Long persistence traces are used for transients, since they keep the fast transient on the screen for observation after the transient has disappeared.

Short persistence is needed for extremely high speed phenomena, to prevent smearing and interference caused when one image persists and overlaps with the next one.

P11 phosphor is considered the best for photographing from the CRT screen.

3. Operating Voltages The CRT requires a heater voltage of 6.3 volts ac or dc at 600 mA.

Several dc voltages are listed below. The voltages vary with the type of tube used.

(i) Negative grid (control) voltage – 14 V to – 200 V.
(ii) Positive anode no. 1 (focusing anode) – 100 V to – 1100 V
(iii) Positive anode no. 2 (accelerating anode) 600 V to 6000 V
(iv) Positive anode no. 3 (accelerating anode) 200 V to 20000 V in some cases

4. Deflection Voltages Either ac or dc voltage will deflect the beam. The distance through which the spot moves on the screen is proportional to the dc, or peak ac amplitude. The deflection sensitivity of the tube is usually stated as the dc voltage (or peak ac voltage) required for each cm of deflection of the spot on the screen.

5. Viewing Screen The viewing screen is the glass face plate, the inside wall of which is coated with phosphor. The viewing screen is a reetangular screen having graticules marked on it. The standard size used nowadays is 8 cm × 10 cm (8 cm on the vertical and 10 cm on horizontal). Each centimeter on the graticule corresponds to one division (div). The standard phosphor colour used nowadays is blue.

7.3.1 Basic Principle of Signal Display (Function of the Sweep Generator)

The amplitude of a voltage may be directly measured on a calibrated viewing screen from the length of the straight line trace it produces. This is entirely satisfactory for dc voltage.

But the straight line tells little, or practically nothing, about the waveform of an ac voltage, pulsating voltage or transient. What is required is a graph of the voltage traced on the screen by the ac spot (a graph of amplitude versus time).

To obtain such a display the signal voltage is applied to the vertical plates (directly or through the vertical amplifier) and it moves the spot vertically to positions, corresponding to the instantaneous values of the signal. Simultaneously, the spot is moved horizontally by a sweep voltage applied to the horizontal plates. The combined action of these two voltages causes the spot to produce a trace on the screen. The horizontal sweep voltage produces the time base by moving the spot horizontally with time, while the signal moves the spot vertically in proportional to the voltage at a particular instant of time.

There are two important sweep generator requirements.

1. The sweep must be linear (the sweep voltage must rise linearly to the maximum value required for full screen horizontal deflection of the spot).
2. The spot must move in one direction only, i.e. from left to right only, else the signal will be traced backwards during the return sweep. This means that the sweep voltage must drop suddenly after reaching its maximum value. These requirements call for a sweep voltage having a linear sawtooth waveform, as shown in Fig. 7.3.

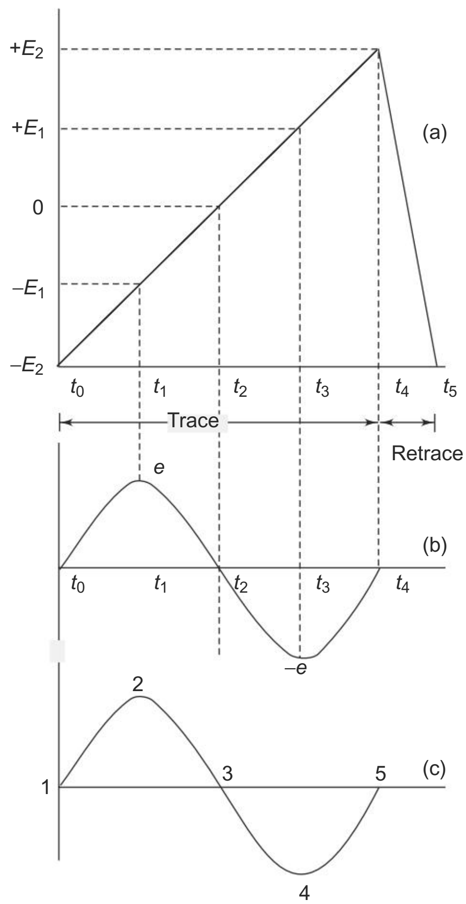


Fig. 7.3 Waveform of sweep voltage

Now at time t_0, the sweep voltage is $-E_2$, and the negative horizontal voltage moves the spot to point 1 on the screen. At this instant, the signal voltage is 0, so the spot rests at the left end of the zero line on the screen.

At time t_1, the linearly increasing sawtooth reaches $-E_1$, which, being more positive than $-E_2$, moves the spot to the screen, point 2. At this instant, the signal voltage is e, the +ve peak value, so the point represents its maximum upward deflection of the spot. At time t_2, the sawtooth voltage is 0, there is no horizontal deflection and the spot is at the centre, point 3. At this instant, the signal voltage is 0, so there is no vertical deflection either. At time t_3, the sawtooth voltage is $+E_1$, moving the spot to point 4.

At this instant, the signal is $-e$, the –ve peak value, so point 4 is the maximum downward deflection of the spot. At time t_4, the sawtooth voltage is $+E_2$, moving the spot to point 5. Now the signal voltage is 0, so the spot is not vertically deflected. Between t_4 and t_5, the sawtooth voltage falls quickly through 0 to its initial value of $-E_2$, snapping the spot back to point 1, in time to sweep forward on the next cycle of signal voltage. When sweep and signal frequencies are equal, a single cycle appears on the screen, when the sweep is lower than the signal, several cycles appear (in the ratio of the two frequencies), and when sweep is higher than signal, less than one cycle appears. The display is stationary only when the two frequencies are either equal or integral multiples of each other. At other frequencies the display will drift horizontally. A sawtooth sweep voltage is generated by a multivibrator, relaxation oscillator or pulse generator. The upper frequency generated by internal devices in the oscilloscope is 50–100 kHz in audio instruments, 500–1000 kHz in TV service instruments and up to several MHz in high quality laboratory instruments. In some oscilloscopes the sweep is calibrated in Hz or kHz, and in others it is calibrated in time units (μs, ms, s). The different types of sweep generated are as follows:

1. Recurrent Sweep When the sawtooths, being an ac voltage alternates rapidly, the display occurs repetitively, so that a lasting image is seen by the eye. This repeated operation is recurrent sweep.

2. Single Sweep The signal under study produces a trigger signal, which in turn produces a single sweep.

3. Driven Sweep The sawtooth oscillator is a free running generator when operated independently. There is a chance that the sweep cycle may start after the signal cycle, thereby missing a part of the signal. Driven sweep removes this possibility because it is fixed by the signal itself. The sweep and signal cycles start at the same time.

4. Triggered Sweep In a recurrent mode, the pattern is repeated again and again. In this mode the voltage rises to a maximum and then suddenly falls to a minimum. The electron beam moves slowly from left to right, retraces rapidly to the left and the pattern is repeated. The horizontal sweep action takes place whether the input signal is applied to the oscilloscope or not, and a horizontal line is displayed on the scope screen.

A triggered sweep, on the other hand, does not start unless initiated by a trigger voltage, generally derived from an incoming signal. In the absence of the input signal, the sweep is held off and the CRT screen is blanked.

The continuous or recurrent sweep uses a free running multivibrator (m/v) which covers a wide frequency range and can be locked into synchronisation by an input signal. Sync takes place when the sweep frequency and the input signal frequency are the same or when the former is a multiple of the latter.

A triggered scope does not use a continuous or recurrent sweep, but uses a monostable multivibrator which is in its off state until a trigger pulse arrives, hence there is no deflection on the screen.

When an input signal is applied, a trigger pulse is generated and applied to the multivibrator, which switches on and produces a sweep signal, and a trace appears on the screen. After a specific voltage, depending on the CRT beam arriving on the RHS, the multivibrator switches back to its off state, causing the beam to return rapidly to the LHS. (The basic difference between recurrent and triggered scopes is that the recurrent sweep locks at the frequency of the input signal, while the triggered scope displays a trace for a specific period of time. Hence, the triggered scope is ON during a specific time interval and will display a waveform or a segment of waveform (e.g. a one shot waveform) regardless of the signal frequency. Hence transients or single clamped oscillations can be observed on the screen.)

Most triggered scopes use a convenient feature of calibrating the sweep speed, in time per cm or division. Sweep frequency is the reciprocal of the time period.

5. Intensity Modulation In some applications an ac signal is applied to the control electrode of the CRT. This causes the intensity of the beam to vary in step with signal alternations. As a result, the trace is brightened during the +ve half cycles and diminished or darkened during –ve half cycles. This process, is called intensity modulation or Z-axis modulation (in contrast to X-axis for horizontal and Y-axis for vertical). It produces bright segments or dots on the trace in response to positive peak or dim segments or holes in response to negative peaks.

BLOCK DIAGRAM OF OSCILLOSCOPE 7.4

The major block circuit shown in Fig. 7.4, of a general purpose CRO, is as follows:

1. CRT
2. Vertical amplifier
3. Delay line
4. Time base
5. Horizontal amplifier
6. Trigger circuit
7. Power supply

The function of the various blocks are as follows.

1. CRT This is the cathode ray tube which emits electrons that strikes the phosphor screen internally to provide a visual display of signal.

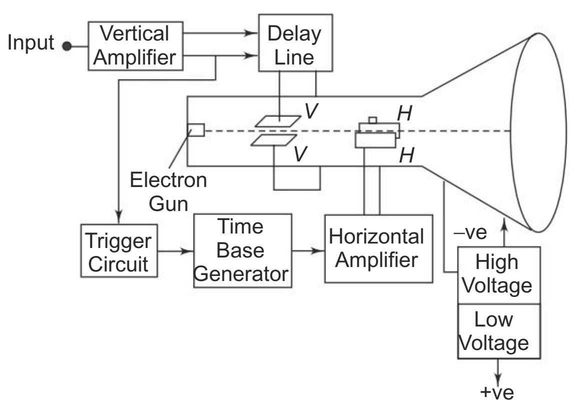


Fig. 7.4 Basic CRO block diagram

2. Vertical Amplifier This is a wide band amplifier used to amplify signals in the vertical section.

3. Delay Line It is used to delay the signal for some time in the vertical sections.

4. Time Base It is used to generate the sawtooth voltage required to deflect the beam in the horizontal section.

5. Horizontal Amplifier This is used to amplify the sawtooth voltage before it is applied to horizontal deflection plates.

6. Trigger Circuit This is used to convert the incoming signal into trigger pulses so that the input signal and the sweep frequency can be synchronised

7. Power Supply There are two power supplies, a –ve High Voltage (HV) supply and a +ve Low Voltage (LV) supply. Two voltages are generated in the CRO. The +ve volt supply is from + 300 to 400 V. The –ve high voltage supply is from – 1000 to – 1500 V. This voltage is passed through a bleeder resistor at a few mA. The intermediate voltages are obtained from the bleeder resistor for intensity, focus and positioning controls.

Advantages of using –ve HV Supply

(i) The accelerating anodes and the deflection plates are close to ground potential. The ground potential protects the operator from HV shocks when making connections to the plates.

(ii) The deflection voltages are measured wrt ground, therefore HV blocking or coupling capacitor are not needed, but low voltage rating capacitors can be used for connecting the HV supply to the vertical and horizontal amplifiers.

(iii) Less insulation is needed between positioning controls and chasis.

SIMPLE CRO 7.5

The basic block diagram of a simple CRO is shown in Fig. 7.5. The ac filament supplies power to the CRT heaters. This also provides an accurate ac calibrating voltage. CRT dc voltage is obtained from the HV dc supply through voltage dividers $R_1 - R_5$. Included along with this voltage divider is a potentiometer (R_3) which varies the potential at the focusing electrode, known as focus control, and one which varies the control grid voltage, called the intensity control (R_5).

Capacitor C_1 is used to ground the deflection plates and the second anode for the signal voltage, but dc isolates these electrodes from the ground.

Normally S_2 is set to its linear position. This connects the sweep generator output to the horizontal input. The sweep voltage is amplified before being applied to the horizontal deflecting plates.

When an externally generated sweep is desired, S_2 is connected to its external position and the external generator is connected to the input. The sweep synchronising voltage is applied to the internal sweep generator through switch S_1, which selects the type of synchronisation.

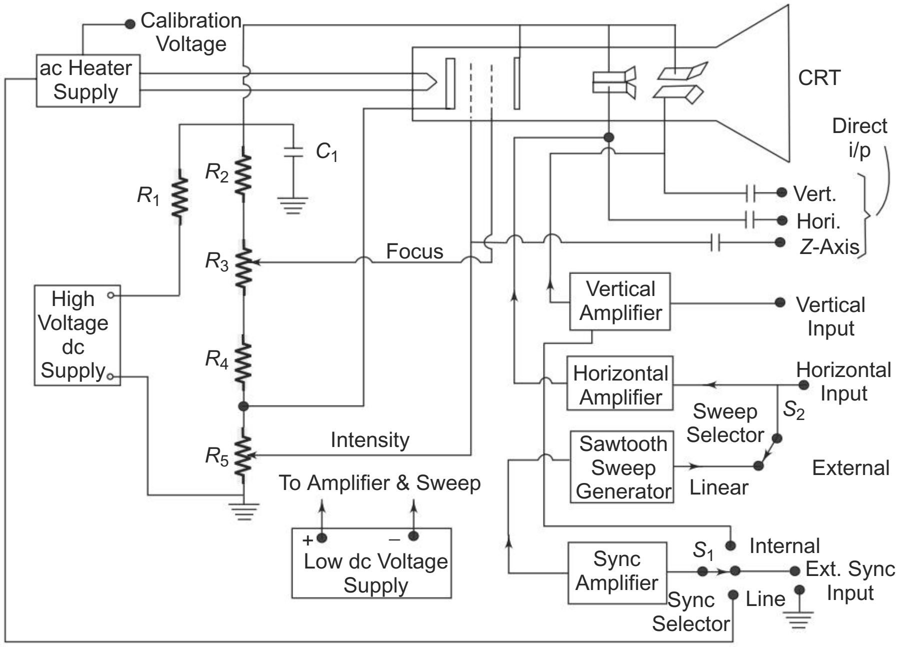


Fig. 7.5 Simple CRO

7.5.1 CRT Showing Power Supply

Figure 7.6 shows the various voltages applied to CRT electrodes. The intensity control controls the number of electrons by varying the control grid voltage. Focusing can be done either electrostatically or electromagnetically. Electrostatic focusing is obtained by using a cylindrical anode, which changes the electrostatic lines of force which controls the beam.

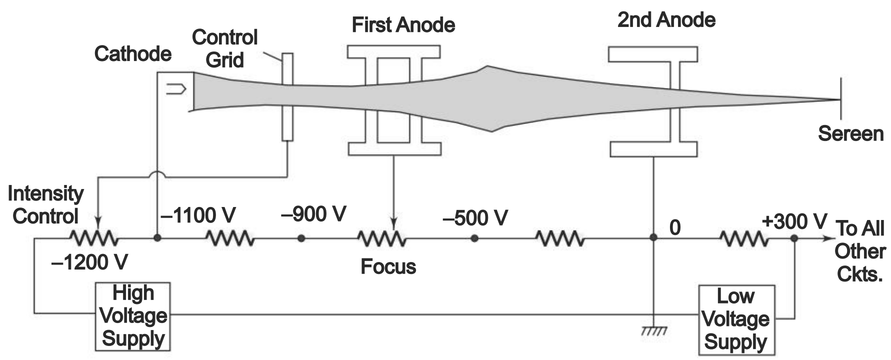


Fig. 7.6 CRT showing power supplies

VERTICAL AMPLIFIER 7.6

The sensitivity (gain) and frequency bandwidth (B.W.) response characteristics of the oscilloscope are mainly determined by the vertical amplifier. Since the gain–

B.W. product is constant, to obtain a greater sensitivity the B.W. is narrowed, or vice-versa.

Some oscilloscopes give two alternatives, switching to a wide bandwidth position, and switching to a high sensitivity position.

Block Diagram of a Vertical Amplifier The block diagram of a vertical amplifier is shown in Fig. 7.7.

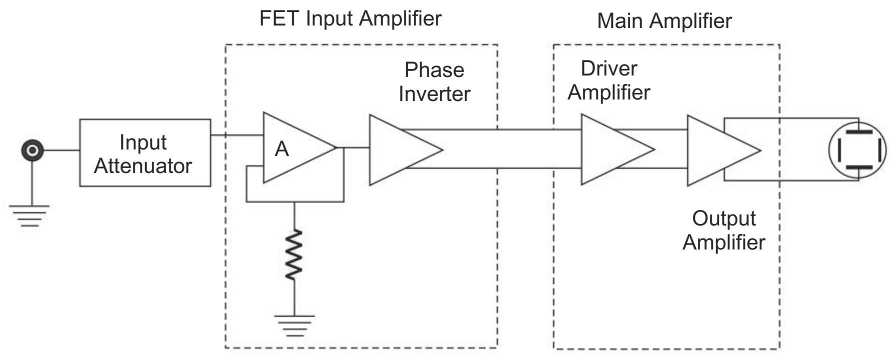


Fig. 7.7 Vertical amplifier

The vertical amplifier consists of several stages, with fixed overall sensitivity or gain expressed in V/div. The advantage of fixed gain is that the amplifier can be more easily designed to meet the requirements of stability and B.W. The vertical amplifier is kept within its signal handling capability by proper selection of the input attenuator switch. The first element of the pre-amplifier is the input stage, often consisting of a FET source follower whose high input impedance isolates the amplifier from the attenuator.

This FET input stage is followed by a BJT emitter follower, to match the medium impedance of FET output with the low impedance input of the phase inverter.

This phase inverter provides two antiphase output signals which are required to operate the push-pull output amplifier. The push-pull output stage delivers equal signal voltages of opposite polarity to the vertical plates of the CRT.

The advantages of push-pull operation in CRO are similar to those obtained from push-pull operation in other applications; better hum voltage cancellation from the source or power supply (i.e. dc), even harmonic suppression, especially the large 2nd harmonic is cancelled out, and greater power output per tube as a result of even harmonic cancellation. In addition, a number of defocusing and non-linear effects are reduced, because neither plate is at ground potential.

HORIZONTAL DEFLECTING SYSTEM 7.7

The horizontal deflecting system consist of a Time Base Generator and an output amplifier.

7.7.1 Sweep or Time Base Generator

A continuous sweep CRO using a UJT as a time base generator is shown in Fig. 7.8. The UJT is used to produce the sweep. When the power is first applied, the UJT is off and the C_T charges exponentially through R_T. The UJT emitter voltage V_E rises towards V_{BB} and when V_E reaches the peak voltage V_P, as shown in Fig. 7.9, the emitter to base '1' (B_1) diode becomes forward biased and the UJT triggers ON. This provides a low resistance discharge path and the capacitor discharges rapidly. The emitter voltage V_E reaches the minimum value rapidly and the UJT goes OFF. The capacitor recharges and the cycle repeats.

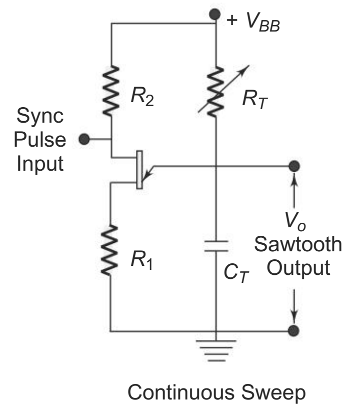


Fig. 7.8 Continuous sweep

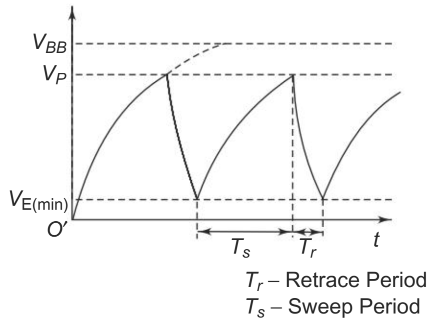


Fig. 7.9 Sawtooth output waveform

To improve sweep linearity, two separate voltage supplies are used, a low voltage supply for UJT and a high voltage supply for the $R_T C_T$ circuit.

R_T is used for continuous control of frequency within a range and C_T is varied or changed in steps for range changing. They are sometimes called as timing resistor and timing capacitor respectively.

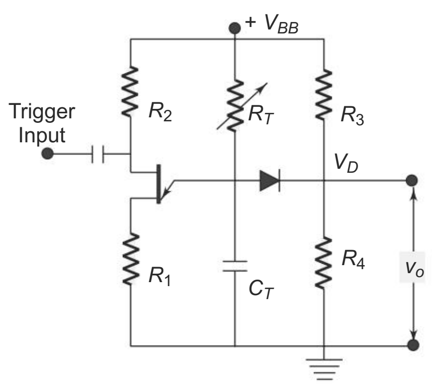
The sync pulse enables the sweep frequency to be exactly equal to the input signal frequency, so that the signal is locked on the screen and does not drift.

TRIGGERED SWEEP CRO 7.8

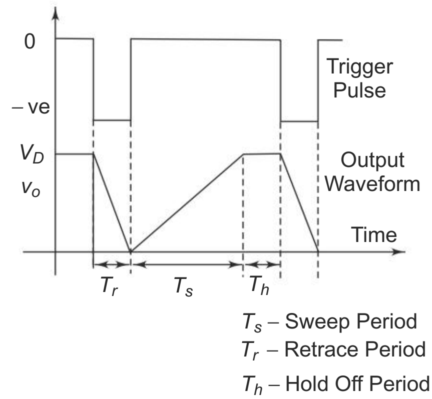
The continuous sweep is of limited use in displaying periodic signals of constant frequency and amplitude. When attempting to display voice or music signals, the pattern falls in and out of sync as the frequency and amplitude of the music varies resulting in an unstable display.

A triggered sweep can display such signals, and those of short duration, e.g. narrow pulses. In triggered mode, the input signal is used to generate substantial pulses that trigger the sweep. Thus ensuring that the sweep is always in step with the signal that drives it.

As shown in Fig. 7.10, resistance R_3 and R_4 form a voltage divider such that the voltage V_D at the cathode of the diode is below the peak voltage V_P for UJT conduction. When the circuit is switched on, the UJT is in the non-conducting stage, and C_T charges exponentially through R_T towards V_{BB} until the diode becomes forward biased and conducts; the capacitor voltage never reaches the peak voltage required for UJT conduction but is clamped at V_D. If now a –ve pulse of sufficient amplitude is applied to the base and the peak voltage V_p is momentarily lowered, the UJT fires. As a result, capacitor C_T discharges rapidly through the UJT until the maintaining voltage of the UJT is reached; at this point the UJT switches off and capacitor C_T charges towards V_{BB}, until it is clamped again at V_D. Figure 7.11 shows the output waveform.



Triggered sweep



Output waveform

TRIGGER PULSE CIRCUIT 7.9

The trigger circuit is activated by signals of a variety of shapes and amplitudes, which are converted to trigger pulses of uniform amplitude for the precision sweep operation. If the trigger level is set too low, the trigger generator will not operate. On the other hand, if the level is too high, the UJT may conduct for too long and part of the leading edge of the input signal may be lost.

The trigger selection is a 3-position switch, Internal-External-Line, as shown in Fig. 7.12. The trigger input signal is applied to a voltage comparator whose reference level is set by the Trigger Level control on the CRO front panel.

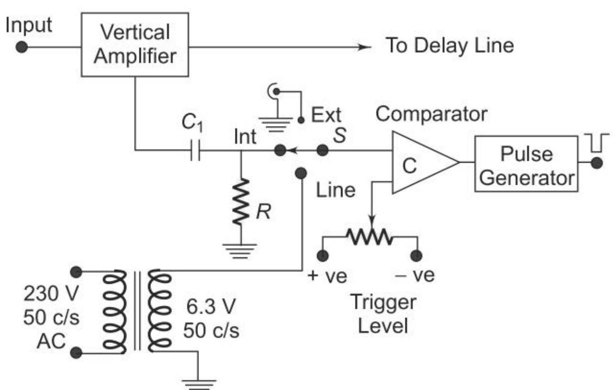


Fig. 7.12 Trigger pulse circuit

The comparator circuit C produces a change in the output whenever the trigger input exceeds the present trigger levels. The pulse generator that follows the comparator produces –ve trigger pulses each time the comparator output crosses its quiescent level, which in turn triggers the sweep generator to start the next sweep. The trigger sweep generator contains the stability or sync control, which prevents the display from jittering or running on the screen. Stability is secured by proper adjustments of the sweep speed. Sweep speed is adjustable by means of a sweep rate control and its multiplier, i.e. range control. The timing resistance R_T is used for sweep rate control and timing capacitor C_T is changed in steps for sweep rate control.

DELAY LINE IN TRIGGERED SWEEP 7.10

Figure 7.13 shows a delay line circuit.

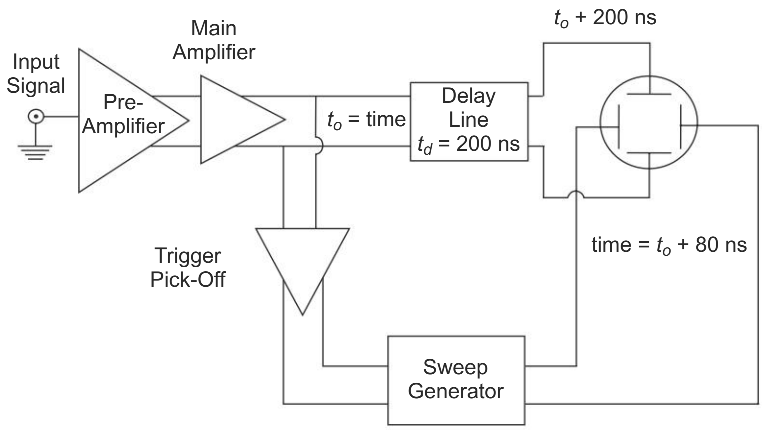


Fig. 7.13 Delay line circuit

Figure 7.14 indicates the amplitude of the signal wrt time and the relative position of the sweep generator output signal. The diagram shows that when the delay line is not used, the initial part of the signal is lost and only part of the signal is displayed. To counteract this disadvantage the signal is not applied directly to the vertical plates but is passed through a delay line circuit, as shown in Fig. 7.13. This gives time for the sweep to start at the horizontal plates before the signal has reached the vertical plates. The trigger pulse is picked off at a time t_o after the signal has passed through the main amplifier. The sweep generator delivers the sweep to the horizontal amplifier and the sweep starts at the HDP at time t_o + 80 ns. Hence the sweep starts well in time, since the signal arrives at the VDP at time t_o + 200 ns.

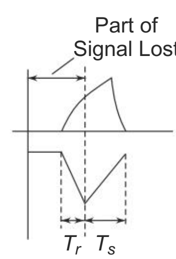


Fig. 7.14 Delay line waveform

SYNC SELECTOR FOR CONTINUOUS SWEEP CRO 7.11

The sync selector is a 3-position switch, Int-Ext-Line. Therefore horizontal sweep can be synchronised with the signals coming from any of the three sources, as shown in Fig. 7.15.

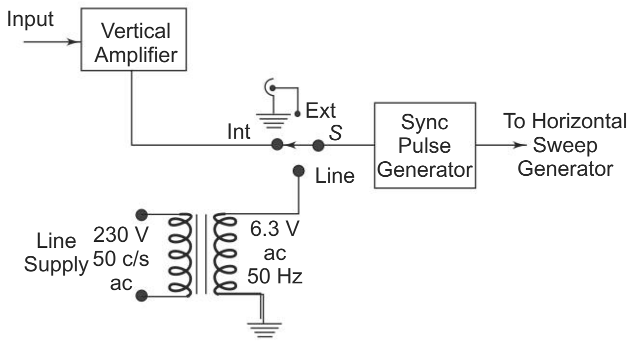


Fig. 7.15 Sync selector

TYPICAL CRT CONNECTIONS 7.12

Figure 7.16 shows various controls and CRT connections.

The following controls are available on CRO panel.

1. Intensity It controls the magnitude of emission of the electron beam, i.e. the electron beam is adjusted by varying the cathode-to-grid bias voltage. This adjustment is done by the 500 kΩ potentiometer.

2. Focus The focusing anode potential is adjusted with respect to the first and final accelerating anodes. This is done by the 2 MΩ potentiometer. It adjusts the negative voltage on the focus ring between –500 V and –900 V.

3. Astigmatism It adjusts the voltage on the acceleration anode with respect to the VDP of the CRT. This arrangement forms a cylindrical lens that corrects any defocusing that might be present. This adjustment is made to obtain the roundest spot on the screen.

4. X-shift or Horizontal Position Control The X-position of the spot is adjusted by varying the voltage between the horizontal plates. When the spot is in the center position, the two horizontal plates have the same potential.

5. Y-shift or Vertical Position Control The Y-position of the spot is adjusted by varying the voltage between the vertical plates. When the spot is in the center position, the two vertical plates have the same potential.

6. Time Base Control This is obtained by varying the C_T and R_T of the time base generator.

7. Sync Selector It can synchronise the sweep to signals coming internally from the vertical amplifier or an external signal or the line supply i.e. the Int-Ext-Line switch.

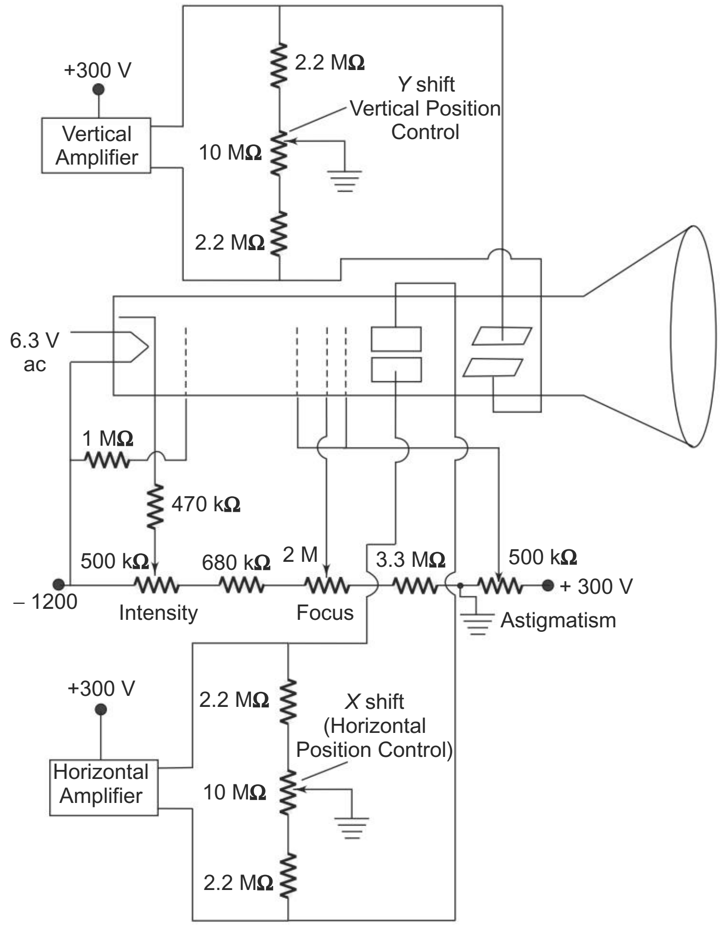


Fig. 7.16 Typical CRT connections

HIGH FREQUENCY CRT OR TRAVELLING WAVE TYPE CRT 7.13

Figure 7.17 illustrates a high frequency CRT.

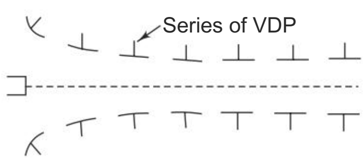


Fig. 7.17 Travelling wave CRO

In an ordinary CRO, there is only one pair of VDPs. When the signal to be displayed is of a very high frequency, the electron beam does not get sufficient time to pick up the instantaneous level of the signal. Also, at high frequencies the numbers of electrons striking the screen in a given time and the intensity of the beam is reduced. Hence, instead of one set of vertical deflection plates, a series of vertical deflection plates are used. The plates are so shaped and spaced that an electron travelling along the CRT receives from each set of plates an additional deflecting force in proper time sequence. This synchronisation is achieved by making the signal travel from one plate to the next at the same speed as the transit time of the electrons. The signal is applied to each pair of plates, and as the electron beam travels the signal also travels through the delay lines. The time delays are so arranged that the same electrons are deflected by the input signal. In this way the electron beam picks up the level of the input signal. The time delays between the plates correspond exactly to the transit times of the electrons. (In addition, new fluorescent materials have now been developed to increase the brightness at HF.)

7.13.1 Characteristics of a HF CRO or (HF Improvement in a CRO)

1. The vertical amplifier must be designed both for high B.W. and high sensitivity or gain. Making the vertical amplifier a fixed gain amplifier simplifies the design. The input to the amplifier is brought to the required level by means of an attenuator circuit. The final stages is the push-pull stage.
2. The LF CRT is replaced by an HF CRT.
3. A probe is used to connect the signals, e.g. a high Z passive probe acts like a compensated attenuator.
4. By using a triggered sweep, for fast rising signals, and by the use of delay lines between the vertical plates, for improvement of HF characteristics.
5. New fluorescent materials that increase the brightness of the display are used.

DUAL BEAM CRO 7.14

Figure 7.18 illustrates a block diagram of a Dual Beam CRO.

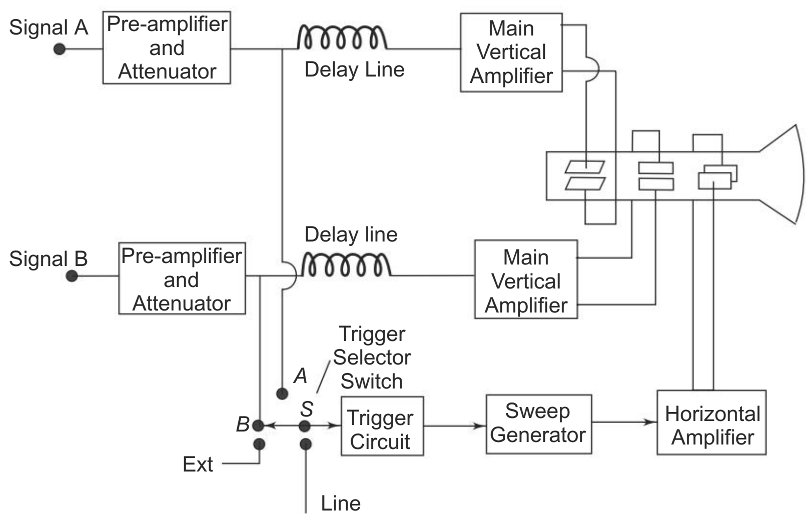


Fig. 7.18 Dual beam CRO

The dual trace oscilloscope has one cathode ray gun, and an electronic switch which switches two signals to a single vertical amplifier. The dual beam CRO uses two completely separate electron beams, two sets of VDPs and a single set of HDPs. Only one beam can be synchronised at one time, since the sweep is the same for both signals, i.e. a common time base is used for both beams. Therefore, the signals must have the same frequency or must be related harmonically, in order to obtain both beams locked on the CRT screen, e.g. the input signal of an amplifier can be used as signal A and its output signal as signal B.

DUAL TRACE OSCILLOSCOPE 7.15

Figure 7.19 (a) shows a block diagram of a dual trace oscilloscope.

This CRO has a single electron gun whose electron beam is split into two by an electronic switch. There is one control for focus and another for intensity. Two signals are displayed simultaneously. The signals pass through identical vertical channels or vertical amplifiers. Each channel has its own calibrated input attenuator and positioning control, so that the amplitude of each signal can be independently adjusted.

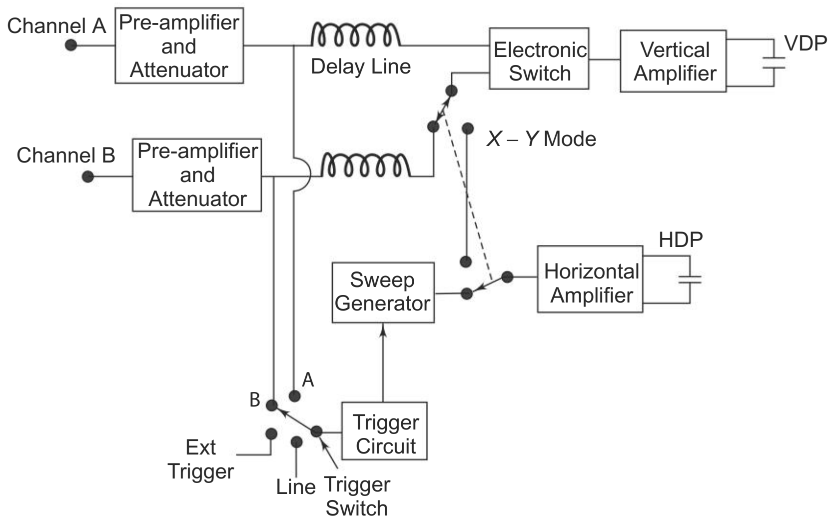


Fig. 7.19 (a) Dual trace oscilloscope

A mode control switch enables the electronic switch to operate in two modes i.e. Alternate and Chop mode. When the switch is in ALTERNATE position, the electronic switch feeds each signal alternately to the vertical amplifier. The electronic switch alternately connects the main vertical amplifier to channels *A* and *B* and adds a different dc component to each signal; this dc component directs the beam alternately to the upper or lower half of the screen. The switching takes place at the start of each new sweep of the sweep generator. The switching rate of the electronic switch is synchronised to the sweep rate, so that the CRT spot traces the channel *A* signal on one sweep and the channel *B* signal on the succeeding sweep [Fig. 7.19 (b)].

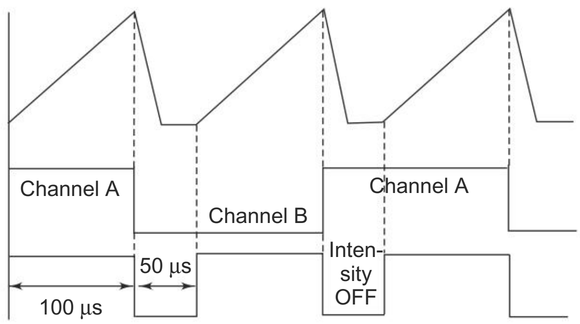


Fig. 7.19 (b) Time relation of a dual-channel vertical amplifier in alternate mode

The sweep trigger signal is available from channels A or B and the trigger pick-off takes place before the electronic switch. This arrangement maintains the correct phase relationship between signals A and B.

When the switch is in the CHOP mode position, the electronic switch is free running at the rate of 100–500 kHz, entirely independent of the frequency of the sweep generator. The switch successively connects small segments of A and B waveforms to the main vertical amplifier at a relatively fast chopping rate of 500 kHz e.g. 1 μs segments of each waveform are fed to the CRT display (Fig. 7.19 (c)).

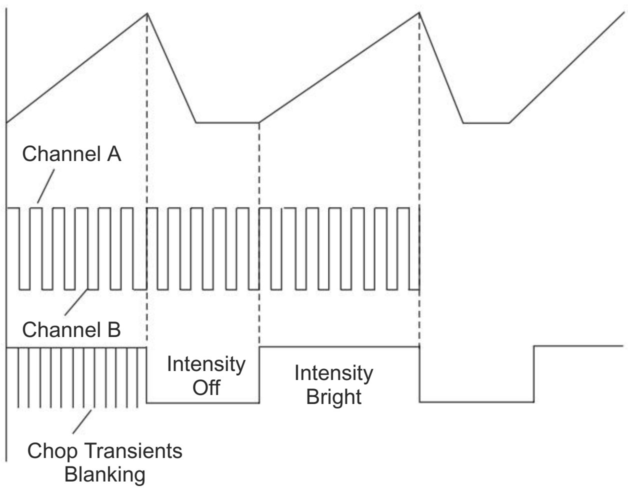


Fig. 7.19 (c) Time relation of a dual-channel vertical amplifier in chop mode

If the chopping rate is slow, the continuity of the display is lost and it is better to use the alternate mode of operation. In the added mode of operation a single image can be displayed by the addition of signal from channels A and B, i.e. $(A + B)$, etc. In the $X - Y$ mode of operation, the sweep generator is disconnected and channel B is connected to the horizontal amplifier. Since both pre-amplifiers are identical and have the same delay time, accurate $X - Y$ measurements can be made.

7.15.1 Dual Trace Oscilloscope (0–15 MHz) Block Description

Y-Channels *A* and *B* vertical channels are identical for producing the dual trace facility. Each comprises an input coupling switch, an input step attenuator, a source follower input stage with protection circuit, a pre-amplifier from which a trigger signal is derived and a combined final amplifier. The input stage protection circuit consists of a diode, which prevents damage to the FET transistors that could occur with excessive negative input potentials, and a resistor network which protects the input stage from large positive voltage swings.

As the transistors are the balanced pre-amplifier stage, they share the same IC block. The resulting stabilisation provides a measure of correction to reduce the drift inherent in high gain amplifiers. The trigger pick-off signal is taken from one side of the balanced pre-amplifier to the trigger mode switch, where either channel *A* or channel *B* triggering can be selected. The supply for the output of the pre-amplifier stage is derived from a constant current source controlled by the channel switching logic. Under the control of channel switching, signals from *A* and *B* channels are switched to the final amplifier. The combined balanced final amplifier is a direct coupled one to the *Y*-plates of the CRT (refer to Fig. 7.20).

Channel Switching The front panel *A* and *B* channel selection (push button or switch), controls an oscillator in the CHOP mode. For channel switching electronic switching logic and a F/F is used. When either *A* or *B* channels are selected, the F/F is switched to allow the appropriate channel.

In the ALTERNATE mode, a pulse from the sweep-gating multivibrator via the electronic switching logic, switches the F/F, thus allowing *A* and *B* channels for alternate sweeps.

In the CHOP mode, the oscillator is switched via the logic stage to provide rapid switching of the channels via the F/F.

Triggering A triggering signal can be obtained from the vertical amplifier of Channels *A* and *B* from an external source or internally from the mains supply (LINE triggering). The triggering signal is selected and normally fed via the amplifier stage to the pulse shaper, which supplies well defined trigger pulses to the sweep-gating multivibrator for starting the sawtooth generator.

Triggering from the TV line and frame signals can be obtained from the sync separator and peak detector stages. The latter stage is switched into circuit in the TOP position.

Time Base The time base generator circuit operates on the constant current integrator principle.

The sweep-gating multivibrator, triggered by pulses from the differentiator and auto circuits, starts the sawtooth generator. Sweep signals are fed to the final X-amplifier.

A gate pulse is supplied by the sweep-gating multivibrator for unblanking the CRT during the forward sweep. In addition this pulse is supplied to an external socket for probe adjustment via a diode network.

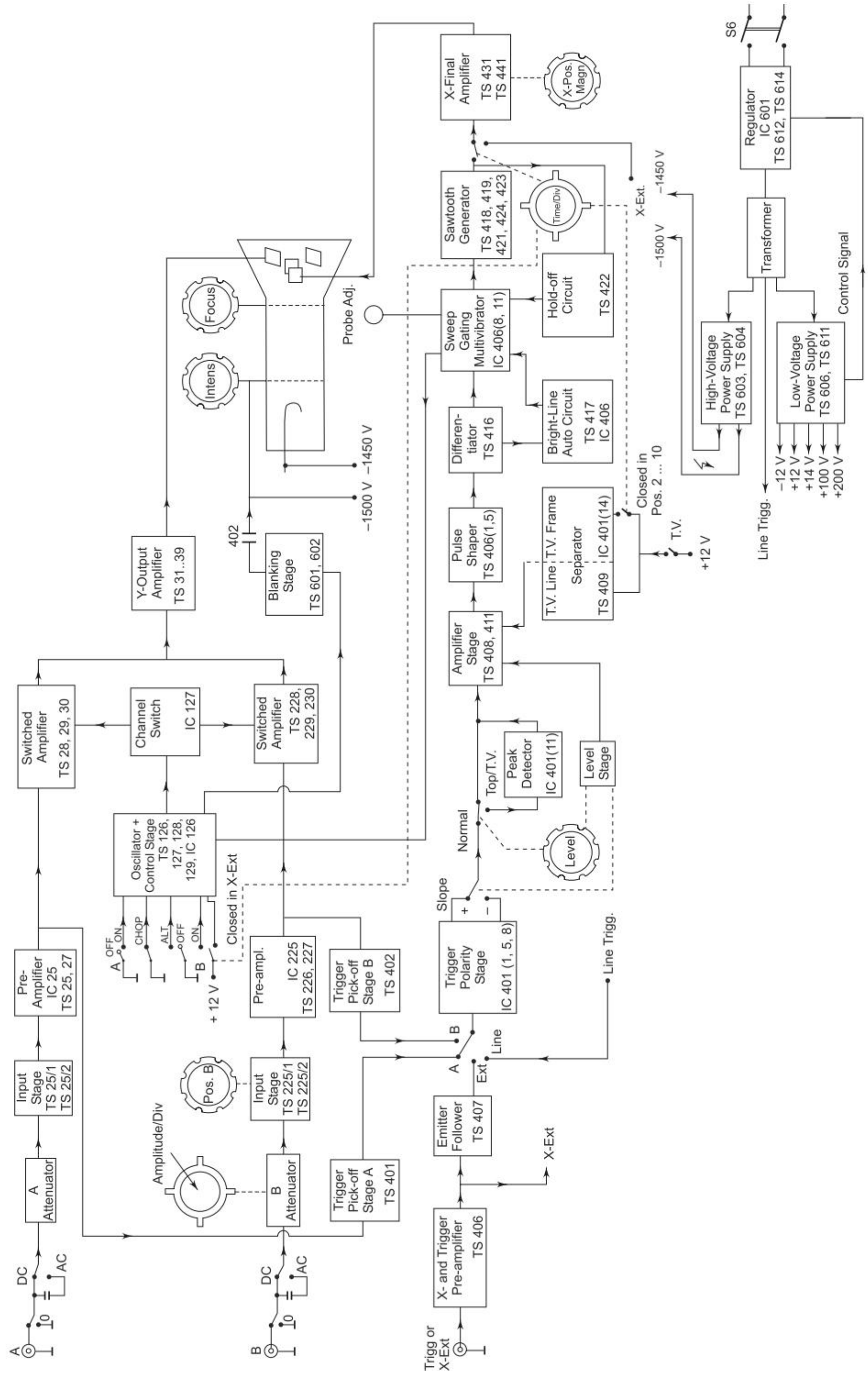


Fig. 7.20 Block diagram of dual trace CRO (Practical)

X-Channel Under the control of diode switching from the TIME/DIV switch, the X-amplifier receives its input signal from either the time base sawtooth generator or from an external source (X-EXT input socket via the *X* and trigger pre-amplifier). The X-MAGN (× 5) circuit is incorporated in the X-final amplifier. The output of this amplifier is direct coupled to the horizontal deflection plates of the CRT.

Cathode-Ray Tube Circuit and Power Supply The high voltages required for the CRT, which has an acceleration potential of 1.5 kV, are generated by a voltage multiplier circuit controlled by a stabilised power supply. The CRT beam current is controlled by:

The intensity potentials network across the Extra High Tension (EHT) supply. During flyback (movement of electron beam from right to left) by the blanking pulses coming from the sawtooth generator via the beam blanking stages to blank the trace during right to left movement of the electron.

Regulation of the mains input voltage is achieved by a diode clipper network controlled by a signal fed back from an LED in the + 14 V rectifier supply.

7.15.2 Dual Trace CRO Specifications

Maximum sensitivity	:	5 mV/div and B.W. 15 MHz
Operating temperature	:	+ 5° to 40°C
CRT	:	
Measuring area	:	8 × 10 Div (1 cm = 1 Div)
Screen type	:	B31 or 3B1
Total acceleration voltage	:	2kV
Vertical Amplifier	:	
Display modes	:	A, A and B, B (chopped in ms) (alternated in µs)
Input coupling	:	AC/DC Bandwidth DC – 0 – 15 MHz (– 3db) AC – 10 Hz – 15 MHz (– 3db)
Deflection accuracy	:	± 5%
Input impedance	:	1 MΩ/35 pF
Maximum rated input voltage	:	400 V (dc + ac peak) (no damage)
Chopper frequency	:	120 kHz approximately
Time Base		
Time coefficients	:	0.2 s/Div to 0.5 s/Div in 2 × 9
	:	calibrated steps (1-2-5 sequence) with
	:	5 × magnifier, max. 0.1 µs/Div
	:	Uncalibrated continuous control 1 : ≥ 2, 5
Coefficient error	:	± 5%
Additional error for × 5 magnifier	:	± 2%

Trigger		
Source	:	CH A, B or Ext
Mode	:	AC/TV
Sensitivity Int	:	0.75 Div/0.75 V – Trigger freq. at 100 kHz.
Ext	:	1 Div/1.0 V – Trigger freq. at 15 MHz.
Trigger frequency	:	10 Hz – 15 MHz
Input impedance	:	1 MΩ/35 pF
Z-Modulation input trace blanking	:	TTL high blanks trace
Power Supply		
Voltage range	:	220 V ± 10%
Frequency	:	50 Hz
Power	:	30 VA
Size	:	378 (L) × 348 (W) × 142 (H)
Weight	:	5 kg Approx.
X-Deflection		
Phase shift	:	3 at 10 kHz
Accuracy	:	± 5%
μs/ms	:	Slide switch in combination with time base
	:	The display is chopped in ms and Alternated in μs

Figure 7.21 illustrates a 30 MHz Dual Trace Oscilloscope with Delayed Sweep and Frequency Counter (I.E. 234).

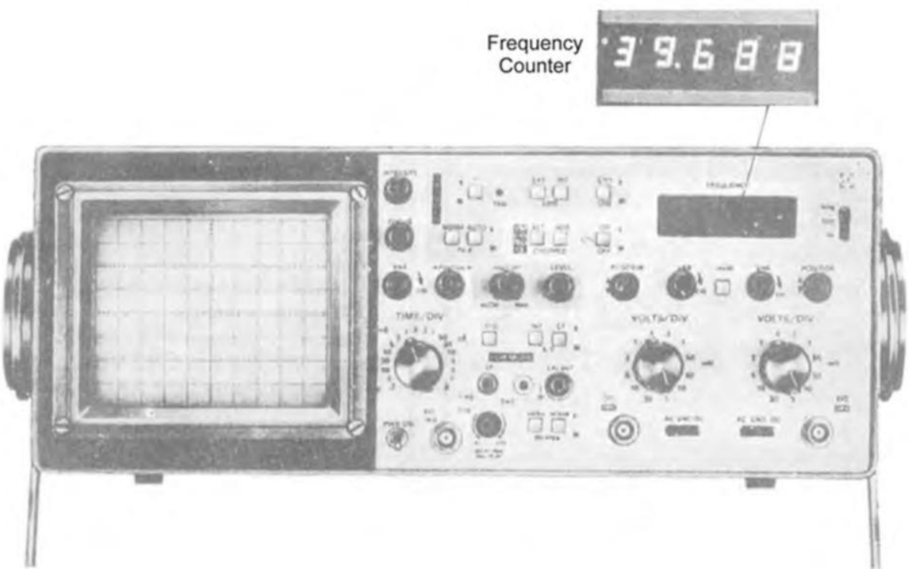


Fig. 7.21 30 MHz Dual trace oscilloscope with delayed sweep and frequency counter (Courtesy: International Electronics Ltd., Bombay; Marketed by: Signetics Electronics Ltd.)

ELECTRONIC SWITCH 7.16

The electronic switch is a device that enables two signals to be displayed simultaneously on the screen by a single gun CRT. The basic block diagram of an electronic switch is shown in Fig. 7.22.

Each signal is applied to a separate gain control and gate stage. The gates stage are alternately biased to cut off by square wave signals from the square wave generator. Therefore only one gate stage is in a condition to pass its signal at any given time.

The outputs of both stages are applied directly to the oscilloscope input.

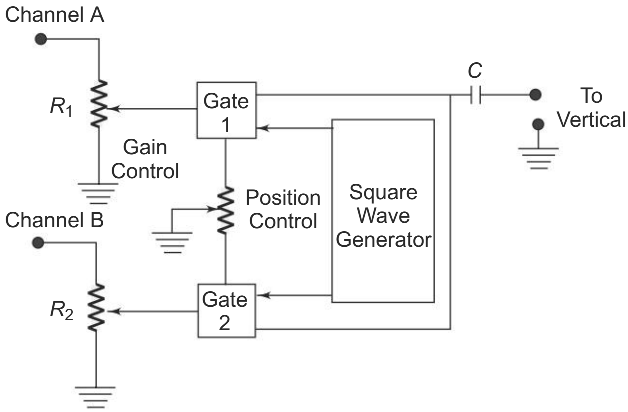


Fig. 7.22 Basic block diagram of an electronic switch

R_1 and R_2 are gain controls used to adjust the amplitudes of Channels A and B.

In the circuit diagram of Fig. 7.23, Q_1 and Q_2 are the amplifiers and Q_3 and Q_4 the switches. Input signal 1 is applied to Q_1 through gain control R_1, and input signal 2 is applied to Q_2 through gain control R_2. The square wave generator

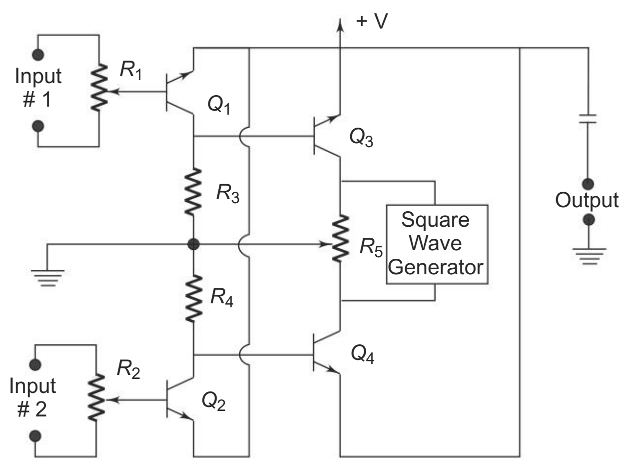


Fig. 7.23 Electronic switch

alternately biases first Q_3 and then Q_4 to cut off. When Q_3 is cut off, Q_4 conducts and transmits signal 2 to the output terminals. When Q_4 is cut off, Q_3 conducts and transmits signal 1 to the output terminals.

When the square wave generator switching frequency is much higher than either signal frequency, bits of each signal are alternately presented to the oscilloscopes vertical input to reproduce the two signals on the screen.

The traces can be moved up or down by the position control R_5. The traces can be overlapped for easy comparison. The heights of the individual signals can be adjusted by means of gain controls R_1 and R_2. The sweep signal produced by this design is very linear and can be calibrated in time per cm or inch, so that accurate time and frequency can be measured.

(VHF) SAMPLING OSCILLOSCOPE 7.17

An ordinary oscilloscope has a B.W. of 10 MHz. The HF performance can be improved by means of sampling the input waveform and reconstructing its shape from the sample, i.e. the signal to be observed is sampled and after a few cycles the sampling point is advanced and another sample is taken. The shape of the waveform is reconstructed by joining the sample levels together. The sampling frequency may be as low as 1/10th of the input signal frequency (if the input signal frequency is 100 MHz, the bandwidth of the CRO vertical amplifier can be as low as 10 MHz). As many as 1000 samples are used to reconstruct the original waveform.

Figure 7.24 shows a block diagram of a sampling oscilloscope. The input waveform is applied to the sampling gate. The input waveform is sampled whenever a sampling pulse opens the sampling gate. The sampling must be synchronised with the input signal frequency. The signal is delayed in the vertical amplifier, allowing the horizontal sweep to be initiated by the input signal. The waveforms are shown in Fig. 7.25.

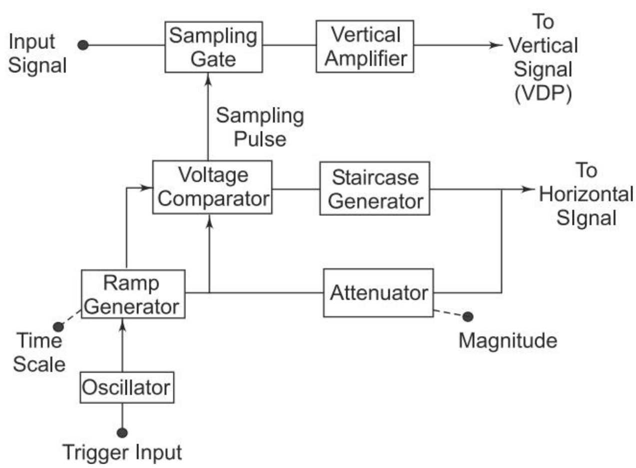


Fig. 7.24 Sampling oscilloscope

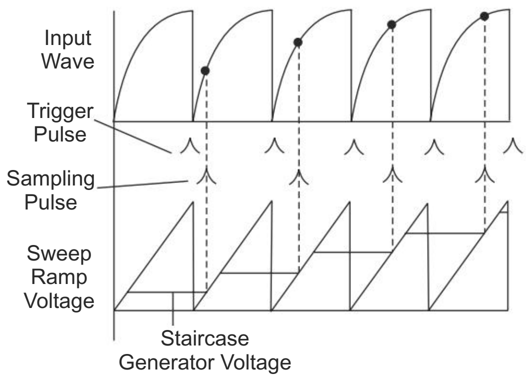


Fig. 7.25 Various waveforms at each block of a sampling oscilloscope

At the beginning of each sampling cycle, the trigger pulse activates an oscillator and a linear ramp voltage is generated. This ramp voltage is applied to a voltage comparator which compares the ramp voltage to a staircase generator. When the two voltages are equal in amplitude, the staircase advances one step and a sampling pulse is generated, which opens the sampling gate for a sample of input voltage.

The resolution of the final image depends upon the size of the steps of the staircase generator. The smaller the size of the steps the larger the number of samples and higher the resolution of the image.

STORAGE OSCILLOSCOPE (FOR VLF SIGNAL) 7.18

Storage targets can be distinguished from standard phosphor targets by their ability to retain a waveform pattern for a long time, independent of phosphor peristence. Two storage techniques are used in oscilloscope CRTs, mesh storage and phosphor storage.

A mesh-storage CRT uses a dielectric material deposited on a storage mesh as the storage target. This mesh is placed between the deflection plates and the standard phosphor target in the CRT. The writing beam, which is the focussed electron beam of the standard CRT, charges the dieletric material positively where hit. The storage target is then bombarded with low velocity electrons from a flood gun and the positively charged areas of the storage target allow these electrons to pass through to the standard phosphor target and thereby reproduce the stored image on the screen. Thus the mesh storage has both a storage target and a phosphor display target. The phosphor storage CRT uses a thin layer of phosphor to serve both as the storage and the display element.

Mesh Storage It is used to display Very Low Frequencies (VLF) signals and finds many applications in mechanical and biomedical fields. The conventional scope has a display with a phosphor peristence ranging from a few micro seconds to a few seconds. The persistence can be increased to a few hours from a few seconds.

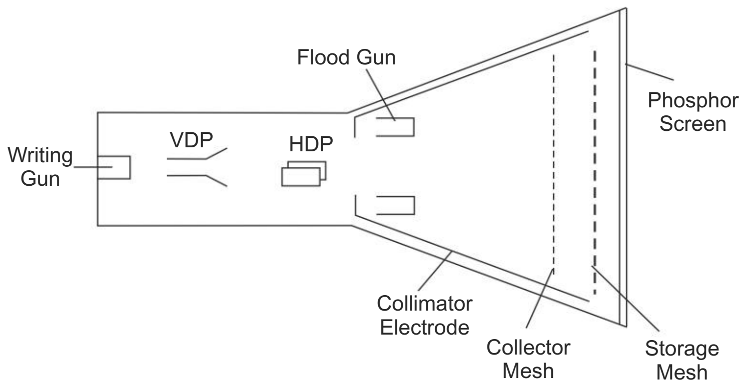


Fig. 7.26 Basic elements of storage mesh CRT

A mesh storage CRT, shown in Fig. 7.26, contains a dielectric material deposited on a storage mesh, a collector mesh, flood guns and a collimator, in addition to all the elements of a standard CRT. The storage target, a thin deposition of a dielectric material such as Magnesium Fluoride on the storage mesh, makes use of a property known as secondary emission. The writing gun etches a positively charged pattern on the storage mesh or target by knocking off secondary emission electrons. Because of the excellent insulating property of the Magnesium Fluoride coating, this positively charged pattern remains exactly in the position where it is deposited. In order to make a pattern visible, a special electron gun, called the flood gun, is switched on (even after many hours). The electron paths are adjusted by the collimator electrode, which constitutes a low voltage electrostatic lens system (to focus the electron beam), as shown in Fig. 7.27. Most of the electrons are stopped and collected by the collector mesh. Only electrons near the stored positive charge are pulled to the storage target with sufficient force to hit the phosphor screen. The CRT will now display the signal and it will remain visible as long as the flood guns operate. To erase the pattern on the storage mesh, a negative voltage is applied to neutralise the stored positive charge.

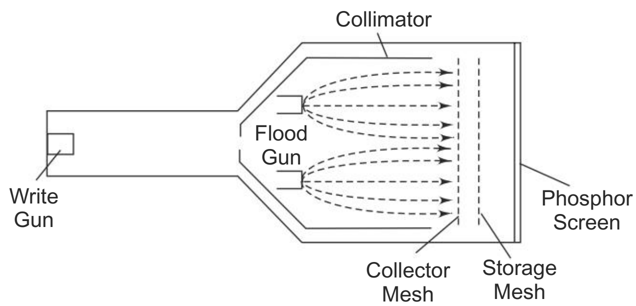


Fig. 7.27 Storage mesh CRT

Since the storage mesh makes use of secondary emission, between the first and second crossover more electrons are emitted than are absorbed by the material, and hence a net positive charge results.

Below the first crossover a net negative charge results, since the impinging electrons do not have sufficient energy to force an equal number to be emitted. In order to store a trace, assume that the storage surface is uniformly charged and write gun (beam emission gun) will hit the storage target. Those areas of the storage surface hit by the deflecting beam lose electrons, which are collected by the collector mesh. Hence, the write beam deflection pattern is traced on the storage surface as a positive charge pattern. Since the insulation of the dielectric material is high enough to prevent any loss of charge for a considerable length of time, the pattern is stored. To view, the stored trace, a flood gun is used when the write gun is turned off. The flood gun, biased very near the storage mesh potential, emits a flood of electrons which move towards the collector mesh, since it is biased slightly more positive than the deflection region. The collimator, a conductive coating on the CRT envelope with an applied potential, helps to align the flood electrons so that they approach the storage target perpendicularly. When the electrons penetrate beyond the collector mesh, they encounter either a positively charged region on the storage surface or a negatively charged region where no trace has been stored. The positively charged areas allow the electrons to pass through to the post accelerator region and the display target phosphor. The negatively charged region repels the flood electrons back to the collector mesh. Thus the charge pattern on the storage surface appears reproduced on the CRT display phosphor just as though it were being traced with a deflected beam.

Figure 7.28 shows a display of the stored charge pattern on a mesh storage.

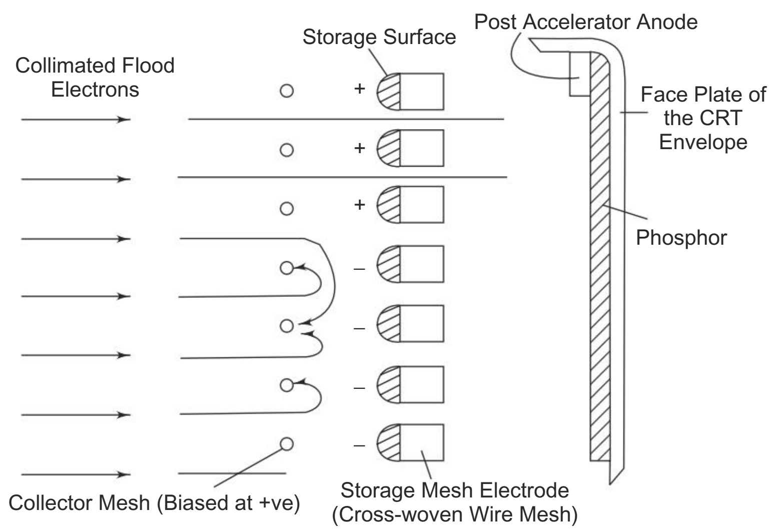


Fig. 7.28 Display of stored charged pattern on a mesh-storage

DIGITAL READOUT OSCILLOSCOPE 7.19

The digital read out oscilloscope instrument has a CRT display and a counter display. The diagram shown is of an instrument where the counter measures the time (Fig. 7.29).

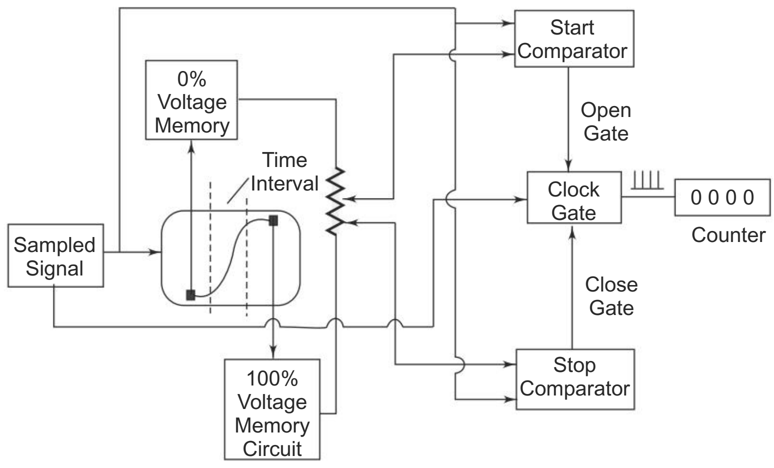


Fig. 7.29 Block diagram of a digital readout oscilloscope when measuring voltage

The input waveform is sampled and the sampling circuit advances the sampling position in fixed increments, a process called strobing. The equivalent time between each sample depends on the numbers of sample taken per cm and on the sweep time/cm, e.g. a sweep rate of 1 nano-sec/cm and a sampling rate of 100 samples/cm gives a time of 10 pico-sec/sample.

Figure 7.29 shows a block diagram of a digital read out oscilloscope when measuring voltage.

Two intensified portions of the CRT trace identify 0% and 100% zones position. Each zone can be shifted to any part of the display. The voltage divider taps between the 0% and 100% memory voltage are set for start and stop timing. The coincidence of any of the input waveforms with the selected percentage point is sensed by this voltage comparator. The numbers of the clock pulse which correspond to the actual sample taken are read out digitally in a Nixie display tube in ns, µs, ms or seconds.

Figure 7.30 shows a block diagram of a digital readout CRO when used for voltage to time conversion.

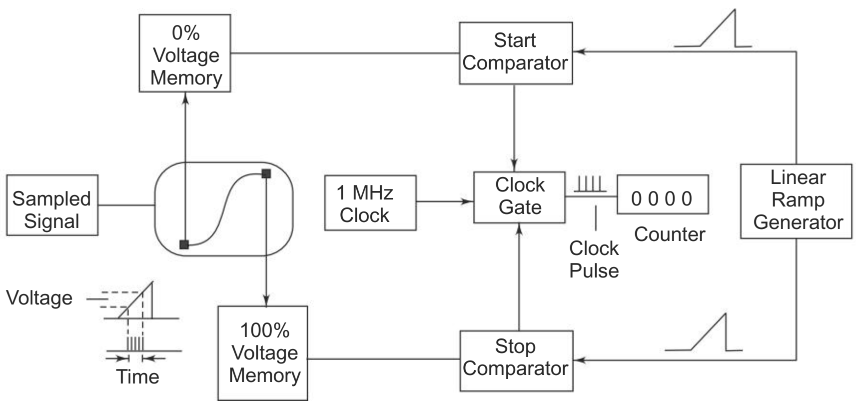


Fig. 7.30 Voltage to time conversion

The CRT display is obtained by sampling the 0% reference voltage as chosen by the memory circuit. A linear ramp generator produces a voltage; when the ramp voltage equals the 0% reference the gate opens. When the ramp equals 100% reference the gate closes. The number of clock pulses that activate the counter is directly proportional to the voltage between the selected reference and is read out in mV or volts by the Nixie tube display.

MEASUREMENT OF FREQUENCY BY LISSAJOUS METHOD 7.20

The oscilloscope is a sensitive indicator for frequency and phase measurements. The techniques used are simple and dependable, and measurement may be made at any frequency in the response range of the oscilloscope.

One of the quickest methods of determining frequency is by using Lissajous patterns produced on a screen. This particular pattern results when sine waves are applied simultaneously to both pairs of the deflection plates. If one frequency is an integral multiple (harmonic) of the other, the pattern will be stationary, and is called a Lissajous figure.

In this method of measurement a standard frequency is applied to one set of deflection plates of the CRT tube while the unknown frequency (of approximately the same amplitude) is simultaneously applied to the other set of plates. However, the unknown frequency is presented to the vertical plates and the known frequency (standard) to the horizontal plates. The resulting patterns depend on the integral and phase relationship between the two frequencies. (The horizontal signal is designated as f_h and the vertical signal as f_v.)

Typical Lissajous figures are shown in Figs 7.31 and 7.32 for sinusoidal frequencies which are equal, integral and in ratio.

7.20.1 Measurement Procedure

Set up the oscilloscope and switch off the internal sweep (change to Ext). Switch off sync control. Connect the signal source as given in Fig. 7.33. Set the horizontal and vertical gain control for the desired width and height of the pattern. Keep frequency f_v constant and vary frequency f_h, noting that the pattern spins in alternate directions and changes shape. The pattern stands still whenever f_v and f_h are in an integral ratio (either even or odd). The $f_v = f_h$ pattern stands still and is a single circle or ellipse. When $f_v = 2f_h$, a two loop horizontal pattern is obtained as shown in Fig. 7.31.

To determine the frequency from any Lissajous figure, count the number of horizontal loops in the pattern, divide it by the number of vertical loops and multiply this quantity by f_h (known or standard frequency).

In Fig. 7.31 (g), there is one horizontal loop and 3 vertical loops, giving a fraction of 1/3. The unknown frequency f_v is therefore 1/3 f_h. An accurately calibrated, variable frequency oscillator will supply the horizontal search frequency for frequency measurement. For the case where the two frequencies are equal and in phase, the pattern appears as a straight line at an angle of 45° with the horizontal. As the phase between the two alternating signals changes, the pattern changes

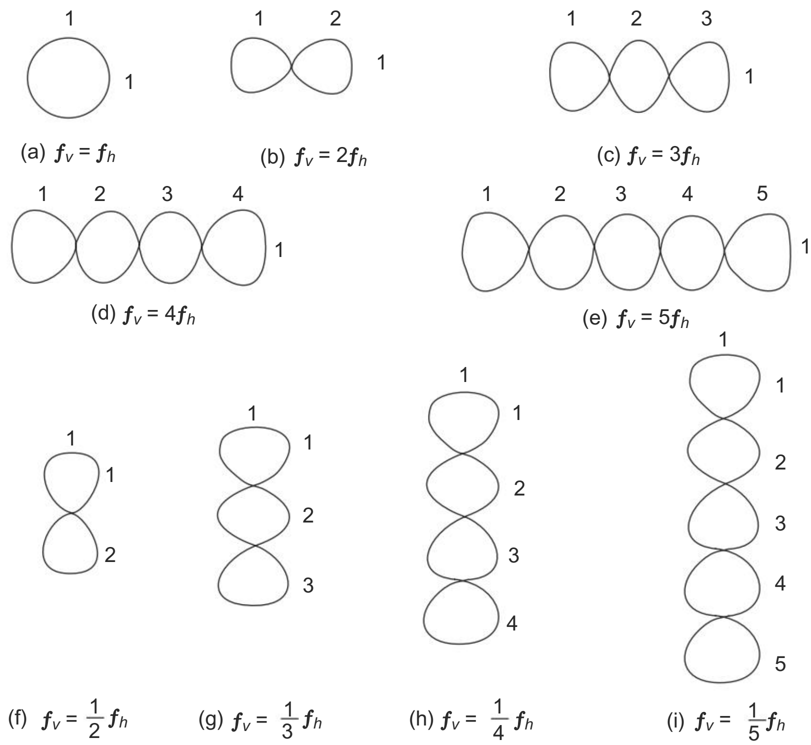


Fig. 7.31 Lissajous patterns for integral frequencies

cyclically, i.e. an ellipse (at 45° with the horizontal) when the phase difference is $\pi/4$, a circle when the phase difference is $\pi/2$ and an ellipse (at 135° with horizontal) when the phase difference is $3\pi/4$, and a straight line pattern (at 135° with the horizontal) when the phase difference is π radians.

As the phase angle between the two signals changes from π to 2π radians, the pattern changes correspondingly through the ellipse-circle-ellipse cycle to a straight line. Hence the two frequencies, as well as the phase displacement can be compared using Lissajous figures techniques.

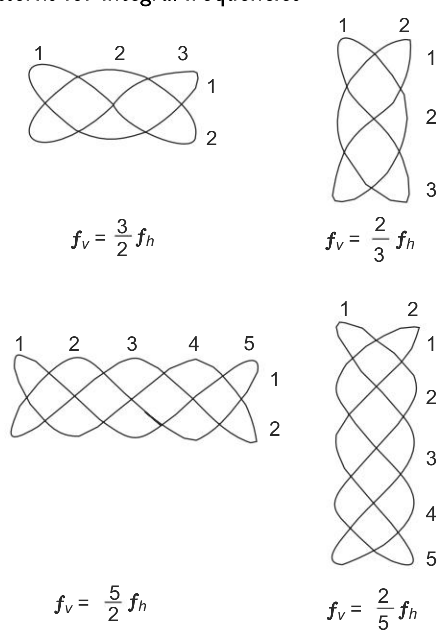


Fig. 7.32 Lissajous patterns for non-integral frequencies

When the two frequencies being compared are not equal, but are fractionally related, a more complex stationary pattern results, whose form is dependent on the frequency ratio and the relative phase between the two signals, as in Fig. 7.32.

The fractional relationship between the two frequencies is determined by counting the number of cycles in the vertical and horizontal.

$$f_v = (\text{fraction}) \times f_h$$

or

$$\frac{f_v}{f_h} = \frac{\text{number of horizontal tangencies}}{\text{number of vertical tangencies}}$$

Figure 7.33 illustrates the basic circuit for comparing two frequencies by the Lissajous method.

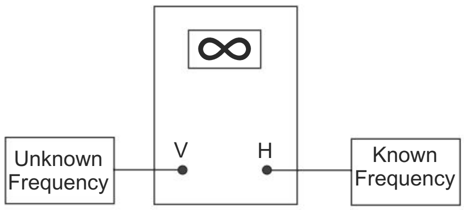


Fig. 7.33 Basic circuit for frequency measurements with lissajous figures

SPOT WHEEL METHOD 7.21

As the ratio of the two frequencies being compared increases, the Lissajous pattern becomes more complicated and hence more difficult to interpret. In such cases, it is advantageous to use the spot-wheel method of display as shown in Fig. 7.34 (a) whereby an intensity modulated circular or elliptical is produced on the face of the CRT. The lesser of the two frequencies being compared is applied to the deflecting plates of the CRT through the resistance-capacitance phase shifter as illustrated, thereby producing the circular (or elliptical) sweep. The higher frequency is applied to the control grid of the CRT (biased slightly below cut off), as shown in Fig. 7.34 (b), thus modulating the intensity of the beam. As a consequence, the pattern appears as a series of alternate bright and dark spots, as shown in Fig. 7.34 (c), where the ratio of the high to low frequencies is given by the number of bright spots.

The application of two sine wave signals, equal in amplitude but 90° out of phase, to the vertical and horizontal deflecting plates results in the formation of a circle. If a single sine wave is applied to a phase shifting circuit, it is possible to obtain two outputs, equal in amplitude and 90° out of phase.

If the voltages across R and C are equal, the pattern is a circle. If they are unequal, the pattern is elliptical.

($R = 1/\omega C$ where $\omega = 2\pi f$. The voltage across C lags the voltage across R by 90°.)

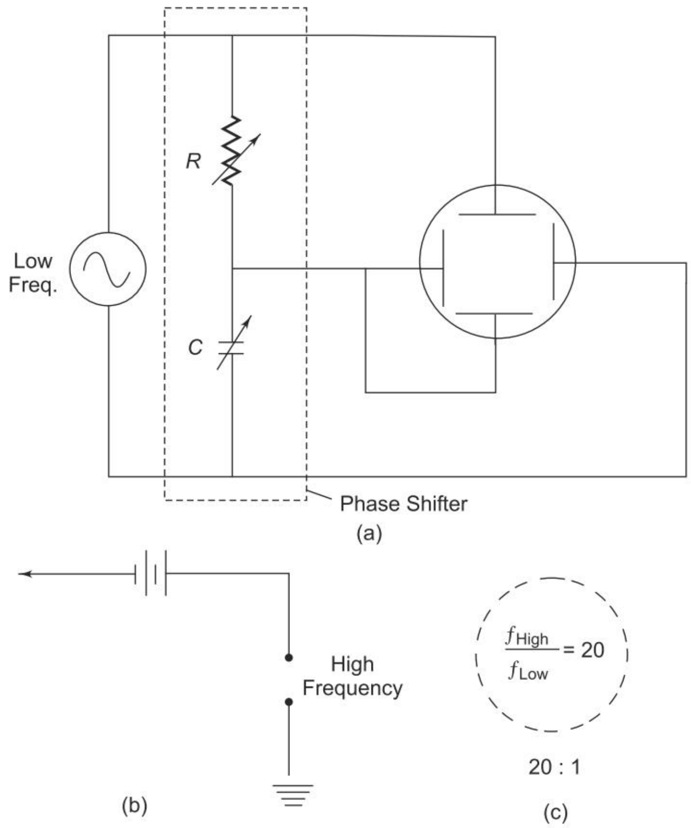


Fig. 7.34 (a) Basic circuit of a spot wheel method for frequency measurement (b) Input circuit to the control grid (c) Spot wheel pattern

Varying the grid-cathode potential changes the density of the electron stream within the CRT and determines the intensity. When the grid bias of the CRT reaches cut off potential, there is no electron stream and fluorescence at the screen cannot occur. This property is used for the comparison of two frequencies, provided their ratio is an integer.

The high frequency signal is applied to the grid. Sufficient grid bias must be present so that the negative peaks of the sine wave cuts off the electron beam. The high frequency signal does not have to be a pure sine wave, it can be a square wave.

In Fig. 7.34 (c), the frequency is 20 : 1 and there are 20 blanks in the pattern.

GEAR WHEEL METHOD 7.22

When a Lissajous figure contains a large number of loops, accurate counting becomes difficult. Figure 7.35 shows a test method that uses a modulated ring pattern in place of the looped figure and permits a higher count. This pattern is also called a Gear wheel or toothed wheel, because of its shape. The unknown frequency is determined by multiplying the known frequency by the number of teeth in the pattern. Figure. 7.36 shows the test set up. Here, a phase shift network (*RC*) introduces a 90° phase shift between the horizontal and vertical channels of the oscilloscope, which is needed to produce a ring or circle pattern with the known frequency f_v.

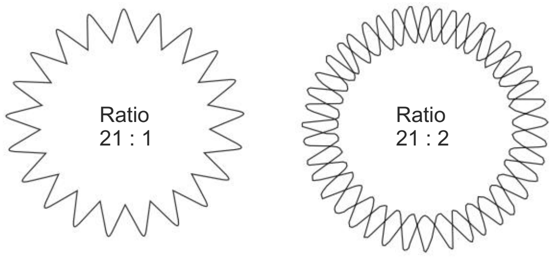


Fig. 7.35 Gear wheel pattern

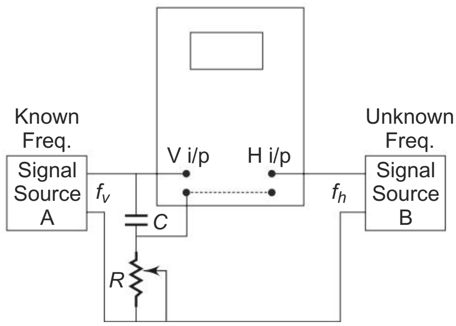


Fig. 7.36 Setup for gear wheel method for frequency measurement

A voltage from the unknown frequency source modulates this ring. When the voltages across *R* and *C* are unequal, the pattern is elliptical. If they are equal, the pattern is a circle. The unknown frequency must be higher than the known frequency and the amplitude of the unknown must be below that of the known to prevent distortion.

7.22.1 Measurement Procedure

Set up the oscilloscope. Switch off the Internal and Sync control. Connect the circuit as in Fig. 7.36, and temporarily switch off signal *B*.

Switch on signal source *A* and adjust *R* to obtain a ring pattern on the screen. Adjust the horizontal and vertical gain controls to spread the ring over the screen as much as possible.

Switch on signal source *B*, noting that the ring becomes toothed by the signal. Adjust the known frequency f_v to stop the ring from rotating. Adjust the amplitude of f_h to obtain a distinct teeth pattern. The unknown frequency is then given by $f_h = nf_v$, where f_h is an integral multiple (odd or even) of f_v. The wheel rotates or spins if f_h is not an exact multiple of f_v. Therefore, adjust the known frequency to stop the wheel from rotating.

CHECKING OF DIODES 7.23

The voltage-current characteristics curve of a crystal diode may be observed using the circuit given in Fig. 7.37 (a).

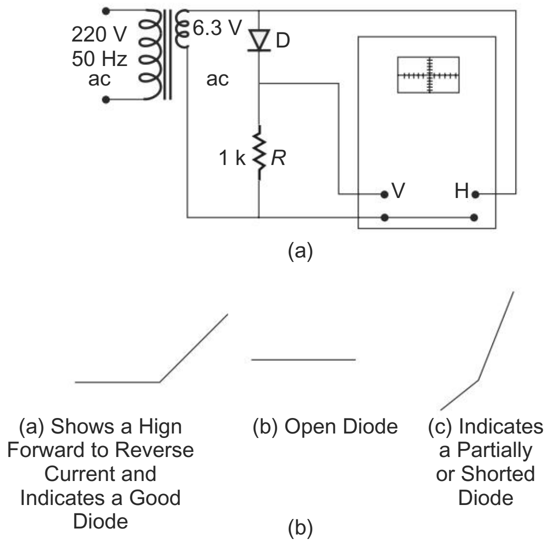


Fig. 7.37 (a) Setup for diode testing (b) Display pattern for diode testing

In this case the internal sweep is not used. Horizonal deflection is obtained by connecting the 6.3 V source directly to the horizontal input and switching the sweep control to horizontal amplifier. This provides an horizontal trace that is proportional to the applied voltage and represents the input signal.

The diode current flowing through the 1 kΩ resistance will develop a voltage across the resistance, which is applied to the vertical input terminals. This represents the rectified signal current, since the voltage drop across R is proportional to the diode current.

If the diode being tested is good, the current through it will be unidirectional, causing the curve to rise vertically from its flat portion. The flat portion represents little or no current in the reverse direction. If the active portion of the curve points downwards, the diode connections should be reversed. The angle between the active and the flat portions of the curve represents the condition of the diode as shown in Fig. 7.37 (b).

BASIC MEASUREMENT OF CAPACITANCE AND INDUCTANCE 7.24

Capacitance may be measured from a capacitor discharge curve. This requires a dc oscilloscope with a triggered single sweep that is of the storage type.

Figure 7.38 (a) shows the test setup. C is the capacitor under test, and E is 1.5 V battery for charging the capacitor. R_1 is the precision 1 W non-inductive resistor. (R_1 should be 100 K for testing capacitances from 10 pf – 0.1 µf, 10 K for 0.1 – 10 μf, 1 K for 10 – 100 μf and 100 Ω for 100 – 1000 μf) R_2 is a 1 MΩ resistor to isolate R_1 from the oscilloscope input resistance.

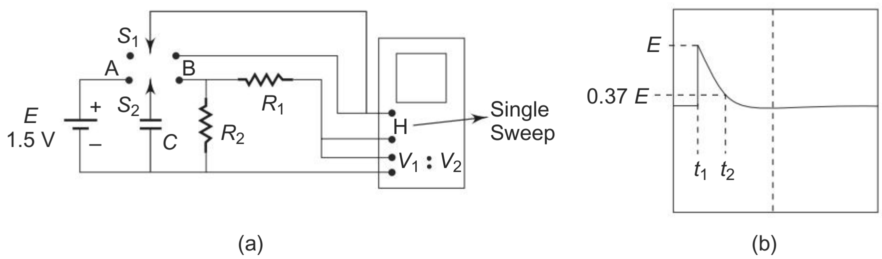


Fig. 7.38 (a) Basic setup for measurement of capacitance (b) Discharge curve

When the DPDT switch ($S_1 - S_2$) is connected to position A, capacitor C is charged by the battery; when this switch is connected to position B, the capacitor discharges through R_1. At this time, section S_1 of the switch triggers a single sweep of the oscilloscope. The shape of the discharge curve which is displayed on the screen is shown in Fig. 7.38 (b). When the ($S_1 - S_2$) switch is connected to position B, the voltage across R_1 rises suddenly to the fully charged value E at time t_1. The capacitor immediately begins to discharge through R_1. At time t_2 the voltage has fallen to 37% of the maximum. The unknown capacitor (in μf) is determined from the time interval between t_1 and t_2 (in ms) and the resistance R_1 in ohms.

Therefore $$C = 1000\left(\frac{t_2 - t_1}{R_1}\right)$$

If a single trace oscilloscope is used, the time interval t_2–t_1 may be measured from (i) the calibrated sweep, and (ii) the calibrated horizontal axis.

Inductances may be measured by means of a time constant method similar to that used in the capacitance case. Figure 7.39 (a) shows the test setup. L is the inductor under test and E is a battery or filtered dc power supply. R_1 is the rheostat for adjusting the current level. A dc ammeter or mA can be temporarily inserted in series with the battery for indicating the maximum safe operating current of the L. The resistance of the rheostat at this setting must be noted.

When the DPST switch is open, no current flows through the inductor and the oscilloscope receives no signal voltage. When the switch is closed, S_1 closes the dc circuit, and current flows through L and R_1 in series. This produces a voltage drop across R_1, which is the signal voltage applied to the oscilloscope. At the same time, S_2 closes the trigger circuit and initiates a single sweep of the oscilloscope. Because of the counter emf generated by the inductor, the current increases exponentially, as shown in Fig. 7.39 (b). At time t_1, the voltage has risen to 63% of its final or maximum value, E. The unknown inductance (in Henries) is determined from the time interval between t_0 and t_1 (t_0 = 0) and resistance $R_1 \cdot t_1$ and t_0 may be measured (i) from the calibrated sweep and (ii) from the time calibrated horizontal axis.

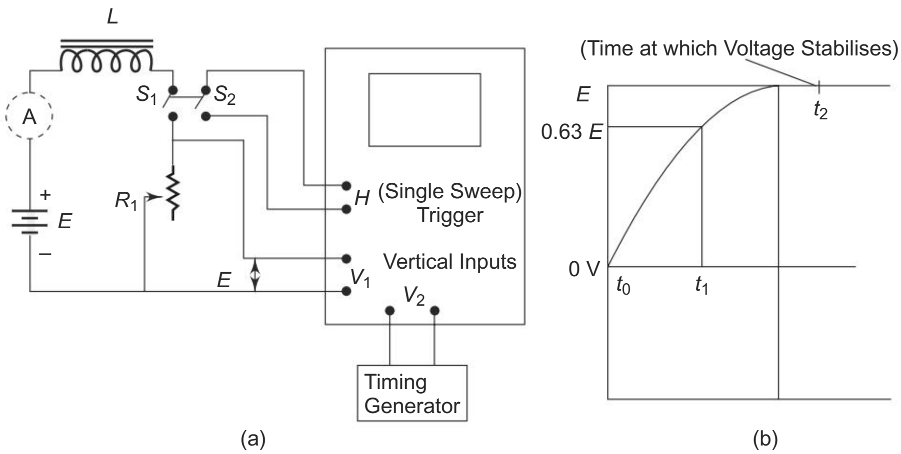


Fig. 7.39 (a) Setup for inductance measurement (b) Display curve

OSCILLOSCOPE AS A BRIDGE NULL DETECTOR 7.25

As a null detector for an ac bridge, the oscilloscope, unlike a meter used for the purpose, gives separate indications for reactive and resistive balances of the bridge.

Figure 7.40 shows how an oscilloscope is connected to the bridge. Here T is the shielded transformer and *C–R* form an adjustable phase shift network. The generator voltage is applied to the horizontal input through a phase shifter, and simultaneously to the bridge input. The bridge output signal is applied to the oscilloscope vertical input. Figure 7.41 shows the type of pattern obtained.

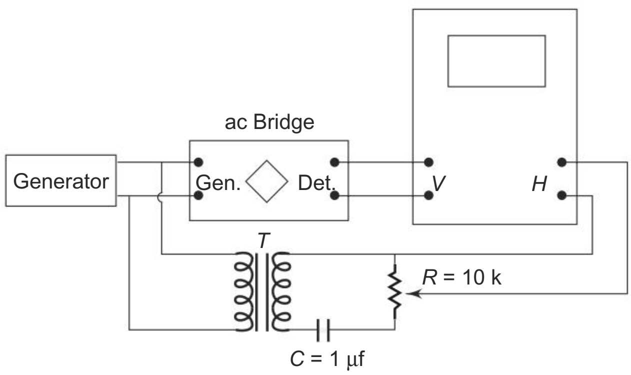


Fig. 7.40 Setup of oscilloscope as a bridge null-detector

7.25.1 Measurement Procedure

Connect the circuit as shown in Fig. 7.40. With the bridge unbalanced, but with test components (*R*, *C*, *L*) connected to the unknown arm of the bridge, adjust *R* in the phase shifter to give an elliptical pattern on the screen. Adjust the vertical and horizontal gain controls to obtain an ellipse of suitable height. Adjust the reactance control of the bridge, noting whether the ellipse tilts to the right or the left. When reactance balance is complete, i.e. reactance = 0, the ellipse will be horizontal.

Adjust the resistance (power factor, dissipation factor, or Q) control of the bridge, noting that the ellipse closes. At complete null (reactance and resistance, both balanced), a straight horizontal line is obtained. If the resistance is balanced while the reactance is not, tilted ellipse collapses, giving a single line trace tilted to the right or left.

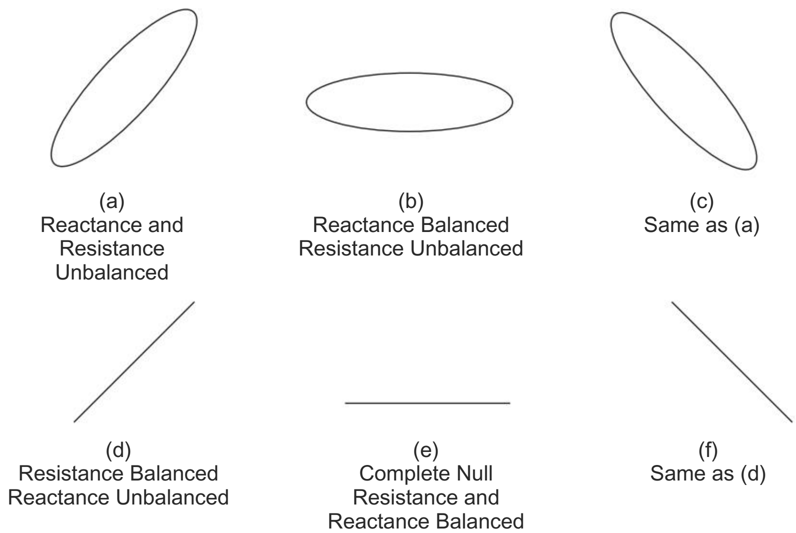


Fig. 7.41 Oscilloscope patterns

A simpler but less effective way of using an oscilloscope as a null detector is to connect the bridge output to the vertical terminals and use the oscilloscope as a voltmeter to give its lowest reading at null.

USE OF LISSAJOUS FIGURES FOR PHASE MEASUREMENT 7.26

When two signals are applied simultaneously to an oscilloscope without internal sweep, one to the horizontal channel and the other to the vertical channel, the resulting pattern is a Lissajous figure that shows a phase difference between the two signals. Such patterns result from the sweeping of one signal by the other.

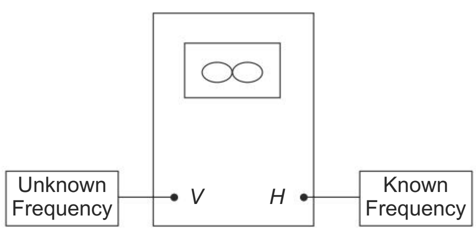


Fig. 7.42 Setup for phase measurement

Figure 7.42 shows the test setup for phase measurement by means of Lissajous figures. Figure 7.43 shows patterns corresponding to

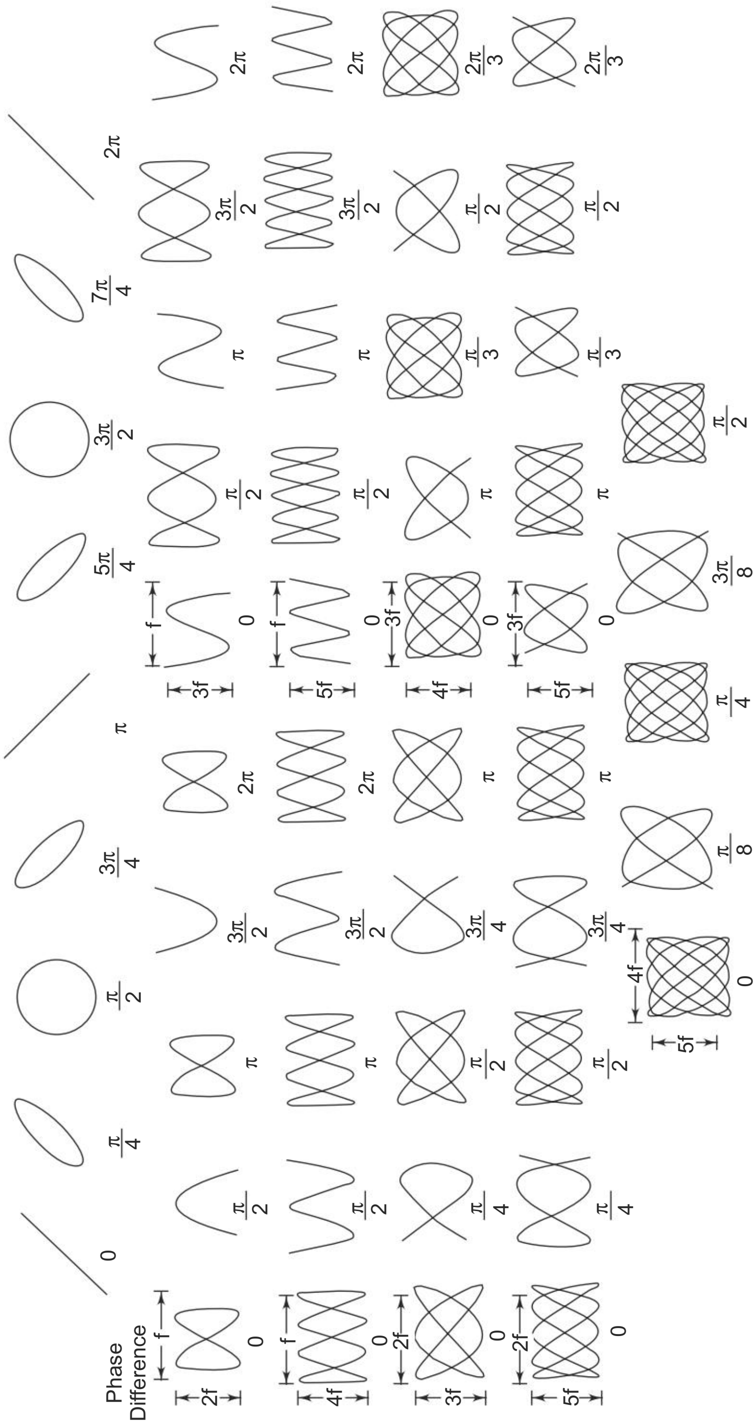


Fig. 7.43 Lissajous pattern

certain phase difference angles, when the two signal voltages are sinusoidal, equal in amplitude and frequency.

A simple way to find the correct phase angle (whether leading or lagging) is to introduce a small, known phase shift to one of the inputs. The proper angle may be then deduced by noting the direction in which the pattern changes.

STANDARD SPECIFICATIONS OF A SINGLE BEAM CRO 7.27

Vertical Amplifier		
Sensitivity	:	5 mV/Div. to 20 V/Div. in 12 calibrated steps in a 1, 2, 5 sequence. Continuous control (uncalibrated) between steps, reduces the sensitivity by a minimum of 2.5 times.
Accuracy	:	± 3%
Bandwidth	:	dc to 20 MHz (– 3 db), dc coupling
	:	0.5 Hz to 20 MHz (– 3 db) ac coupling
Rise time	:	Better than 18 ns
Input Impedance	:	1 MΩ/40 pf
Maximum input voltage	:	400 V (dc + ac peak)
Signal delay	:	Built in delay line sufficient to display leading edge of the waveform
Time Base		
Sweep ranges	:	0.1 μs/Div. to 0.5 s/Div. in 21 calibrated steps in a 1, 2, 5 sequence. Continuous uncalibrated control between steps extending slowest speed to 1.5 s/Div.
Accuracy	:	± 5%
Magnification	:	5 times. Takes the highest speed to 20 ns/Div.
Triggering		
Auto mode	:	Free running in the absence of a trigger signal. Triggers to the input signal automatically.
Level	:	Continuously adjustable on the + ve and – ve going slopes to trigger signal. Level adjustable over 8 Divs.
Source	:	Internal-External-Line
Polarity	:	Positive or negative
Maximum trigger input	:	250 V (dc + ac peak) short term
Input impedance	:	1 MΩ/30 pf
Internal trigger level	:	3 Div from 2 Hz to 20 MHz (1 Div, 30 Hz to 20 MHz in Auto mode)
External trigger level	:	3 V peak to peak, 2 Hz to 20 MHz (1 V, 30 Hz to 20 MHz in Auto mode)
Horizontal Amplifier		
Bandwidth	:	dc – 2 MHz (– 3 db)
Sensitivity	:	100 mV and 0.5 V/Div

Input impedance	:	1 MΩ/50 pf
Maximum input voltage	:	250 V (dc + ac peak)
Calibration	:	200 mV peak to peak square wave at 1 kHz
Cathode ray tube	:	Flat faced medium persistance
Accelerating Potential	:	4.5 kV
Graticule	:	8 × 10 Div of 8 mm each
Power requirements	:	230 V ac, 50 Hz, 50 W
Dimensions	:	220 × 275 × 430 mm
Weight	:	10kg approximately
Optional accessories	:	(i) × 1 probe
		(ii) Oscilloscope trolley
		(iii) × 10 probe (10 MΩ/12 pf)

PROBES FOR CRO 7.28

7.28.1 Direct Probes (1 : 1)

The simplest types of probe (one can hardly call it a probe) is the test lead. Test leads are simply convenient lengths of wire for connecting the CRO input to the point of observation. At the CRO end, they usually terminate with lugs, banana tips or other tips to fit the input jacks of the scope, and at the other end have a crocodile clip or any other convenient means for connection to the electronic circuit.

Since a CRO has high input impedance and high sensitivity, the test leads should be shielded to avoid hum pickup, unless the scope is connected to low impedance high level circuits.

Although the input impedances of most CROs are relatively very high compared to the circuits where they are connected, it is often desirable to increase their impedance to avoid loading of the circuits or causing unstable effects.

The input capacitance of the scope, plus the stray capacitance of the test leads, may be just enough to cause a sensitive circuit to break into oscillation when the CRO is connected. This effect can be prevented by an isolation probe made by placing a carbon resistor in series with the test lead, as shown in Fig. 7.44.

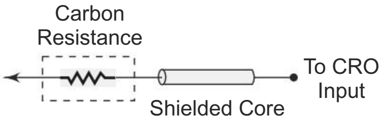


Fig. 7.44 Isolation probe

A slight reduction in the amplitude of the waveform and a slight change in the waveshape occurs with this probe. To avoid this possibility, a high impedance compensated probe, called a low capacitance probe or a 10 : 1 probe, is used.

7.28.2 Passive Voltage (High Z) Probe

Figure 7.45 (a) shows a 10 : 1 probe. Figure 7.45 (b) shows the equivalent circuit. Referring to Fig. 7.45 (b). The capacitor is adjusted so that the elements of the bridge are balanced. Under conditions of balance we have

$$R_1 X_{(C_2 + C_{in})} = R_{in} X(C_1)$$

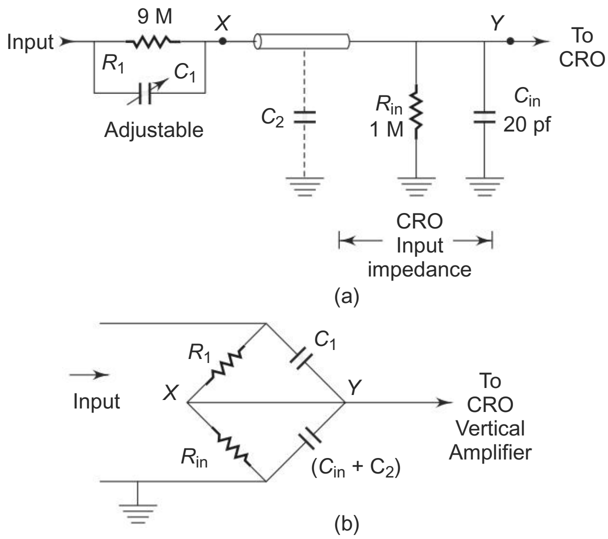


Fig. 7.45 (a) 10 : 1 Probe (b) Equivalent Circuit of 10 : 1 Probe

$$\therefore \qquad \frac{R_1}{\omega(C_2 + C_{in})} = \frac{R_{in}}{\omega C_1}$$

$$R_1 C_1 = R_{in}(C_2 + C_{in})$$

Therefore, X and Y are equipotential and the effect of the probe is equivalent to placing a potential divider consisting of R_1 and R_{in} across the input circuit. The attenuation of the signal is 10 : 1, i.e. $(R_1 + R_{in})/R_1 = 10 : 1$ over a wide frequency range. Therefore, it is called a compensated 10×1 probe. As far as dc voltage inputs are concerned, the coaxial capacitance equals 30 pf per foot. (Assuming a coaxial length of 3.5 ft, the total coaxial length capacitance is 105 pf). Substituting this value in the balance bridge equation, we have

$$C_1 = \frac{R_{in}(C_{in} + C_2)}{R_1} = \frac{1\text{ M}(105 + 20)\text{ pf}}{9\text{ M}} = 13.88\text{ pf}$$

Therefore, the input capacitance of a CRO can range from 15–50 pf. C_1 should be adjusted from 13–47 pf. It must be adjusted to obtain optimum frequency response from the probe-CRO combination. The C_1 adjustment is done by connecting the probe tip to a square wave of 1 kHz and observing the CRT display. When the CRT display has optimum response, the C_1 value is deemed to be appropriate.

Therefore $\qquad V_{out} = (0.1)V_{in} = \dfrac{V_{in} \times R_{in}}{R_1 + R_{in}}$

7.28.3 Active Probes

Active probes are designed to provide an efficient method of coupling high frequency, fast rise time signals to the CRO input. Usually active probes have very high input impedance, with less attenuation than passive probes. Active devices may be diodes, FETs, BJTs, etc.

Active probes are more expensive and bulky than passive probes, but they are useful for small signal measurements, because their attenuation is less.

Active Probes Using FETs Figure 7.46 shows a basic circuit of an active probe using a FET.

The FET is used as the active element to amplify the input signal. Although the voltage gain of the FET follower circuit shown is unity, the follower circuit provides a power gain so that the input impedance can be increased. To be effective the FET must be mounted directly in the voltage probe tip, so that the capacitance of the interconnecting cable can be eliminated. This requires that the power for the FET be supplied from the oscilloscope to the FET in the probe tip. The FET voltage follower drives a coaxial cable, but instead of the cable connecting directly to the high input impedance of the oscilloscope, it is terminated in its characteristic impedance.

There is no signal attenuation between the FET Amplifier and the probe tip. The range of the signals that can be handled by the FET probe is limited to the dynamic range of the FET amplifier and is typically less than a few volts. To handle a larger dynamic range, external attenuators are added at the probe tip. Active probes have limited use because the FET probe effectively becomes an FET attenuator. Therefore, oscilloscopes are typically used with a 10 to 1 attenuator probe.

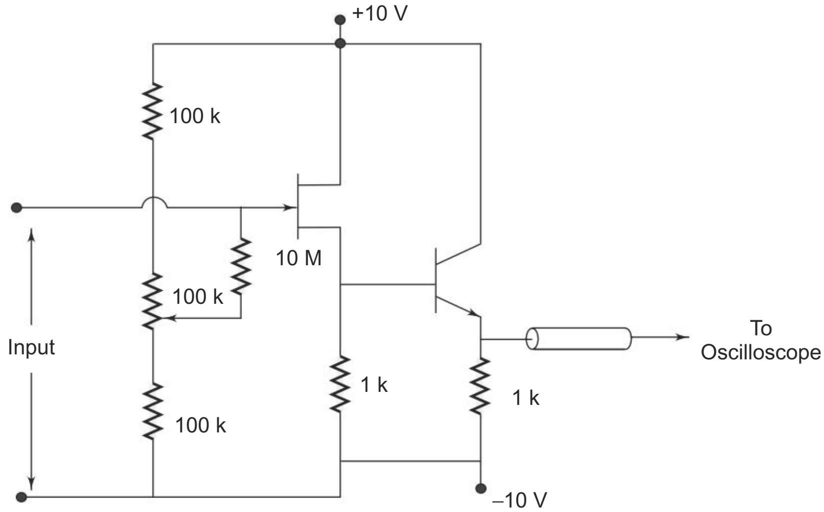


Fig. 7.46 FET probe

ATTENUATORS 7.29

Attenuators are designed to change the magnitude of the input signal seen at the input stage, while presenting a constant impedance on all ranges at the attenuator input.

A compensated RC attenuator is required to attenuate all frequencies equally. Without this compensation, HF signal measurements would always have to take the input circuit RC time constant into account.

The input attenuator must provide the correct 1-2-5 sequence while maintaining a constant input impedance, as well as maintain both the input impedance and attenuation over the frequency range for which the oscilloscope is designed.

7.29.1 Uncompensated Attenuators

The circuit diagram shown in Fig. 7.47 gives a resistive divider attenuator connected to an amplifier with a 10 pf input capacitance. If the input impedance of the amplifier is high, the input impedance of the attenuator is relatively constant, immaterial of the switch setting of the attenuator.

The input impedance, as seen by the amplifier, changes greatly depending on the setting of the attenuator. Because of this, the RC time constant and frequency response of the amplifier are dependent on the setting of the attenuator, which is an undesirable feature.

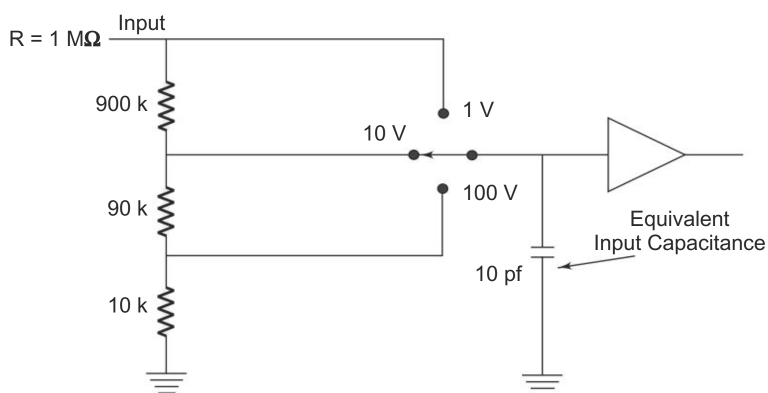


Fig. 7.47 Uncompensated attenuator

7.29.2 Simple Compensated Attenuator

The diagram in Fig. 7.48 shows an attenuator with both resistive and capacitive voltage dividers. The capacitive voltage dividers improve the HF response of the attenuator. This combination of capacitive and resistive voltage dividers is known as a compensated attenuator. For oscilloscopes where the frequency range extends to 100 MHz and beyond, more complex dividers are used.

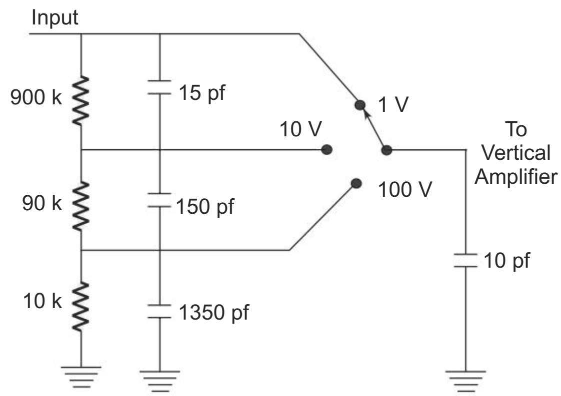


Fig. 7.48 Simple Compensated Attenuator

Figure 7.49 shows an attenuator divider between the input and output of the vertical deflection pre-amplifier. The input attenuator provides switching powers of 10, while attenuators at the output of the vertical preamplifier provides 1–2–5 attenuation.

Practically all oscilloscopes provide a switchable input coupling capacitor, as shown in Fig. 7.49.

The input impedance of an oscilloscope is 1 MΩ which is shunted with an input capacitance of 10–30 pf. If a probe were connected to the oscilloscope, the input impedance at the probe tip would have a greater capacitance because of the added capacitance of the probe assembly and of the connecting shielded cable. If it is desired for HF oscilloscopes to have an input capacitance of much less than 20–30 pf, an attenuator probe is used. Figure 7.48 shows a 10 to 1 attenuator probe connected to the input of the oscilloscope.

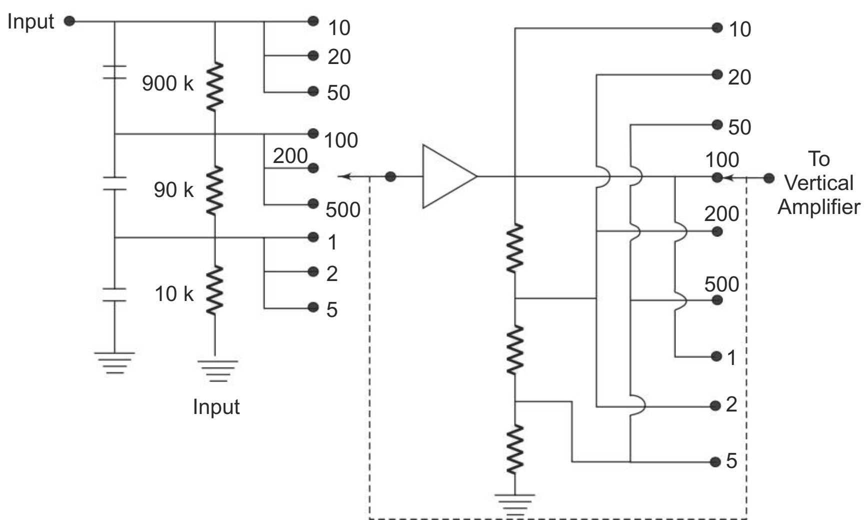


Fig. 7.49 Switchable Input Attenuator

Within the probe tip is a 9 MΩ resistor and shunted across this resistor is a capacitor. This capacitor is adjusted so that the ratio of the shunt capacitance to the series capacitance is exactly 10 to 1.

The attenuator probe, often called a 10 to 1 probe, provides an approximately 10 to 1 reduction in the input capacitance. However, it also gives a 10 to 1 reduction in overall oscilloscope sensitivity.

The input capacitance is not constant from one oscilloscope to another hence the probe is provided with an adjustable compensating capacitor. If the ratio of the series to shunt is not adjusted precisely to 10 to 1, the frequency response of the oscilloscope will be flat.

APPLICATIONS OF OSCILLOSCOPE 7.30

The range of applications of an oscilloscope varies from basic voltage measurements and waveform observation to highly specialised applications in all areas of science, engineering and technology.

7.30.1 Voltage Measurements

The most direct voltage measurement made with the help of an oscilloscope is the peak to peak (*p–p*) value. The rms value of the voltage can then be easily calculated from the *p–p* value.

To measure the voltage from the CRT display, one must observe the setting of the vertical attenuator expressed in V/div and the peak to peak deflection of the beam, i.e. the number of divisions. The peak to peak value of voltage is then computed as follows.

$$V_{p-p} = \left(\frac{\text{volts}}{\text{div}}\right) \times \left(\frac{\text{no. of div}}{1}\right)$$

Example 7.1 *The waveform shown in Fig. 7.50 is observed on the screen of an oscilloscope. If the vertical attenuation is set to 0.5 V/div, determine the peak to peak amplitude of the signal.*

Solution Using the equation,

$$V_{p-p} = \left(\frac{\text{volts}}{\text{div}}\right) \times \left(\frac{\text{no. of div}}{1}\right)$$

$$V_{p-p} = 0.5\ V \times 3 = 1.5\ V_{p-p}$$

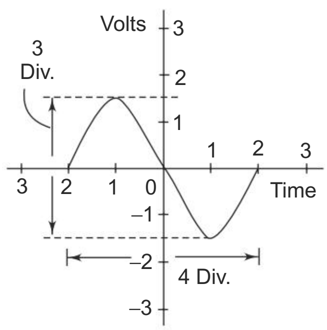


Fig.7.50

7.30.2 Period and Frequency Measurements

The period and frequency of periodic signals are easily measured with an oscilloscope. The waveform must be displayed such that a complete cycle is displayed on the CRT screen. Accuracy is generally improved if a single cycle displayed fills as much of the horizontal distance across the screen as possible.

The period is calculated as follows.

$$T = \left(\frac{\text{time}}{\text{div}}\right) \times \left(\frac{\text{No. of div}}{\text{cycle}}\right)$$

The frequency is then calculated as $f = 1/T$.

Example 7.2 *If the time/div control is set to 2µs/div when the waveform in Fig. Ex. 7.1 is displayed on the CRT screen, determine the frequency of the signal.*

Solution The period of the signal is calculated using the equation

$$T = \left(\frac{\text{time}}{\text{div}}\right) \times \left(\frac{\text{No. of div}}{\text{cycle}}\right)$$

$$T = 2\ \mu\text{s} \times 4 = 8\ \mu\text{s}$$

Hence frequency is calculated as

$$f = 1/T = 1/8\ \mu\text{s} = 125\ \text{kHz}$$

DELAYED SWEEP 7.31

Many oscilloscopes of laboratory quality include a delayed sweep feature. This feature increases the versatility of the instrument by making it possible to magnify a selected portion of an undelayed sweep, measure waveform jitter or rise time, and check pulse time modulation, as well as many other applications.

Delayed sweep is a technique that adds a precise amount of time between the trigger point and the beginning of the scope sweep. When the scope is being used in the sweep mode, the start of the horizontal sweep can be delayed, typically from a few µs to perhaps 10 seconds or more. Delayed sweep operation allows the user to view a small segment of the waveform, e.g. an oscillation or ringing that occurs during a small portion of a low frequency waveform.

The most common approaches used by oscilloscpe manufacturers for delayed sweep operations are, the following.

1. Normal triggering sweep after the desired time delay, which is set from the panel controls.
2. A Delay Plus Trigger mode, where a visual indication, such as light, indicates that the delay time has elapsed and the sweep is ready to be triggered.
3. Intensified sweep, where the delayed sweep acts as a positional magnifier.

DIGITAL STORAGE OSCILLOSCOPE (DSO) 7.32

Digital storage oscilloscope are available in processing and non-processing types. Processing types include built in computing power, which takes advantage of the fact that all data is already in digital form.

The inclusion of interfacing and a microprocessor provides a complete system for information acquisition, analysis and output. Processing capability ranges from simple functions (such as average, area, rms, etc.) to complete Fast Fourier Transform (FFT) spectrum analysis capability.

(Units with built in hard copy plotters are particularly useful, since they can serve as digital scope high speed recorders, tabular printers and *X*–*Y* plotters, all in one unit, with computing power and an 8½″ × 11″ paper/ink printout.)

Non-processing digital scopes are designed as replacements for analog instruments for both storage and non-storage types. Their many desirable features may lead to replace analog scopes entirely (within the Bandwidth range where digitization in feasible).

The basic principle of a digital scope is given in Fig. 7.51. The scope operating controls are designed such that all confusing details are placed on the back side and one appears to be using a conventional scope. However, some digital scope panels are simpler also, most digital scopes provide the facility of switching selectable to analog operation as one of the operating modes.

The basic advantage of digital operation is the storage capability, the stored waveform can be repetitively read out, thus making transients appear repetitively and allowing their convenient display on the scope screen. (The CRT used in digital storage is an ordinary CRT, not a storage type CRT.)

Furthermore, the voltage and time scales of display are easily changed after the waveform has been recorded, which allows expansion (typically to 64 times) of selected portions, to observe greater details.

A cross-hair cursor can be positioned at any desired point on the waveform and the voltage/time values displayed digitally on the screen, and/or readout electrically.

Some scopes use 12 bit converters, giving 0.025% resolution and 0.1% accuracy on voltage and time readings, which are better than the 2–5% of analog scopes.

Split screen capabilities (simultaneously displaying live analog traces and replayed stored ones) enable easy comparison of the two signals.

Pretrigger capability is also a significant advantage. The display of stored data is possible in both amplitude versus time and *X*-*Y* modes. In addition to the fast memory readout used for CRT display, a slow readout is possible for producing hard copy with external plotters.

When more memory than the basic amount (typically 4096 points/words) is needed, a magnetic disk accessory allows expansion to 32,000 points.

All digital storage scopes are limited in bandwidth by the speed of their A/D converters. However, 20 MHz digitizing rates available on some scopes yield a 5 MHz bandwidth, which is adequate for most applications.

Consider a single channnel of Fig. 7.51. The analog voltage input signal is digitised in a 10 bit A/D converter with a resolution of 0.1% (1 part in 1024) and frequency response of 25 kHz. The total digital memory storage capacity is 4096 for a single channel, 2048 for two channels each and 1024 for four channels each.

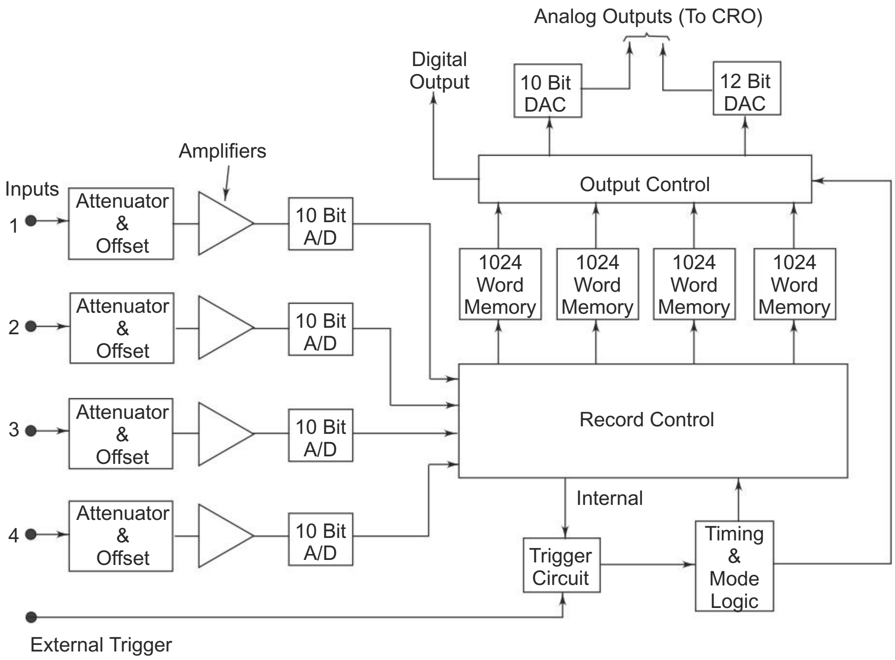


Fig. 7.51 Digital storage CRO

The analog input voltage is sampled at adjustable rates (up to 100,000 samples per second) and data points are read onto the memory. A maximum of 4096 points are storable in this particular instrument. (Sampling rate and memory size are selected to suit the duration and waveform of the physical event being recorded.)

Once the sampled record of the event is captured in memory, many useful manipulations are possible, since memory can be read out without being erased.

If the memory is read out rapidly and repetitively, an input event which was a single shot transient becomes a repetitive or continuous waveform that can be observed easily on an ordinary scope (not a storage scope). The digital memory also may be read directly (without going through DAC) to, say, a computer where a stored program can manipulate the data in almost any way desired.

Pre-triggering recording allows the input signal preceding the trigger points to be recorded. In ordinary triggering the recording process is started by the rise of the input (or some external triggering) above some preset threshold value.

As in digital recorder, DSO can be set to record continuously (new data coming into the memory pushes out old data, once memory is full), until the trigger signal is received; then the recording is stopped, thus freezing data received prior to the trigger signal in the memory.

An adjustable trigger delay allows operator control of the stop point, so that the trigger may occur near the beginning, middle or end of the stored information.

7.32.1 Commercial DSO

IE-522 shown in illustration Fig. 7.52, is a 25 MHz analog digital storage dual trace four display oscilloscope with computer interface.

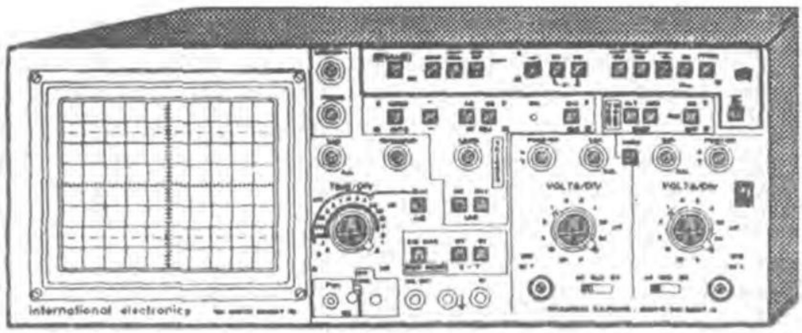


Fig. 7.52 25 MHz digital storage oscilloscope (IE-522) with computer interfacing [*Courtesy:* Inernational Electronics Ltd; Marketed: Signetics Electronics Ltd. Bombay]

The IE-522 Digital Storage Oscilloscopes (DSO) has the following features.

1. Sampling rate 20 Mega-samples per second per channel. Max. (simultaneous) capture of both channels.
2. Pre-trigger: 25%, 50%, 75%, for Single Shot, Roll normal.
3. Roll mode: (Continuous and Single Shot with Pre-trigger of 25%, 50%, 75%)
4. Single Shot (0.5 µs Single shot @ 10 pts./div resolution with pre-trigger 25%, 50%, 75%)
5. Digital Sweep rate: 0.5 µs/cm to 50 sec/cm, (event as long as 8.33 minutes can be captured)
6. Computer built in Interface: (RS 232 Serial port and Centronics Parallel interface).

FIBRE OPTIC CRT RECORDING OSCILLOSCOPE 7.33

The familiar CRT oscilloscope has an extremely HF response, but permanent records, normally photography of the screen, cannot be taken, since they are time limited to one sweep.

By combining a special fibre optic CRT with an oscillograph type paper drive (which passes the paper over the CRT face, where it is exposed by light from the CRT phosphor), a recording oscilloscope with uniquely useful characteristics is obtained. If the paper is held still, conventional CRO operation allows single sweep, i.e. amplitude versus time and X-Y recording of signals from dc to 1

MHz, much like the use of standard camera recording techniques, except that ordinary direct print oscillograph paper is used.

By employing a CRT sweep and simultaneous paper drive (speeds up to about 100 in/s can be selected), the 1 MHz frequency response is retained. Additionally, we can now get as many records as we wish, because they are simply stacked one above the other on the recording paper. This technique gives an equivalent paper speed of 40,000 in/s.

Time skew, resulting from paper motion during a single CRT sweep, is corrected electronically, but there will be small gaps in the data between sweep because of the CRT sweep retrace time.

Since the CRT allows beam intensity (*z*-axis) modulation, the instrument can produce grey-scale pictures, such as video images.

A multichannel version utilises sampling techniques to obtain direct current to 5 kHz response for up to 18 channels of data.

Digitest Lab The Digitest Lab (IE-549) shown in Fig. 7.53 has the following features,

1. 25 MHz Dual Trace Oscilloscope with rectangular CRT
2. Digital Storage Oscilloscope-5 Mega-samples/second Sampling Rate
3. Frequency Counter-70 MHz : will read internal, external and function generator frequency
4. 100 kHz Function Generator—Sine, Square, Triangular
5. Power Supply : + 12, – 12 V/0.2 A and 5 V/1 A
6. Component Tester: Test capacitors, Resistors, Diodes, etc.
7. Continuity Tester:
   Test for breaks/shorts in PCB tracks and cable harnesses.
8. Curve Tracer:

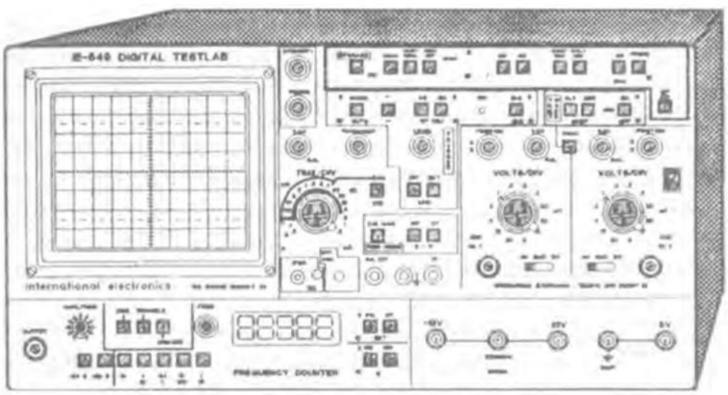


Fig. 7.53 8 in 1 digitest lab [IE-549]
[*Courtesy:* International Electronics Ltd.] [Marketed: Signetics Electronics Ltd.]

Test semiconductors like Transistors, FET's, Diodes, etc. You can check for selected parameters and obtain matched pairs of devices for production and service.

OSCILLOSCOPE OPERATING PRECAUTIONS 7.34

In addition to the general safety precautions, the following specific precautions should be observed when operating any type of oscilloscope. Most of the precautions also apply to recorders.

1. Always study the instruction manual of any oscilloscope with which you are not familiar even if you have had considerable experience with oscilloscopes.
2. Use the procedure of Section 7.35 to place the oscilloscope in operation. It is a good practice to go through the procedures each time the oscilloscope is used. This is especially true when the oscilloscope is used by other persons. The operator cannot be certain that position, focus and (especially) intensity control are at safe positions and the oscilloscope CRT could be damaged by switching on immediately.
3. As for any cathode ray tube device (such as a TV receiver), the CRT spot should be kept moving on the screen. If the spot must remain in one position, keep the intensity control as low as possible.
4. Always keep the minimum intensity necessary for good viewing.
5. If possible, avoid using the oscilloscope in direct sunlight or in a brightly lighted room. This will permit a low intensity setting. If the oscilloscope must be used in bright light, use the viewing hood.
6. Make all measurements in the centre area of the screen; even if the CRT is flat, there is a chance of reading errors caused by distortion at the edges.
7. Use only shielded probes. Never allow your fingers to slip down to the metal probe tip when the probe is in contact with a hot circuit.
8. Avoid operating an oscilloscope in a strong magnetic field. Such fields can causes distortion of the display. Most quality oscilloscopes are well shielded against magnetic interference. However, the face of the CRT is exposed and is subjected to magnetic interference.
9. Most oscilloscopes and their probes have some maximum input voltage specified in the instruction manual. Do not exceed this maximum value. Also, do not exceed the maximum line voltage or use a different power frequency.
10. Avoid operating the oscilloscope with the shield or case removed. Besides the danger of exposing high voltage circuits (several thousand volts are used on the CRT), there is the hazard of the CRTs imploding and scattering glass at high velocity.
11. Avoid vibration and mechanical shock. Like most electronic equipment, an oscilloscope is a delicate instrument.
12. If an internal fan or blower is used, make sure that it is operating. Keep ventilation air filters clean.
13. Do not attempt repair of an oscilloscope unless you are a qualified instrument technician. If you must adjust any internal circuits, follow the instruction manual.

14. Study the circuit under test before making any test connections. Try to match the capability of the oscilloscope to the circuit under test. For example, if the circuit has a range of measurements to be made (ac, dc, RF, pulse), you must use a wide-band DC oscilloscope, with a low capacitance probe and possibly a demodulator probe. Do not try to measure 3 MHz signals with a 100 kHz bandwidth oscilloscope. On the other hand, it is wasteful to use a dual trace 50 MHz laboratory oscilloscope to check out the audio sections of transistor radios.

PLACING AN OSCILLOSCOPE IN OPERATION 7.35

After the setup instruction of the oscilloscope manual have been digested, they can be compared with the following general or typical procedures.

1. Set the power switch to Off
2. Set the internal recurrent sweep to OFF
3. Set the focus, gain, intensity and sync controls to their lowest position (usually fully counterclockwise)
4. Set the sweep selector to External
5. Set the vertical and horizontal position controls to their approximate midpoint
6. Set the power switch to ON. It is assumed that the power cord has been connected.
7. After a suitable warmup period (as recommended by the manual) adjust the intensity control until the trace spot appears on the screen. If a spot is not visible at any setting of the intensity control, the spot is probably off screen (unless the oscillsocope is defective). If necessary, use the vertical and horizontal position controls to bring the spot into view. Always use the longest setting of the intensity control needed to see the spot, so as to prevent burning of the oscilloscope screen.
   It should be noted that dc oscilloscopes need longer warmup times than ac oscilloscopes because of drift problems associated with dc amplifiers.
8. Set the focus control for a sharp fine dot
9. Set the vertical and horizontal position controls to centre the spot on the screen
10. Set the sweep selector to Internal. This should be the linear internal sweep, if more than one internal sweep is available.
11. Set the internal recurrent sweep to ON. Set the sweep frequency to any frequency, or a recurrent rate higher than 100 Hz.
12. Adjust the horizontal gain control and check that the spot is expanded into a horizontal trace or line. The line length should be controllable by adjusting the horizontal gain control.
13. Return the horizontal gain control to zero, set the internal recurrent sweep to OFF.
14. Set the vertical gain control to its approximate midpoint, and touch the vertical input with you finger. The stray signal pickup should cause the

spot to be deflected vertically into a trace of line. Check that the line length is controllable by adjustment of the vertical gain control.

15. Return the vertical gain control to zero (or its lowest setting).
16. Set the internal recurrent sweep to ON. Advance the horizontal gain control to expand the spot into a horizontal line.
17. If required, connect a probe to the vertical input.
18. The oscilloscope should now be ready for immediate use. Depending on the test to be performed the oscilloscope may require calibration.

Review Questions

1. List the major components of a CRT.
2. What does the term phosphorescence mean?
3. Explain the principle of operation of a single beam CRO.
4. Draw the basic block diagram of an oscilloscope and state the functions of each block.
5. How is an electron beam focused onto a fine spot on the face of the CRT.
6. Why are the operating voltages of a CRT arranged so that the deflection plates are at nearly ground potential?
7. How is the vertical axis of an oscilloscope deflected? How does it differ from the horizontal axis?
8. How does the *X*-shift and *Y*-shift function.
9. Explain the function of a trigger circuit.
10. State the function of a delay line used in the vertical section of an oscilloscope.
11. List the various controls on the front panel of a CRO. State the function of various controls on the front panel of a CRO.
12. State the need of a time base generator.
13. Describe with a diagram the operation of a continuous sweep generator. List the drawbacks of a continuous sweep generator.
14. Explain with a diagram the operation of a triggered sweep generator.
15. List the advantages of using negative supply in a CRO.
16. Define intensity, focus and astigmatism.
17. Describe with a diagram the operation of a vertical amplifier. State the function of using FET stage in the vertical amplifier.
18. State the advantages of dual trace over dual beam?
19. Describe with diagram the operation of a dual beam CRO.
20. Explain the working principle of a dual trace CRO.
21. Describe with a diagram and waveforms the operation of a dual trace CRO in alternate and Chop mode. State the functions of each block.
22. How does alternate sweep compare with chopped sweep? When would one method be selected over the other?
23. State the function of the electronic switch. Explain with a diagram the working of an electronic switch.
24. Explain with a diagram the operation of a dual trace in X-Y mode.
25. Compare dual beam and dual trace CRO.
26. Explain with diagram the operation of a delayed sweep CRO. State the advantages of using delayed sweep CRO.

27. How does the sampling CRO increase the apparent frequency response of an oscilloscope.
28. Describe with diagram the operation of a sampling CRO. State the function of the staircase generator used in a sampling CRO.
29. Explain with a diagram the principle of analog storage CRO.
30. What is the speciality of a storage CRO.
31. Describe with a diagram the operation of an analog storage CRO.
32. State the advantages and disadvantages of a phosphor storage oscilloscope.
33. Describe with a block diagram the operation of a digital storage CRO. State the functions of each block.
34. Explain how frequency can be measured by a CRO using lissajous figures.
35. Explain with a diagram how frequency can be measured using spot wheel method. Explain with a diagram how frequency can be measured using a gear wheel method.
36. Compare the spot wheel method with that of the gear wheel method.
37. Explain with a diagram how CRO can be used to check diodes, inductors and capacitors.
38. State the standard specifications of a simple CRO.
39. State the function of a probe. State the function and explain with a diagram the operation of a 10:1 probe.
40. Compare passive probes with active probes.
41. State the advantages of using a probe.
42. State the function of attenuators in CRO.
43. What do you understand by compensation in attenuators? Explain with a diagram the operation of a simple compensated attenuator.
44 State the various applications of an oscilloscope.

Multiple Choice Questions

1. Post deflection acceleration is used to
   (a) enhance the intensity of the beam
   (b) focus the beam
   c) repel the electron beam
   (d) increase the velocity of the electron beam
2. Trigger pulses in the CRO are used
   (a) to generate high voltage required for the CRT
   (b) to synchronise the input with the time base generator
   (c) to synchronise the input and the vertical amplifier
   (d) to generate low voltages required for the CRT
3. The function of the sync section in CRO is
   (a) to match the horizontal sweep rate with the frequency of the vertical signal
   (b) to start the horizontal sweep at the same relative point on the vertical signal
   (c) to adjust the intensity control
   (d) to control the gain of the amplifier
4. The amplitude read on CRO set of 1 V/div is 1.5 cm on the vertical axis. The value of amplitude in V is
   (a) 1.5 V (b) 5 V
   (c) 1 V (d) 0.15 V
5. The distance between two peaks measured on the *X*-axis is 2 cm, at 1 ms/div. The frequency of the signal is
   (a) 50 Hz (b) 5 Hz
   (c) 1 kHz (d) 500 Hz

6. A dual beam CRO uses
   (a) electronic switch
   (b) two electron guns
   (c) one electron gun
   (d) two time base generator circuits
7. A dual trace CRO uses
   (a) one electron gun
   (b) two electron guns
   (c) two pairs of VDPs
   (d) two pairs of HDPs
8. An electronic switch is used in a
   (a) single beam CRO
   (b) dual beam CRO
   (c) dual Trace CRO
   (d) sampling CRO
9. A sampling CRO is used for
   (a) HF (b) VLF (c) VHF (d) LF
10. An analog storage CRO is used for displaying waveforms in the frequency range of
    (a) VHF (b) VLF (c) HF (d) LF

Practice Problems

1. A CRO with a sensitivity of 5 V/cm is used. An ac voltage is applied to the y-input. A 10 cm long straight line is observed. Determine the ac voltage.
2. The Lissajous pattern on an CRO is stationary and has five horizontal and two vertical tangencies. The frequency of the horizontal input is 1000 Hz. Determine the frequency of vertical input.
3. A CRO is set to a time base of 0.1 ms/cm with a 10 cm amplitude. Sketch the display of the pulse signal waveform with a pulse repetition rate of 2000 Hz and a duty cycle of 25%.

Further Reading

1. Refuse P. Turner, Practical Oscilloscope Handbook, Vol. 2, D.B. Taraporewala Sons & Co., 1985.
2. John D. Lenk, Handbook of Oscilloscope: Theory and Application, Prentice-Hall of India, 1968.
3. John D. Lenk, Handbook of Electronic Meters (Theory and Applications), D.B. Taraporewala Sons & Co., Prentice-Hall, 1980)
4. Oliver Cage, Electronic Measurements & Instrumentation, McGraw-Hill, 1975.
5. Rudolf F. Graf, ABCs of Electronic Test Probes, Howard.W.Sams & Co., 1966.
6. Vestor Robinson, Handbook of Electronic Instrumentation Testing & Troubleshooting, D.B. Taraporewala Sons & C0., 1979.

Chapter 8
Signal Generators

INTRODUCTION 8.1

A signal generator is a vital component in a test setup, and in electronic troubleshooting and development, whether on a service bench or in a research laboratory. Signal generators have a variety of applications, such as checking the stage gain, frequency response, and alignment in receivers and in a wide range of other electronic equipment.

They provide a variety of waveforms for testing electronic circuits, usually at low powers. The term oscillator is used to describe an instrument that provides only a sinusoidal output signal, and the term generator to describe an instrument that provides several output waveforms, including sine wave, square wave, triangular wave and pulse trains, as well as an amplitude modulated waveform. Hence, when we say that the oscillator generates a signal, it is important to note that no energy is created; it is simply converted from a dc source into ac energy at some specific frequency.

There are various types of signal generator but several requirements are common to all types.

1. The frequency of the signal should be known and stable.
2. The amplitude should be controllable from very small to relatively large values.
3. Finally, the signal should be distortion-free

The above mentioned requirements vary for special generators, such as function generators, pulse, and sweep generators.

Various kinds of signals, at both audio and radio frequencies, are required at various times in an instrumentation system. In most cases a particular signal required by the instrument is internally generated by a self-contained oscillator. The oscillator circuit commonly appears in a fixed frequency form (e.g. when it provides a 1000 c/s excitation source for an ac bridge). In other cases, such as in a *Q*-meter, oscillators in the form of a variable frequency arrangement for covering *Q*-measurements over a wide range of frequencies, from a few 100 kHz to the MHz range, are used.

In contrast with self-contained oscillators that generate only the specific signals required by the instrument, the class of generators that are available as separate instruments to provide signals for general test purposes are usually designated as signal generators. These AF and RF generators are designed to provide extensive and continuous coverage over as wide a range of frequencies as is practical.

In RF signal generators, additional provision is generally made to modulate the continuous wave signal to provide a modulated RF signal. The frequency band limits are listed in Table 8.1.

Table 8.1

Band	*Approximate Range*		
AF	20 Hz	–	20 kHz
RF	above 30 kHz		
VLF – Very Low Frequency	15	–	100 kHz
LF – Low Frequency	100	–	500 kHz
Broadcast	0.5	–	1.5 MHz
Video	DC	–	5 MHz
HF	1.5	–	30 MHz
VHF	30	–	300 MHz
UHF	300	–	3000 MHz
Microwave	beyond 3000 MHz (3 GHz)		

Most of the service type AF generators commonly cover from 20 Hz to 200 kHz, which is far beyond the AF range.

In more advanced laboratory types of AF generators, the frequency range extends quite a bit further e.g. a Hewlett Packard model covers 5 Hz – 600 kHz and a Marconi model generates both sine and square waves and has a very wide range of 10 Hz – 10 MHz.

FIXED FREQUENCY AF OSCILLATOR 8.2

In many cases, a self-contained oscillator circuit is an integral part of the instrument circuitry and is used to generate a signal at some specified audio frequency. Such a fixed frequency might be a 400 Hz signal used for audio testing or a 1000 Hz signal for exciting a bridge circuit.

Oscillations at specified audio frequencies are easily generated by the use of an iron core transformer to obtain positive feedback through inductive coupling between the primary and secondary windings.

VARIABLE AF OSCILLATOR 8.3

A variable AF oscillator for general purpose use in a laboratory should cover atleast the full range of audibility (20 Hz to 20 kHz) and should have a fairly constant pure sinusoidal wave output over the entire frequency range.

Hence, variable frequency AF generators for laboratory use are of the *RC* feedback oscillator type or Beat Frequency Oscillator (BFO) type.

BASIC STANDARD SIGNAL GENERATOR (SINE WAVE) 8.4

The sine wave generator represents the largest single category of signal generator. This instrument covers a frequency range from a few Hertz to many Giga-Hertz. The sine wave generator in its simplest form is given in Fig. 8.1.

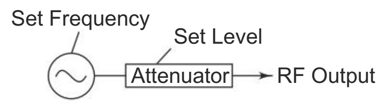


Fig. 8.1 Basic sine wave generator

The simple sine wave generator consists of two basic blocks, an oscillator and an attenuator. The performance of the generator depends on the success of these two main parts. The accuracy of the frequency, stability, and freedom from distortion depend on the design of the oscillator, while the amplitude depends on the design of the attenuator.

STANDARD SIGNAL GENERATOR 8.5

A standard signal generator produces known and controllable voltages. It is used as power source for the measurement of gain, signal to noise ratio (S/N), bandwidth, standing wave ratio and other properties. It is extensively used in the testing of radio receivers and transmitters.

The instrument is provided with a means of modulating the carrier frequency, which is indicated by the dial setting on the front panel. The modulation is indicated by a meter. The output signal can be Amplitude Modulated (AM) or Frequency Modulated (FM). Modulation may be done by a sine wave, square wave, triangular wave or a pulse. The elements of a conventional signal generator are shown in Fig. 8.2 (a).

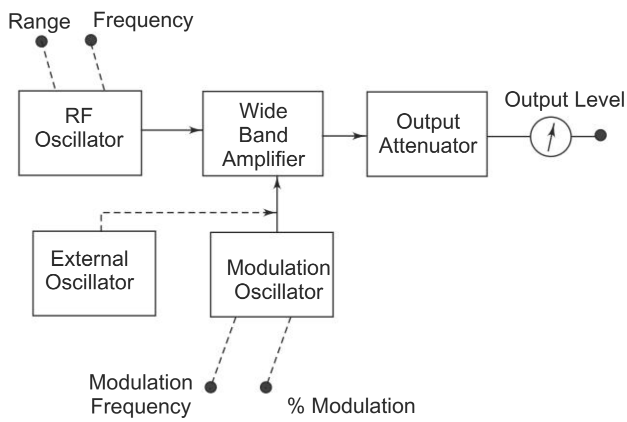


Fig. 8.2 (a) Conventional standard signal generator

The carrier frequency is generated by a very stable RF oscillator using an LC tank circuit, having a constant output over any frequency range. The frequency of oscillations is indicated by the frequency range control and the vernier dial setting. AM is provided by an internal sine wave generator or from an external source.

(Modulation is done in the output amplifier circuit. This amplifier delivers its output, that is, modulation carrier, to an attenuator. The output voltage is read by an output meter and the attenuator output setting.)

Frequency stability is limited by the LC tank circuit design of the master oscillator. Since range switching is usually accomplished by selecting appropriate capacitors, any change in frequency range upsets the circuit design to some extent and the instrument must be given time to stabilise at the new resonant frequency.

In high frequency oscillators, it is essential to isolate the oscillator circuit from the output circuit. This isolation is necessary, so that changes occurring in the output circuit do not affect the oscillator frequency, amplitude and distortion characteristics. Buffer amplifiers are used for this purpose.

Figure 8.2(b) illustrates a commercial AM/FM signal generator (IE900A) having a frequency range from 100 kHz – 110 MHz, along with a frequency counter and TTL output.

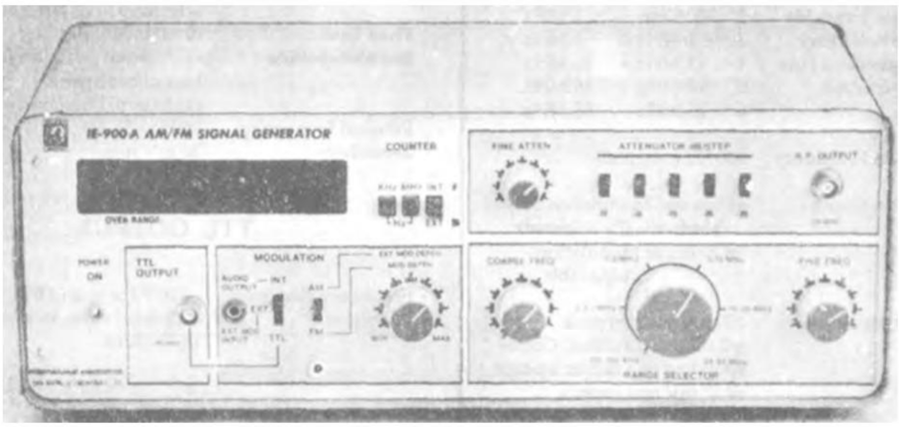


Fig. 8.2 (b) Commercial 110 MHz AM/FM Signal Generator (IE900A) along with a Frequency Counter [Courtesy: International Electronics Ltd., Bombay]

MODERN LABORATORY SIGNAL GENERATOR 8.6

To improve the frequency stability, a single master oscillator is optimally designed for the highest frequency range and frequency dividers are switched in to produce lower ranges. In this manner the stability of the top range is imparted to all the lower ranges.

The master oscillator is made insensitive to temperature variations and also to the influence of the succeeding stages by careful circuit design. Temperature compensation devices are used for any temperature changes. The block diagram of the modern standard signal generator is given in Fig. 8.3.

The highest frequency range of 34 – 80 MHz, is passed through B_1, an untuned buffer amplifier. B_2 and B_3 are additional buffer amplifiers and A is the main amplifier. The lowest frequency range produced by the cascaded frequency divider (9 frequency dividers of 2:1 ratio are used), is the highest frequency range divided by 512, or 2^9, or 67 – 156 kHz. Thus, the frequency stability of the highest range is imparted to the lower frequency ranges.

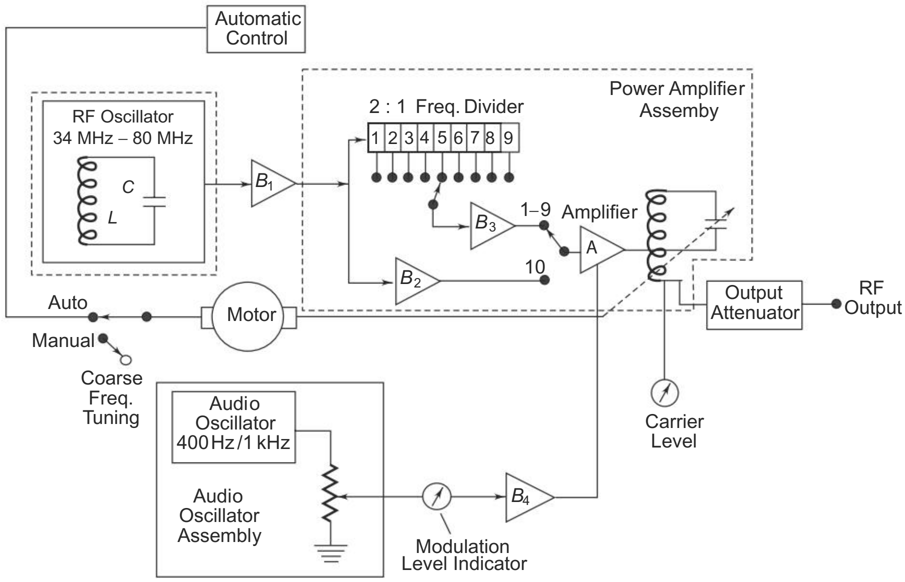


Fig. 8.3 Modern signal generator

The use of buffer amplifiers provides a very high degree of isolation between the master oscillator and the power amplifier, and almost eliminates all the frequency effects (distortion) between the input and output circuits, caused by loading.

Range switching effects are also eliminated, since the same oscillator is used on all bands. The master oscillator is tuned by a motor driven variable capacitor. For fast coarse tuning, a rocker switch is provided, which sends the indicator gliding along the slide rule scale of the main frequency dial at approximately 7% frequency changes per second. The oscillator can then be fine tuned by means of a large rotary switch (control), with each division corresponding to 0.01% of the main dial setting.

The master oscillator has both automatic and manual controllers. The availability of the motor driven frequency control is employed for programmable automatic frequency control devices.

Internal calibration is provided by the 1 MHz crystal oscillator. The small power consumption of the instruments makes it relatively easy to obtain excellent regulation and Q stability with very low ripple. The supply voltage of the master oscillator is regulated by a temperature compensated reference circuit.

The modulation is done at the power amplifier stage. For modulation, two internally generated signals are used, that is, 400 Hz and 1 kHz. The modulation level may be adjusted up to 95% by a control device. Flip-flops can be used as frequency dividers to get a ratio of 2:1.

AF SINE AND SQUARE WAVE GENERATOR 8.7

The block diagram of an AF Sine-Square wave audio oscillator is illustrated in Fig. 8.4.

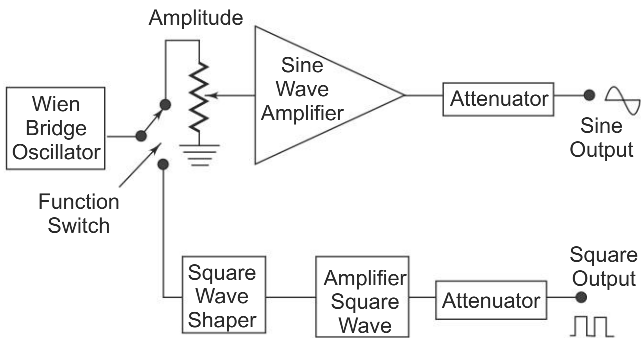


Fig. 8.4 AF sine and square wave generator

The signal generator is called an oscillator. A Wien bridge oscillator is used in this generator. The Wien bridge oscillator is the best for the audio frequency range. The frequency of oscillations can be changed by varying the capacitance in the oscillator. The frequency can also be changed in steps by switching in resistors of different values.

The output of the Wien bridge oscillator goes to the function switch. The function switch directs the oscillator output either to the sine wave amplifier or to the square wave shaper. At the output, we get either a square or sine wave. The output is varied by means of an attenuator.

The instrument generates a frequency ranging from 10 Hz to 1 MHz, continuously variable in 5 decades with overlapping ranges. The output sine wave amplitude can be varied from 5 mV to 5 V (rms).The output is taken through a push-pull amplifier. For low output, the impedance is 600Ω. The square wave amplitudes can be varied from 0 – 20 V (peak). It is possible to adjust the symmetry of the square wave from 30 – 70%. The instrument requires only 7 W of power at 220 V – 50 Hz.

The front panel of a signal generator consists of the following.

1. *Frequency selector* It selects the frequency in different ranges and varies it continuously in a ratio of 1 : 11. The scale is non-linear.
2. *Frequency multiplier* It selects the frequency range over 5 decades, from 10 Hz to 1 MHz.
3. *Amplitude multiplier* It attenuates the sine wave in 3 decades, × 1, × 0.1 and × 0.01.
4. *Variable amplitude* It attenuates the sine wave amplitude continuously.
5. *Symmetry control* It varies the symmetry of the square wave from 30% to 70%.
6. *Amplitude* It attenuates the square wave output continuously.
7. *Function switch* It selects either sine wave or square wave output.

8. *Output available* This provides sine wave or square wave output.
9. *Sync* This terminal is used to provide synchronisation of the internal signal with an external signal.
10. On-Off Switch

FUNCTION GENERATOR 8.8

A function generator produces different waveforms of adjustable frequency. The common output waveforms are the sine, square, triangular and sawtooth waves. The frequency may be adjusted, from a fraction of a Hertz to several hundred kHz.

The various outputs of the generator can be made available at the same time. For example, the generator can provide a square wave to test the linearity of an amplifier and simultaneously provide a sawtooth to drive the horizontal deflection amplifier of the CRO to provide a visual display.

Capability of Phase Lock The function generator can be phase locked to an external source. One function generator can be used to lock a second function generator, and the two output signals can be displaced in phase by adjustable amount.

In addition, the fundamental frequency of one generator can be phase locked to a harmonic of another generator, by adjusting the amplitude and phase of the harmonic, almost any waveform can be generated by addition.

The function generator can also be phase locked to a frequency standard and all its output waveforms will then have the same accuracy and stability as the standard source.

The block diagram of a function generator is illustrated in Fig. 8.5. Usually the frequency is controlled by varying the capacitor in the LC or RC circuit. In this instrument the frequency is controlled by varying the magnitude of current which drives the integrator. The instrument produces sine, triangular and square waves with a frequency range of 0.01 Hz to 100 kHz.

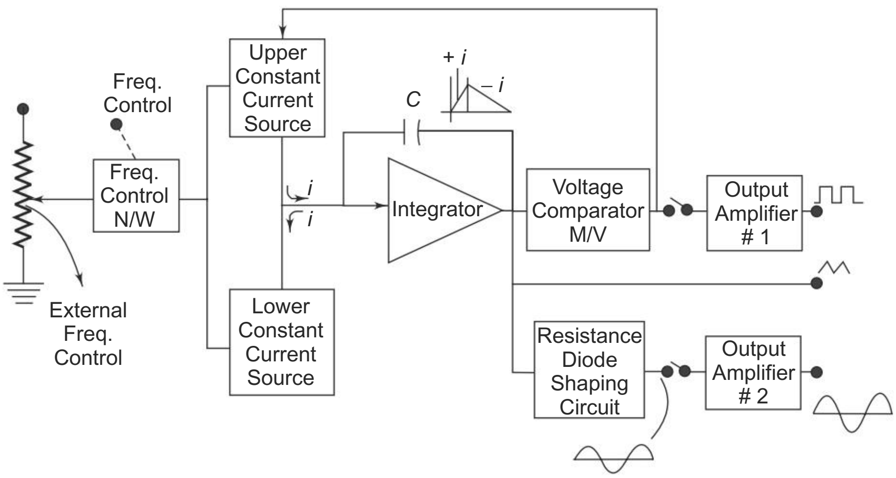


Fig. 8.5 Function generator

The frequency controlled voltage regulates two current sources. The upper current source supplies constant current to the integrator whose output voltage increases linearly with time, according to the equation of the output signal voltage.

$$e_{out} = -\frac{1}{C}\int_0^t i\,dt$$

An increase or decrease in the current increases or decreases the slope of the output voltage and hence controls the frequency.

The voltage comparator multivibrator changes states at a pre-determined maximum level of the integrator output voltage. This change cuts off the upper current supply and switches on the lower current supply.

The lower current source supplies a reverse current to the integrator, so that its output decreases linearly with time. When the output reaches a pre-determined minimum level, the voltage comparator again changes state and switches on the upper current source.

The output of the integrator is a triangular waveform whose frequency is determined by the magnitude of the current supplied by the constant current sources.

The comparator output delivers a square wave voltage of the same frequency. The resistance diode network alters the slope of the triangular wave as its amplitude changes and produces a sine wave with less than 1% distortion.

SQUARE AND PULSE GENERATOR (LABORATORY TYPE) 8.9

These generators are used as measuring devices in combination with a CRO. They provide both quantitative and qualitative information of the system under test. They are made use of in transient response testing of amplifiers. The fundamental difference between a pulse generator and a square wave generator is in the duty cycle.

$$\text{Duty cycle} = \frac{\text{pulse width}}{\text{pulse period}}$$

A square wave generator has a 50% duty cycle.

8.9.1 Requirements of a Pulse

1. The pulse should have minimum distortion, so that any distortion, in the display is solely due to the circuit under test.
2. The basic characteristics of the pulse are rise time, overshoot, ringing, sag, and undershoot.
3. The pulse should have sufficient maximum amplitude, if appreciable output power is required by the test circuit, e.g. for magnetic core memory. At the same time, the attenuation range should be adequate to produce small amplitude pulses to prevent over driving of some test circuit.

4. The range of frequency control of the pulse repetition rate (PRR) should meet the needs of the experiment. For example, a repetition frequency of 100 MHz is required for testing fast circuits. Other generators have a pulse-burst feature which allows a train of pulses rather than a continuous output.
5. Some pulse generators can be triggered by an externally applied trigger signal; conversely, pulse generators can be used to produce trigger signals, when this output is passed through a differentiator circuit.
6. The output impedance of the pulse generator is another important consideration. In a fast pulse system, the generator should be matched to the cable and the cable to the test circuit. A mismatch would cause energy to be reflected back to the generator by the test circuit, and this may be re-reflected by the generator, causing distortion of the pulses.
7. DC coupling of the output circuit is needed, when dc bias level is to be maintained.

The basic circuit for pulse generation is the asymmetrical multi-vibrator. A laboratory type square wave and pulse generator is shown in Fig. 8.6.

The frequency range of the instrument is covered in seven decade steps from 1 Hz to 10 MHz, with a linearily calibrated dial for continuous adjustment on all ranges.

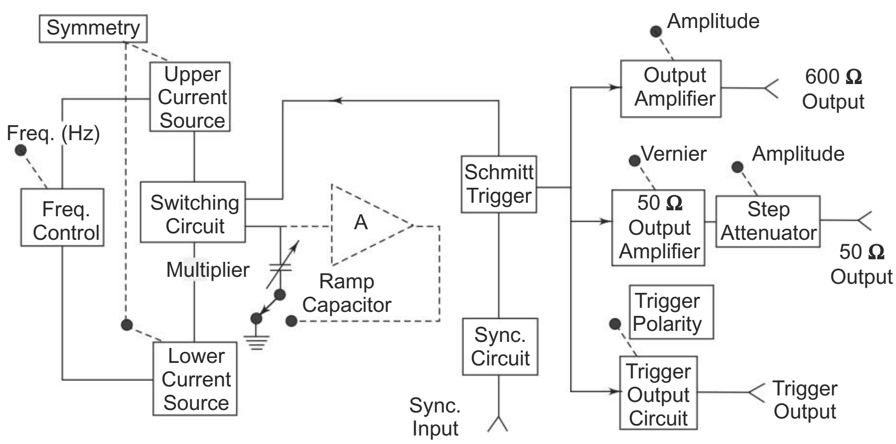


Fig. 8.6 Block diagram of a pulse generator

The duty cycle can be varied from 25 – 75%. Two independent outputs are available, a 50 Ω source that supplies pulses with a rise and fall time of 5 ns at 5 V peak amplitude and a 600 Ω source which supplies pulses with a rise and fall time of 70 ns at 30 V peak amplitude. The instrument can be operated as a free-running generator, or it can be synchronised with external signals.

The basic generating loop consists of the current sources, the ramp capacitor, the Schmitt trigger and the current switching circuit, as shown in Fig. 8.7.

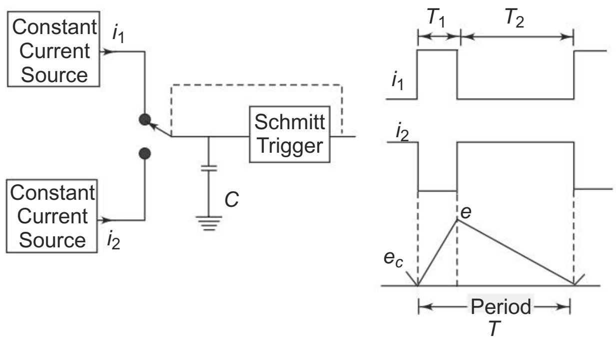


Fig. 8.7 Basic generating loop

The upper current source supplies a constant current to the capacitor and the capacitor voltage increases linearly. When the positive slope of the ramp voltage reaches the upper limit set by the internal circuit components, the Schmitt trigger changes state. The trigger circuit output becomes negative and reverses the condition of the current switch. The capacitor discharges linearly, controlled by the lower current source. When the negative ramp reaches a predetermined lower level, the Schmitt trigger switches back to its original state. The entire process is then repeated. The ratio i_1/i_2 determines the duty cycle, and is controlled by symmetry control. The sum of i_1 and i_2 determines the frequency. The size of the capacitor is selected by the multiplier switch.

The unit is powered by an internal supply that provides regulated voltages for all stages of the instrument.

RANDOM NOISE GENERATOR 8.10

A simplified block diagram used in the audio frequency range is shown in Fig. 8.8.

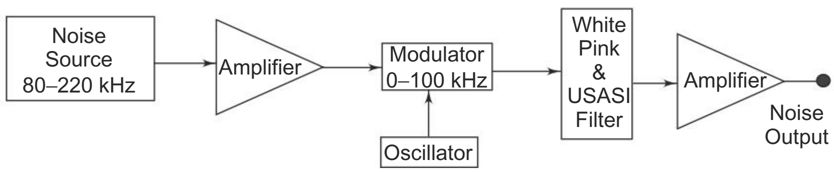


Fig. 8.8 Random noise generator

The instrument offers the possibility of using a single measurement to indicate performance over a wide frequency band, instead of many measurements at one frequency at a time. The spectrum of random noise covers all frequencies and is referred to as White noise, i.e. noise having equal power density at all frequencies (an analogy is white light). The power density spectrum tells us how the energy of a signal is distributed in frequency, but it does not specify the signal uniquely, nor does it tell us very much about how the amplitude of the signal varies with time. The spectrum does not specify the signal uniquely because it contains no phase information.

The method of generating noise is usually to use a semi conductor noise diode, which delivers frequencies in a band roughly extending from 80 – 220 kHz. The output from the noise diode is amplified and heterodyned down to the audio frequency band by means of a balanced symmetrical modulator. The filter arrangement controls the bandwidth and supplies an output signal in three spectrum choices, white noise, pink noise and Usasi noise.

From Fig. 8.9, it is seen that white noise is flat from 20 Hz to 25 kHz and has an upper cutoff frequency of 50 kHz with a cutoff slope of –12 dbs/octave.

Pink noise is so called because the lower frequencies have a larger amplitude, similar to red light. Pink noise has a voltage spectrum which is inversely proportional to the square root of frequency and is used in bandwidth analysis.

Usasi noise ranging simulates the energy distribution of speech and music frequencies and is used for testing audio amplifiers and loud speakers.

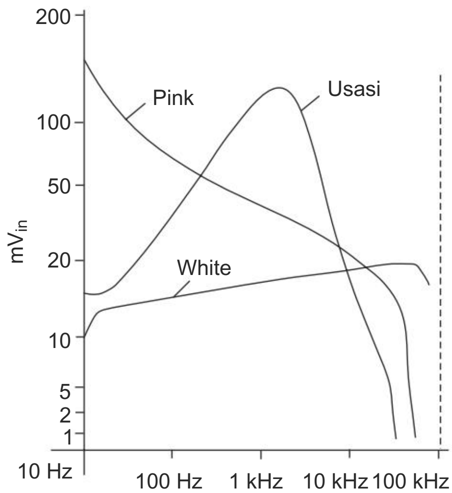


Fig. 8.9 Random noise generator

SWEEP GENERATOR 8.11

It provides a sinusoidal output voltage whose frequency varies smoothly and continuously over an entire frequency band, usually at an audio rate. The process of frequency modulation may be accomplished electronically or mechanically.

It is done electronically by using the modulating voltage to vary the reactance of the oscillator tank circuit component, and mechanically by means of a motor driven capacitor, as provided for in a modern laboratory type signal generator. Figure 8.10 shows a basic block diagram of a sweep generator.

The frequency sweeper provides a variable modulating voltage which causes the capacitance of the master oscillator to vary. A representative sweep rate could be of the order of 20 sweeps/second. A manual control allows independent adjustment of the oscillator resonant frequency.

The frequency sweeper provides a varying sweep voltage for synchronisation to drive the horizontal deflection plates of the CRO. Thus the amplitude of the response of a test device will be locked and displayed on the screen.

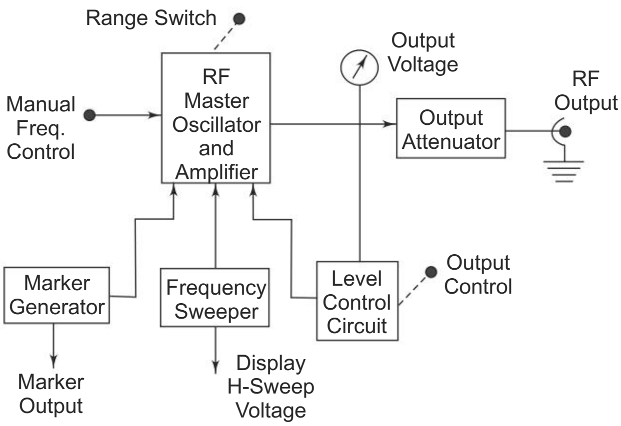


Fig. 8.10 Sweep generator

To identify a frequency interval, a marker generator provides half sinusoidal waveforms at any frequency within the sweep range. The marker voltage can be added to the sweep voltage of the CRO during alternate cycles of the sweep voltage, and appears superimposed on the response curve.

The automatic level control circuit is a closed loop feedback system which monitors the RF level at some point in the measurement system. This circuit holds the power delivered to the load or test circuit constant and independent of frequency and impedance changes. A constant power level prevents any source mismatch and also provides a constant readout calibration with frequency.

TV SWEEP GENERATOR 8.12

An RF generator, when used for alignment and testing of the RF and IF stages of a TV receiver, permits recording of circuit performance at one frequency at a time. Therefore, plotting the total response curve point by point over the entire channel bandwidth becomes a laborious process and takes a long time. To overcome this difficulty, a special RF generator, known as a sweep generator, is used. It delivers RF output voltage at a constant amplitude which sweeps across a range of frequencies and continuously repeats at a predetermined rate.

The sweep generator is designed to cover the entire VHF and UHF range. Any frequency can be selected as the centre frequency by a dial on the front panel of the instrument.

Frequency sweep is obtained by connecting a varactor diode across the HF oscillator circuits. A modified triangular voltage at 50 Hz is used to drive the varactor diode. Thus the frequency sweeps on either side of the oscillator centre frequency, at the rate of driving voltage frequency. The amplitude of the driving voltage applied across the varactor diodes can be varied to control maximum frequency deviation on either side of the carrier frequency. This is known as the sweep width and can be adjusted to the desired value, up to a maximum of about ± 15 MHz. A width control is provided for this procedure.

Alignment Procedure The output of the sweep generator is connected to the input terminals of the tuned circuit under test.

The frequency and sweep width dials are adjusted to a sweep range which lies in the pass-band of the circuit. With an input signal of constant amplitude, the output voltage varies in accordance with the frequency gain characteristics of the circuit. The magnitude of the output voltage varies with time as the oscillator frequency sweeps back and forth through the centre frequency. The RF output is detected either by a video detector in the receiver or by a demodulator probe. The detected output varies at the sweep rate of 50 Hz and its instantaneous amplitude changes in accordance with the circuit characteristics. Thus the output signal is a low frequency signal at 50 Hz and can be displayed on an ordinary scope, provided the low frequency response of the vertical amplifier is flat, down to atleast 50 Hz.

In order to obtain a linear display of the detected signal on the CRO, the scope sweep voltage must vary accordingly and have a frequency of 50 Hz.

The basic circuit configuration used for the alignment of any tuned circuit with a sweep generator and scope is shown in Fig. 8.11.

The time base switch of the CRO is set on External and its horizontal input terminals are connected to the sweep output on the sweep generator. Thus, the application of a 50 Hz triangular voltage to the horizontal deflecting plates results in a linear display on the scope screen, both during trace and retrace periods.

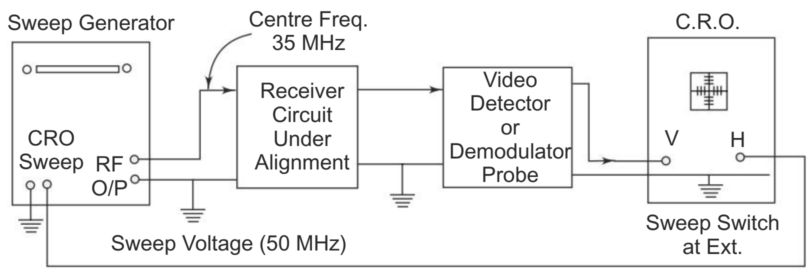


Fig. 8.11 Test Equipment connections for alignment of RF/IF sections of the receiver

MARKER GENERATOR 8.13

The sweep generator provides a visual display of the characteristics of the circuit or amplifier, but this is inadequate because it does not give any precise information of the frequency on the traced curve. For this, a separate RF generator, known as the Marker generator is used. This generator though essentially an RF signal generator in the VHF and UHF bands, is of much higher accuracy than other signal generators.

The output of the Marker generator is set at the desired frequency within the pass-band of the circuit under test.

As given in Fig. 8.12, the outputs from two generators are mixed together before being applied to the input terminals of the circuit under test.

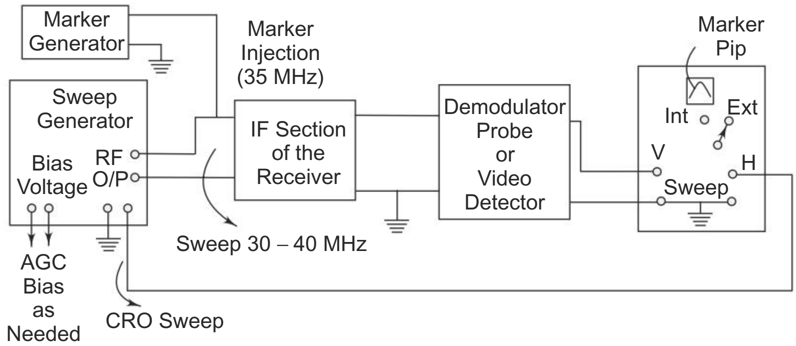


Fig. 8.12 Marker generator

The two signals are heterodyned to produce outputs at the sum and difference of two frequencies. These appear at the detector output and represent the amplitude characteristics of the circuit, because of the limited frequency response of the CRO vertical amplifier, the sum and difference frequency signals are produced when the sweep frequency varies over a wide range. When no difference frequency is produced, no vertical deflection is produced on the screen. However, as the frequency approaches and just crosses the marker frequency, the difference frequency signals which are produced, lie within the pass-band of the vertical amplifier. Thus, they produce a pip on the screen along the trace generated by the low frequency output of the detector. In order to get a sharp pip, a suitable capacitor is shunted across the vertical input terminals of the scope.

Since the pip is produced at the marker frequency, it can be shifted to any point on the response curve by varying the marker generator frequency.

The marker generator, besides providing dial controlled frequency, often has a provision to generate crystal controlled fixed frequency outputs at several important frequencies.

The additional frequency in a generator designed for a CCIR 625 lines system consists of the following.

31.4 MHz (Band-edge)
31.9 MHz (Trap-frequency)
33.4 MHz (Sound IF)
34.47 MHz (Colour IF)
36.15 MHz (Band centre)
38.9 MHz (Picture IF)
40.4 MHz (Trap frequency)
41.4 MHz (Band-edge)

These birdy-type markers can be switched in, either individually or simultaneously by the use of toggle switches on the front panel.

SWEEP-MARKER GENERATOR 8.14

Sweep and Marker generators were earlier manufactured as separate units. Now the two instruments are combined into a single instrument known as a Sweep-Marker generator.

This includes a built-in Marker adder for post-injection of the marker signal. Also provided is a variable dc bias source for feeding a fixed bias to the RF and IF sections of the receiver.

The frequency spectrum covered is in the VHF and UHF band ranges in both manual and auto modes. The RF output is about 0.5 V rms across a 75 Ω with a continuously variable attenuator, up to 50 db or 0 – 60 db in 10 db steps. The marker section provides crystal controlled output at all important frequencies. The output can be internally modulated with a 1 kHz tone when necessary. A demondulator probe, a balun and special cables with matched terminations are the additional accessories.

WOBBLUSCOPE 8.15

This instrument combines a Sweep generator, a Marker generator, and an oscilloscope, as shown in Fig. 8.13. It is a very useful unit for the alignment of RF, IF and video sections of a TV receiver. It may not have all the features of a high quality sweep generator but is an economical and compact piece of equipment specially designed for TV servicing. The oscilloscope usually has a provision for TV-V (vertical) and TV-H (horizontal) sweep modes. An RF output, down to 1 MHz, is available for video amplifier testing.

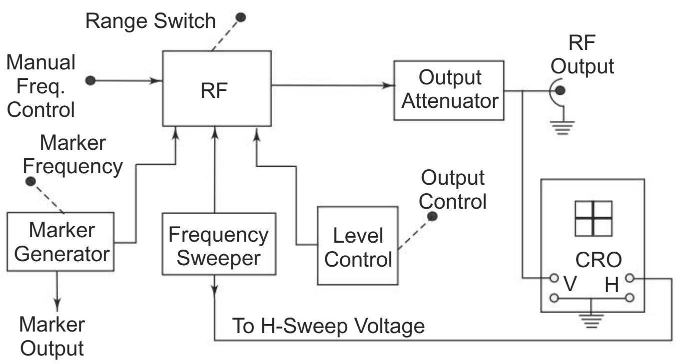


Fig. 8.13 Basic block diagram of a wobbluscope

VIDEO PATTERN GENERATOR 8.16

A pattern generator provides video signals directly, and with RF modulation, on standard TV channels for alignment, testing and servicing of TV receivers. The output signal is designed to produce simple geometric patterns like vertical and horizontal bars, checkerboard, cross-hatch, dots, etc.

These patterns are used for linearity and video amplifier adjustment. In addition to this, an FM sound signal is also provided in pattern generators for aligning sound sections of the receiver.

A simplified functional block diagram of a pattern cum sound signal generator is shown in Fig. 8.14.

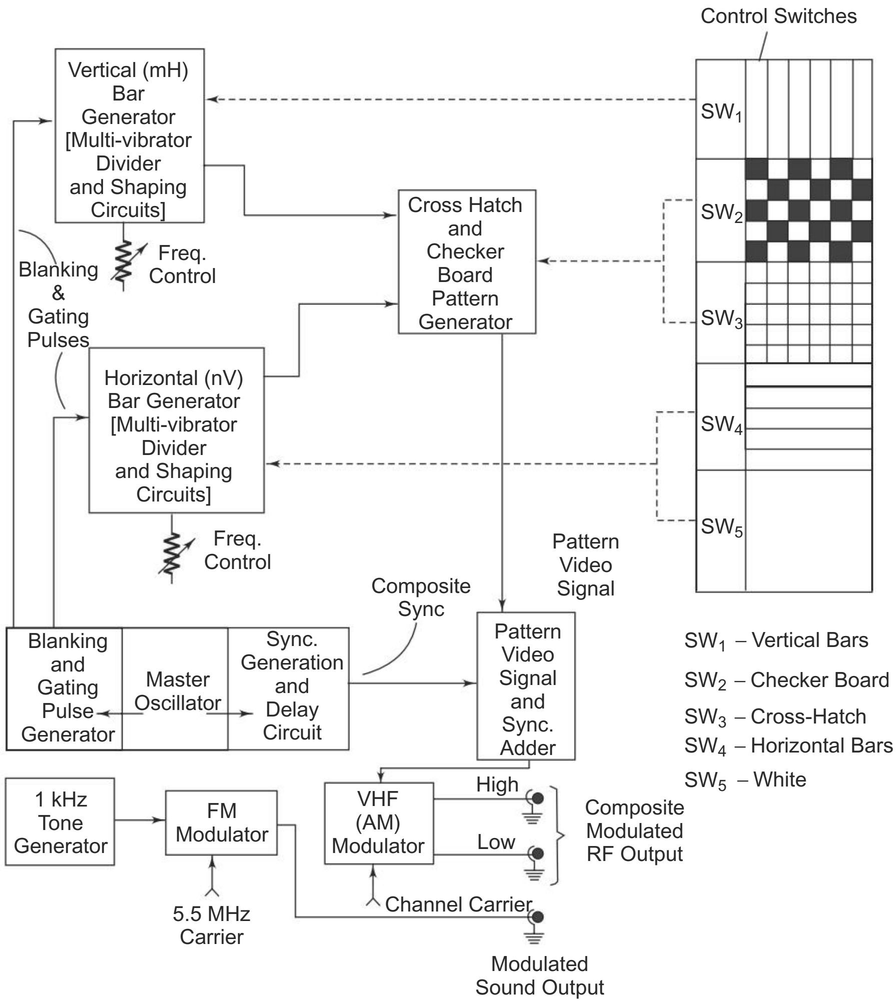


Fig. 8.14 Simplified functional block diagram of a pattern cum sound signal generator

The generator employs two stable chains of multivibrators, dividers and pulse shaping circuits, one below the line frequency to produce a series of horizontal bars, and another above 15625 Hz to produce vertical bars. The signals are modified into short duration pulses, which when fed to the video section of the receiver along with the sync pulse train, produce fine lines on the screen.

Multivibrators produce a square wave video signal at *m* times the horizontal frequency to provide *m* vertical black and white bars. After every *m* cycles, the horizontal blanking pulse triggers the multivibrators for synchronising the bar signal on every line. A control on the front panel of the pattern generator enables variation of multivibrators frequency to change the number of bars.

Similarly, square wave pulses derived either from 50 Hz mains or from the master oscillator are used to trigger another set of multivibrator to generate square wave video signals that are *n* times the vertical frequency. On feeding the video amplifier these produce horizontal black and white bars. The number of horizontal bars can also be varied by a potentiometer that controls the switching rate of the corresponding multivibrator. (The bar pattern signal is combined with the sync and blanking pulses in the video adder to produce composite video signals before being fed to the modulator). The provision of switches in the signal path of the two multivibrators enables the generation of various patterns. If both mH and nV switches are off, a blank white raster is produced. With only the mH switch on, vertical bars are produced, and with only the nV switch on, horizontal bars are generated. With both switches on, a cross-hatch pattern will be produced (Fig. 8.14).

The horizontal bar pattern is used for checking vertical linearity. These bars should be equally spaced throughout the screen for linearity. Similarly, the vertical bar pattern can be used for checking and setting horizontal linearity. With the cross-hatch pattern formed by the vertical and horizontal lines, linearity can be adjusted more precisely, because any unequal spacing of the lines can be discerned.

Picture centering and aspect ratio can also be checked with the cross-hatch pattern by counting the number of squares on the vertical and horizontal sides of the screen.

The pattern generator can also be used for detecting any spurious oscillations in the sweep generation circuits, interaction between the two oscillators, poor interlacing, and barrel and pin cushion effects.

Modulated picture signals are available on limited channels for injecting into the RF section of the receiver.

Similarly, an FM sound signal with a carrier frequency of 5.5 MHz ±100 kHz, modulated by a 1 kHz tone, is provided for aligning sound IF and discriminator circuits. A 75/300 Ω VHF balun is usually available as a standard accessory with the pattern generator.

COLOUR BAR GENERATOR 8.17

The composite video signal at the output of a video detector consists of luminance Y signals, the chrominance signal, the colour burst, sync pulses and blanking pulses.

The amplitude of the video signal varies continuously due to the changing picture content, such a waveform is not useful for adjustment and trouble shooting purposes.

The colour bar generator acts as a substitute transmitter and supplies to the receiver, a known constant amplitude colour pattern signal for alignment and servicing purposes.

8.17.1 The Gated Rainbow Colour Bar Generator

The gated colour bar generator develops a composite video signal that produces a rainbow colour bar pattern on the receiver screen. The pattern consists of 10 colour bars ranging in colour shades from red on the left side through blue in the centre, to green on the far right.

The colour bar pattern, the associated video waveform and the corresponding phase relationship of the gated pattern is given in Figs. 8.15 and 8.16 respectively.

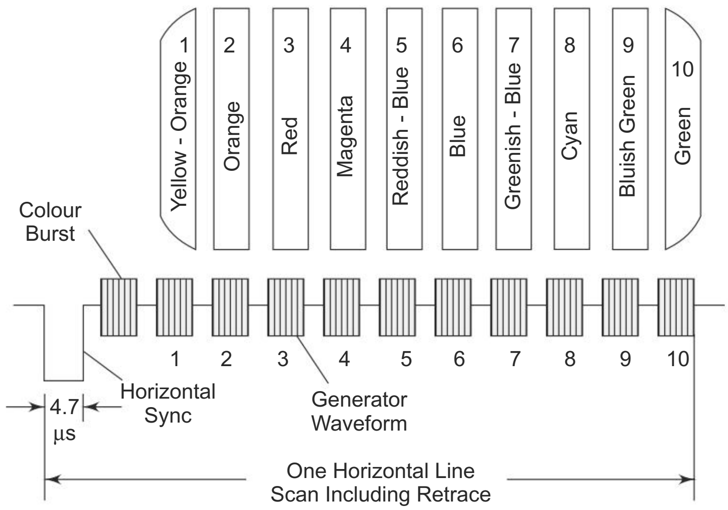


Fig. 8.15 Colour bar pattern on the screen of a colour picture tube

Each colour in the bar pattern has been identified and lined up with the associated modulated voltage.

The composite video signal for the pattern consists of a horizontal sync pulse and 11 equal amplitude bursts of colour sub-carrier frequency. The burst to the right is a colour burst; and other bursts from 1 to 10 differ in phase from one another and correspond to different colours in the bar pattern.

The basic principle of a colour bar generator is quite simple. Any two signals at different frequencies have a phase difference that changes continuously.

As shown in Fig. 8.17, a crystal oscillator is provided to generate a frequency of 4.41799375 MHz. This is 15625 Hz lower than the colour sub-carrier frequency (4.43361875 – 4.41799375 MHz = 15625 Hz). Since the difference in frequency is equal to the horizontal scanning rate, the relative phase between the two carrier frequencies changes by 360° per horizontal line. Thus the effective carrier signal at 4.41799375 MHz will appear as a signal that is continuously changing in phase (360° during each H-line) when compared to the 4.43361875 MHz reference oscillator in the TV receiver. It is the phase of the chrominance signal which determines the colour seen, and therefore has a frequency relationship with the subcarrier frequency that provides the colour bar signal. Since there is a

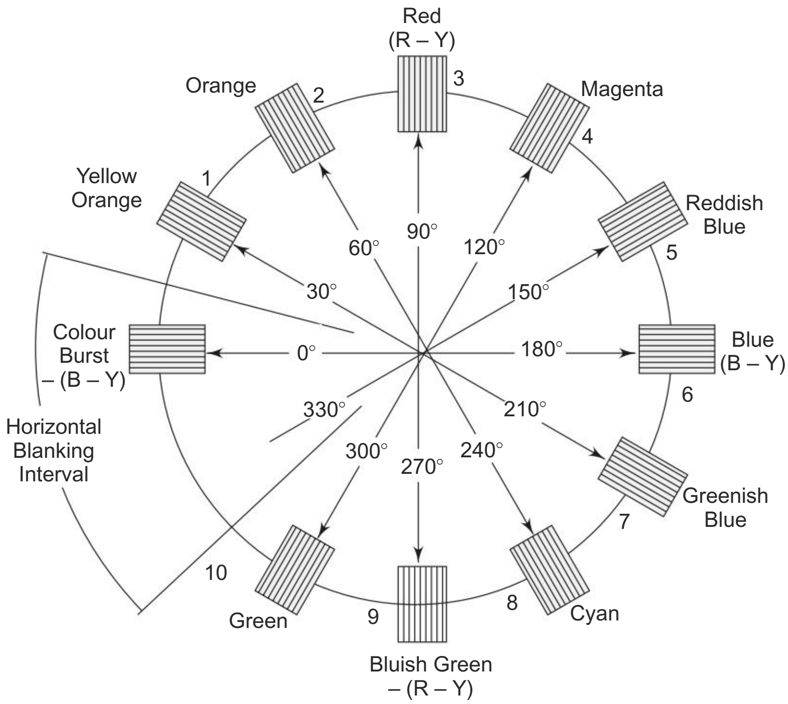


Fig. 8.16 Colour signal phase relationship with respect to colour burst

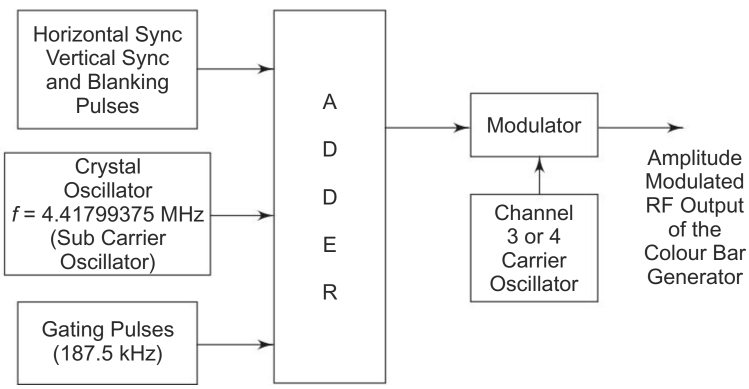


Fig. 8.17 Simplified block diagram of a gated rainbow colour bar generator

complete change of phase of 360° for each H-sweep, a complete range of colours is produced during each H-line. Each line displays all the colours simultaneously, since the phase between the frequency of the crystal oscillator and the horizontal scanning rate frequency is zero at the beginning of each such line and advances to become 360° at the end of each horizontal sweep stroke.

The colour bar pattern is produced by gating On and Off the 4.41799375 MHz oscillator at a rate 12 times higher than the H-sweep frequency (15625 × 12

= 187.5 kHz). The gating at a frequency of 187.5 kHz produces colour bars with blanks between them. The colour bars have a duration corresponding to 15°, and are 30° apart all around the colour spectrum. When viewed on the picture tube screen of a normally operated colour receiver, these bars appear as shown in Fig. 8.15.

Only 10 colour bars are shown in the screen, because one of the bursts occurs at the same time as the H-sync pulse and is thus eliminated. The adder, while gating the crystal oscillator output, also combines H-sync, V-sync and blanking pulses to the oscillator output. The composite colour video signal available at the output of the adder can be fed directly to the chrominance band-pass amplifier in the TV receiver. This signal is usually AM modulated with the carrier of either channel 3 or 4.

The main technical specifications of a colour bar pattern generator are as follows.

Test signals

1. 8 bars, linearised, grey scale
2. Cross-hatch pattern
3. 100% white pattern (with burst)
4. Red pattern (50% saturated)
5. Standard colour bar with white reference. 75% contrast (internally changeable to full bars).

Video carrier

1. VHF B-III (170 MHz - 230 MHz)
2. UHF B-IV (470 MHz - 600 MHz)

RF output

>> 100 mV peak to peak (75 Ω impedance)

Video modulation

Amplitude modulation (negative)

Sound carrier

Frequency	–	5.5 MHz (or 6 MHz by internal adjustment)
Modulation	–	Frequency modulation
Internal signal	–	1 kHz sine wave.

FM Sweep 40 kHz on 5.5 MHz
Chroma-PAL-G and I standards

Power

115 – 230 V; 50 – 60 Hz, 6 W

Dimension

23 × 11 × 21 cm – ($w \times h \times d$)

Weight

1.25 kg

Figure 8.18 illustrates a commercial VHF colour pattern generator (IE-1044) along with the various test patterns generated.

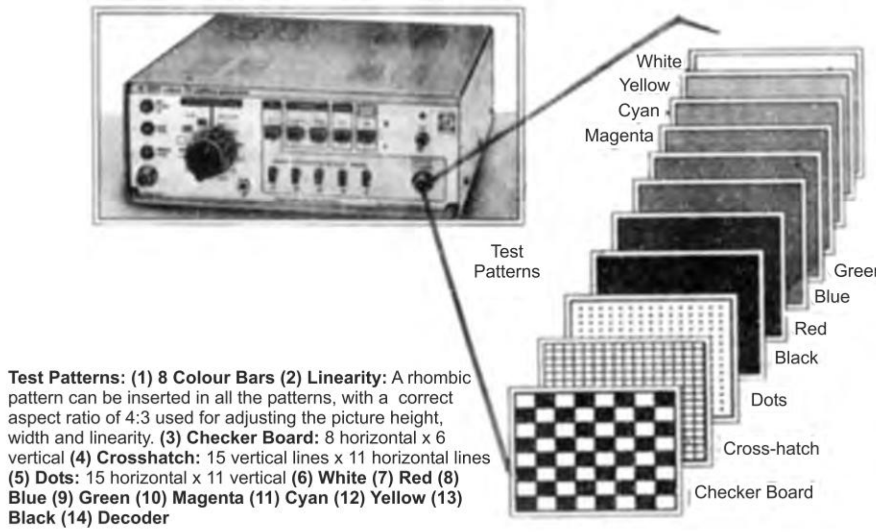


Fig. 8.18 VHF colour pattern generator (IE-1044) along with the test patterns (Courtesy: International Electronics Ltd. Marketed by Signetics Electronics Ltd., Bombay)

VECTROSCOPE 8.18

This test instrument combines a keyed colour bar generator with an oscilloscope and is used for alignment and testing the colour section of a TV receiver. The amplitude and phase of the chrominance signal represents the colour saturation and hue of the scene. This information can also be displayed on the oscilloscope screen in the form of Lissajous patterns. The resultant display is called a vectrogram.

A separate colour bar generator with a conventional CRO can be used to produce a vectrogram. The necessary circuit connections and the resulting pattern are shown in Fig. 8.19.

With the gated colour bar generator connected at the input terminals of the receiver, serrated (R – Y) and (B – Y) video signals become available at corresponding control grids of the colour picture tube. These two outputs are connected to the vertical and horizontal inputs of the CRO. Since both (R – Y) and (B – Y) inputs are interrupted sine waves and have a phase difference of 90° with each other, the resultant Lissajous pattern is in the basic form of a circle which collapses towards the centre during the serrations in the signals.

Assuming ideal input signal waveforms, the formation of a vectrogram is given in Fig. 8.20.

Since there are 10 colour bursts, the pattern displays ten petals. The horizontal sync and colour burst do not appear in the display because these are blanked out during retrace intervals.

The position of each petal represents the phase angle of each colour in the colour bar pattern.

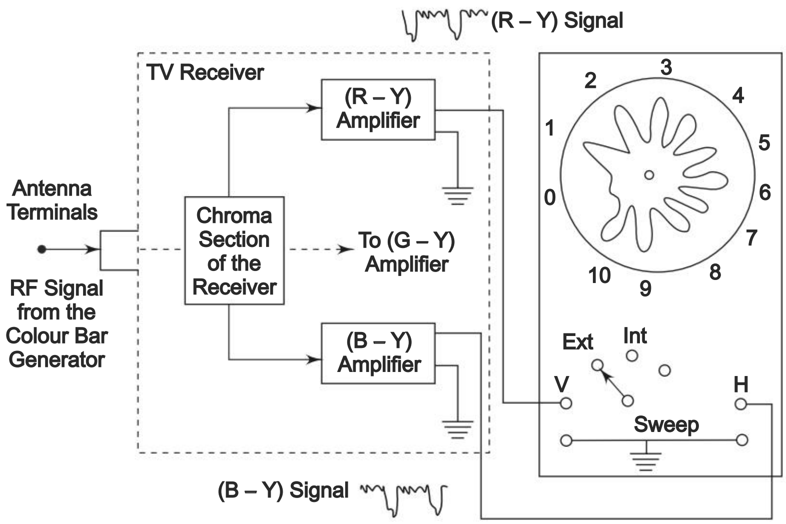


Fig. 8.19 Circuit connections for producing a vectrogram on the CRO

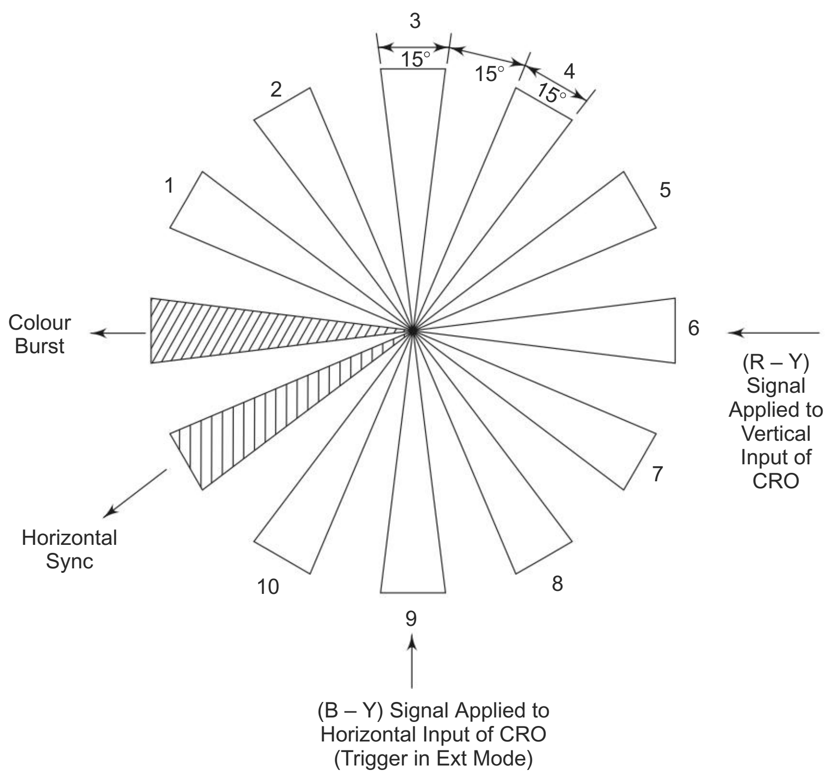


Fig. 8.20 Lissajous pattern

For example, petal 1, 3, R – Y, 6, B – Y and 10 (G – Y) correspond to angles 30°, 90°, 180°, and 360° respectively.

Vectroscopes usually have an overlay sheet on the scope screen, marked with segment numbers and corresponding phase angles. This enables the user to identify different colours and interpret the size and shape of each petal.

In actual practice, the top of (R – Y) and (B – Y) bar signals do not have sharp corners and the resultant pattern is somewhat feathered and rounded at the periphery. This is depicted in the actual pattern shown on the oscilloscope screen in Fig. 8.20.

Rounding of corners and feathering occurs due to limited HF response and non-zero rise-time of the amplifier in a colour bar generator and oscilloscope.

With a vectogram display, chroma troubles can be ascertained and servicing expedited.

For example the loss of an (R – Y) signal causes the vectrogram to be H-line only. Similarly, the absence of (B – Y) results in a single vertical line on the screen. Any change of the receiver colour control will alter the amplitude of both (R – Y) and (B – Y) signals and cause the diameter of the pattern to change.

The receiver fine tuning affects the size and shape of the reproduced pattern. Proper fine tuning produces the largest and best shaped vectrogram.

If some of the petals are longer than others, non-linear distortion is indicated. If the petal tops are flattened, some circuit overloading is occuring in the receiver.

It is also possible to check for defective colour stages, mistuned band-pass amplifier, misadjusted circuitry in the sub-carrier oscillator section and inoperative colour stages.

Vectroscopes are also used in TV recording studios to adjust the white balance of various cameras and to monitor colour signals during recording.

BEAT FREQUENCY OSCILLATOR (BFO) 8.19

In this circuit, the outputs of two RF oscillators are applied to a square law detector and the resulting difference frequency is amplified. The main advantage of this type of oscillator is that a stable continuous output covering the entire AF range can be realised by simple variation of the tuning capacitor in one of the oscillators.

In the circuit given in Fig. 8.21, the voltages obtained from two RF oscillators operating at slightly different frequencies are combined and applied to a mixer circuit. The difference frequency current that is thus produced represents the desired oscillations. The practical value of a BFO arises from the fact that a small or moderate percentage variation in the frequency of one of the individual oscillators (such as can be obtained by the rotation of the shaft controlling a variable tuning capacitor) varies the beat or difference output continuously from a few c/s to throughout the entire AF or video frequency range. At the same time,

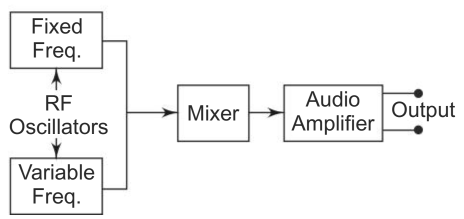


Fig. 8.21 Beat frequency oscillator

the amplitude of the difference frequency output is largely constant as frequency is varied. The principal factors involved in the performance of a BFO are the frequency stability of individual oscillators, the tendency of the oscillators to synchronise at very low difference frequencies, the wave shape of the difference frequency output, and the tendency for spurious beat notes to be produced.

Frequency stability of the individual oscillators is important, because a slight change in their relative frequency would cause a relatively large change in the difference frequency. To minimise the drift of the difference frequency with time, the individual oscillators should have high inherent frequency stability with respect to changes in temperature and to supply voltage variations, and they should be as alike electrically, mechanically and thermally, as possible.

In this way, frequency changes are minimised and the frequency changes that do take place tend to be the same in each of the individual oscillators and so have little effect on the difference in their frequencies.

The two RF oscillators must be completely isolated from each other. If coupling of any type exists between them, they will synchronise when the difference is small. Hence, low values of difference frequency are impossible to obtain, and in addition cause interaction between the oscillators that results in a highly distorted wave shape.

(To ensure low distortion, one of the voltages applied to the mixer, preferably the one derived from the fixed frequency oscillator, should be considerably smaller than the voltage derived from the other oscillator, and preferably free of harmonics.)

BFOs are commonly affected with spurious beat notes, sometimes called whistles. These effects are usually the result of cross-modulation in the AF amplifier between high order RF harmonics generated by the mixer. These spurious whistles often appear when the output frequency is high.

Whistles can be eliminated by operating the mixer so as to minimise the production of RF harmonics, and by using a filter and shielding to prevent the harmonics that are generated in the mixer from reaching the amplifier circuit.

STANDARD SPECIFICATIONS OF A SIGNAL GENERATOR 8.20

Table 8.2 gives the standard specifications of a signal generator

Table 8.2

Properties	*Specifications*
Calibration accuracy	± 2% under normal conditions
Frequency response	Within ± 1 db (of a 1 kHz reference) over the entire frequency range
Frequency stability	Negligible shift in output frequency for ± 10% line voltage variation
Distortion	Less than 0.5% below 500 kHz, (less than 1% above 500 kHz), independent of load impedance

Balanced output	May be obtained (at maximum output), with better than 1% balance, or may be operated single-ended (with low side grounded), at an internal impedance of 600 Ω, for any portion of output attenuation

Review Questions

1. Explain the operation of a basic signal generator.
2. Explain the operating principle of a modern standard signal generator.
3. Describe with the help of a neat block diagram the working of a standard signal Generator. State the limitations of a standard signal generator.
4. Describe with the help of a neat block diagram the operation of a modern laboratory signal generator. Explain the technique used to improve its stability.
5. Explain the operation and use of frequency dividers.
6. Describe with diagram the operation of an AF sine and wave generator. State the various controls on the front panels of a sine and square wave generator.
7. Explain the operating principle of a function generator.
8. Explain with the help of a block diagram the operation of a function generator.
9. Explain the method of producing sine waves in a function generator. Explain the operation of a resistance diode network.
10. Explain with the help of block diagram the working of a standard sweep generator.
11. How are broadband sweep frequencies generated using a sweep generator?
12. State with a diagram the working principle of a pulse generator. Describe with the help of a block diagram the operation of a pulse generator.
13. Define duty cycle. List the requirements of a pulse.
14. List the various control on the front panel of a pulse generator. Mention their uses.
15. State the function of symmetry control in a pulse generator.
16. State the function of frequency sweeper and marker generator in a sweep generator.
17. Explain with diagram the working of a marker generator. What is the necessity of using a marker generator?
18. State the application of a sweep generator.
19. Explain the term 'wobbluscope'.
20. Describe with a diagram the operation of a wobbluscope. State the applications of a wobbluscope.
21. Differentiate between a sweep generator and wobbluscope.
22. State the front panel controls of a wobbluscope.
23. Describe the procedure of alignment RF and IF sections using a wobbluscope.
24. Explain with a block diagram the operation of a pattern generator. State the various applications of a pattern generator.
25. List the various controls on the front panel of a pattern generator. List the various patterns generated by a pattern generator.
26. Differentiate between a function generator and pulse and square wave generator.
27. Compare a wobbluscope and sweep generator.

28. Explain in brief the alignment procedure of a TV receiver using a sweep generator.
29. What are the different methods of obtaining colour bar patterns? How are the colour bar patterns generated?
30. Explain with block diagram the operation of a gated rainbow colour bar generator.
31. List the main technical specifications of a colour bar pattern generator.
32. What is a vectroscope? Where is it used?
33. Explain with a diagram the operation of a vectroscope.
34. How is the vectrogram produced on the CRO? Explain with a diagram.
35. State the applications of a vectroscope.
36. Explain the working principle of a beat frequency oscillator. State its applications.
37. Explain with a diagram the working of a beat frequency oscillator.

Multiple Choice Questions

1. Modulation in modern signal generator is done internally by signals of frequency
   (a) 400 Hz and 1000 Hz
   (b) 600 Hz and 2000 Hz
   (c) 100 Hz and 5000 Hz
   (d) 10000 Hz and 4000 Hz
2. AF sine and square wave generator has an output impedance of
   (a) 600 Ω (b) 200 Ω
   (c) 1000 Ω (d) 50 Ω.
3. A frequency divider used in a modern signal generator
   (a) divides the frequency by 2
   (b) doubles the frequency
   (c) divides the frequency by 10
   (d) multiply the frequency by 2
4. Internal calibration in a modern signal generator is obtained by using
   (a) 2 MHz crystal oscillator
   (b) 1 MHz crystal oscillator
   (c) 5 MHz crystal oscillator
   (d) 5.5 MHz crystal oscillator
5. Frequency dividers are obtained by the use of
   (a) LC network (b) AND gate
   (c) flip-flop's (d) RC network
6. A Wien bridge oscillator is suitable for
   (a) RF generator
   (b) function generator
   (c) pulse generator
   (d) AF generators
7. In a function generator, the resistance diode network is used to produce
   (a) square wave (b) sine wave
   (c) triangular wave (d) pulse wave
8. The principle used in the operation of a function generator is by using an
   (a) *LC* oscillator (b) *RC* oscillator
   (c) integrator (d) derivation
9. The frequency of a function generator is varied by varying
   (a) *LC* network
   (b) *RC* network
   (c) constant current sources
   (d) constant voltage sources
10. A pulse generator generating a square wave has a duty cycle of
    (a) 25% (b) 50%
    (c) 75% (d) 40%
11. A pulse generating, generated a pulse waveform has a duty cycle of
    (a) 25% (b) 40%
    (c) 50% (d) 75%
12. The comparator used in a function generator produces
    (a) square wave
    (b) triangular wave
    (c) sine wave
    (d) pulse wave

13. The frequency sweeper provides the modulating voltage which varies the
    (a) inductance (b)capacitance
    (c) resistance (d) voltage
14. A sweep generator is used for
    (a) fault finding
    (b) frequency generation
    (c) amplification
    (d) alignment
15. A wobbluscope is used for alignment of a/an
    (a) radio receiver (b) TV receiver
    (c) oscilloscope (d) wave analyzer
16. Picture centering and aspect ratio using a pattern generator can be checked by the pattern
    (a) horizontal bar
    (b) vertical bar
    (c) cross hatch
    (d) checker board
17. In a function generator, the spectral purity of a sine wave is poor as compared to the Wein bridge type oscillator because
    (a) it is obtained using piecewise linear approximation
    (b) it uses crystal reactive element
    (c) inductor element used has a non-linear B-H relation
18. A digital counter IC is to be checked for it's performance up to 25 MHz. It can be done
    (a) RF signal generator
    (b) function generator
    (c) pulse generator
    (d) pattern generator
19. The number of colour burst generated by the colour bar pattern generated is
    (a) 10 (b) 20 (c) 5 (d) 15
20. If the number of colour bursts present are ten then the number of petals displayed are
    (a) 15 (b) 10 (c) 20 (d) 5

Further Reading

1. Oliver Cage, *Electronic Measurements and Instrumentation*, McGraw-Hill, 1975.
2. R.R. Gulati, *Monochrome and Colour TV*, Wiley Eastern, 1983.
3. Larry D. Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley & Sons, 1987.

Wave Analyzers and Harmonic Distortion

Chapter 9

INTRODUCTION 9.1

It can be shown mathematically that any complex waveform is made up of a fundamental and its harmonics.

It is often desired to measure the amplitude of each harmonic or fundamental individually. This can be performed by instruments called wave analyzers. This is the simplest form of analysis in the frequency domain, and can be performed with a set of tuned filters and a voltmeter. Wave analyzers are also referred to as frequency selective voltmeters, carrier frequency voltmeters, and selective level voltmeters. The instrument is tuned to the frequency of one component whose amplitude is measured.

This instrument is a narrow band superheterodyne receiver, similar to a spectrum analyzer (discussed later). It has a very narrow pass-band. A meter is used for measurement, instead of a CRT. Wave analyzers are used in the low RF range, below 50 MHz and down through the AF range. They provide a very high frequency resolution.

Some wave analyzers have the facility of automatic frequency control, in which the tuning automatically locks to a signal. This makes it possible to measure the amplitude of signals that are drifting in frequency by amounts that would carry them outside the widest pass-band available.

When a sinusoidal signal is applied to the input of an ideal linear amplifier, it produces a sinusoidal output waveform. However, in most cases the output waveform is not an exact replica of the input signal because of different types of distortion. The amount by which the output waveform of an amplifier differs from the input waveform is a measure of the distortion introduced by the inherent non-linear characteristics of the active devices.

Harmonic distortion analyzers measure the total harmonic content in the waveforms. It can be shown mathematically that an amplitude distorted sine wave is made up of pure sine wave components, including the fundamental frequency f of the input signal, and harmonic multiples of the fundamental frequency, $2f$, $3f$, $4f$ etc.

Harmonic distortion can be quantitatively measured very accurately with a harmonic distortion analyzer, generally called a distortion analyzer.

The total harmonic distortion or factor is given by

$$D = \sqrt{D_2^2 + D_3^2 + D_4^2 \cdots}$$

where D_2, D_3, D_4 … represent the second harmonic, third harmonic, etc. respectively.

The distortion analyzer measures the total harmonic distortion without indicating the amplitude and frequency of each component waves.

Signal analysis of both random and periodic signals in the frequency domain is used extensively in electronic and telecommunications. The frequency stability and spectral purity of signal sources can be measured by the use of these signal analyzers.

These signal analyzers can be used along with a frequency generator or a source of white or pseudo-random noise to measure the frequency response of amplifiers, filters or other networks.

The operational characteristics of a transreceiver and communication system are determined by measuring various parameters, such as spectral purity of the carrier wave, spectral power distribution of the amplitude or frequency modulated wave, signal distortion, and the systems signal to noise ratio.

Such analysis is provided by a wave analyzer, distortion analyzer, spectrum analyzer, and digital fourier analyzer.

BASIC WAVE ANALYZER 9.2

A basic wave analyzer is shown in Fig. 9.1(a). It consists of a primary detector, which is a simple LC circuit. This LC circuit is adjusted for resonance at the frequency of the particular harmonic component to be measured.

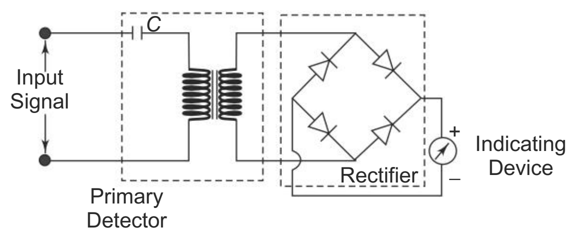


Fig. 9.1 (a) Basic wave analyzer

The intermediate stage is a full wave rectifier, to obtain the average value of the input signal. The indicating device is a simple dc voltmeter that is calibrated to read the peak value of the sinusoidal input voltage.

Since the LC circuit is tuned to a single frequency, it passes only the frequency to which it is tuned and rejects all other frequencies. A number of tuned filters, connected to the indicating device through a selector switch, would be required for a useful Wave analyzer.

FREQUENCY SELECTIVE WAVE ANALYZER 9.3

The wave analyzer consists of a very narrow pass-band filter section which can be tuned to a particular frequency within the audible frequency range (20 Hz – 20 kHz). The block diagram of a wave analyzer is as shown in Fig. 9.1(b).

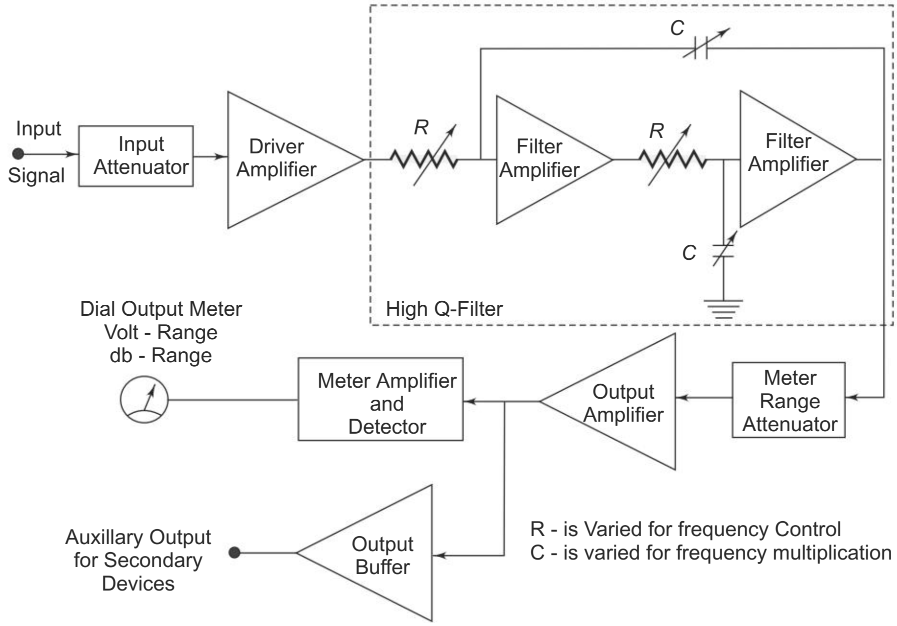


Fig. 9.1 (b) Frequency selective wave analyzer

The complex wave to be analyzed is passed through an adjustable attenuator which serves as a range multiplier and permits a large range of signal amplitudes to be analyzed without loading the amplifier.

The output of the attenuator is then fed to a selective amplifier, which amplifies the selected frequency. The driver amplifier applies the attenuated input signal to a high-Q active filter. This high-Q filter is a low pass filter which allows the frequency which is selected to pass and reject all others. The magnitude of this selected frequency is indicated by the meter and the filter section identifies the frequency of the component. The filter circuit consists of a cascaded RC resonant circuit and amplifiers. For selecting the frequency range, the capacitors generally used are of the closed tolerance polystyrene type and the resistances used are precision potentiometers. The capacitors are used for range changing and the potentiometer is used to change the frequency within the selected pass-band, Hence this wave analyzer is also called a Frequency selective voltmeter.

The entire AF range is covered in decade steps by switching capacitors in the RC section.

The selected signal output from the final amplifier stage is applied to the meter circuit and to an untuned buffer amplifier. The main function of the buffer amplifier is to drive output devices, such as recorders or electronics counters.

The meter has several voltage ranges as well as decibel scales marked on it. It is driven by an average reading rectifier type detector.

The wave analyzer must have extremely low input distortion, undetectable by the analyzer itself. The bandwidth of the instrument is very narrow, typically about 1% of the selective band given by the following response characteristics shown in Fig. 9.2).

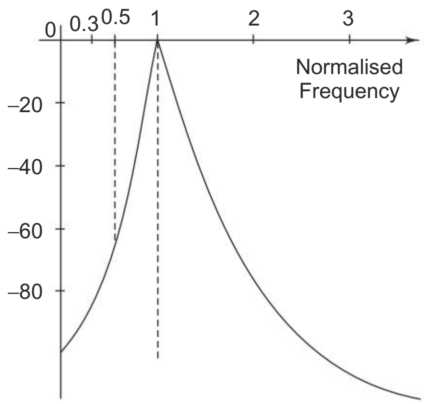


Fig. 9.2 Relative response in dBs

HETERODYNE WAVE ANALYZER 9.4

Wave analyzers are useful for measurement in the audio frequency range only. For measurements in the RF range and above (MHz range), an ordinary wave analyzer cannot be used. Hence, special types of wave analyzers working on the principle of heterodyning (mixing) are used. These wave analyzers are known as Heterodyne wave analyzers.

In this wave analyzer, the input signal to be analyzed is heterodyned with the signal from the internal tunable local oscillator in the mixer stage to produce a higher IF frequency.

By tuning the local oscillator frequency, various signal frequency components can be shifted within the pass-band of the IF amplifier. The output of the IF amplifier is rectified and applied to the meter circuit.

An instrument that involves the principle of heterodyning is the Heterodyning tuned voltmeter, shown in Fig. 9.3.

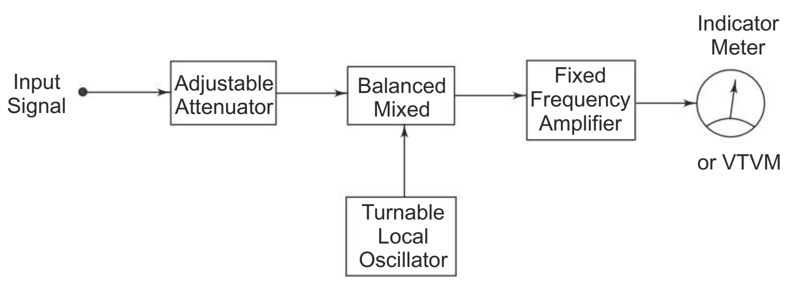


Fig. 9.3 Heterodyne wave analyzer

The input signal is heterodyned to the known IF by means of a tunable local oscillator. The amplitude of the unknown component is indicated by the VTVM or output meter. The VTVM is calibrated by means of signals of known amplitude.

The frequency of the component is identified by the local oscillator frequency, i.e. the local oscillator frequency is varied so that all the components can be identified. The local oscillator can also be calibrated using input signals of known frequency. The fixed frequency amplifier is a multistage amplifier which can be designed conveniently because of its frequency characteristics. This analyzer has good frequency resolution and can measure the entire AF frequency range. With the use of a suitable attenuator, a wide range of voltage amplitudes can be covered. Their disadvantage is the occurrence of spurious cross-modulation products, setting a lower limit to the amplitude that can be measured.

Two types of selective amplifiers find use in Heterodyne wave analyzers. The first type employs a crystal filter, typically having a centre frequency of 50 kHz. By employing two crystals in a band-pass arrangement, it is possible to obtain a relatively flat pass-band over a 4 cycle range. Another type uses a resonant circuit in which the effective *Q* has been made high and is controlled by negative feedback. The resultant signal is passed through a highly selective 3-section quartz crystal filter and its amplitude measured on a *Q*-meter.

When a knowledge of the individual amplitudes of the component frequency is desired, a heterodyne wave analyzer is used.

A modified heterodyne wave analyzer is shown in Fig. 9.4. In this analyzer, the attenuator provides the required input signal for heterodyning in the first mixer stage, with the signal from a local oscillator having a frequency of 30 – 48 MHz.

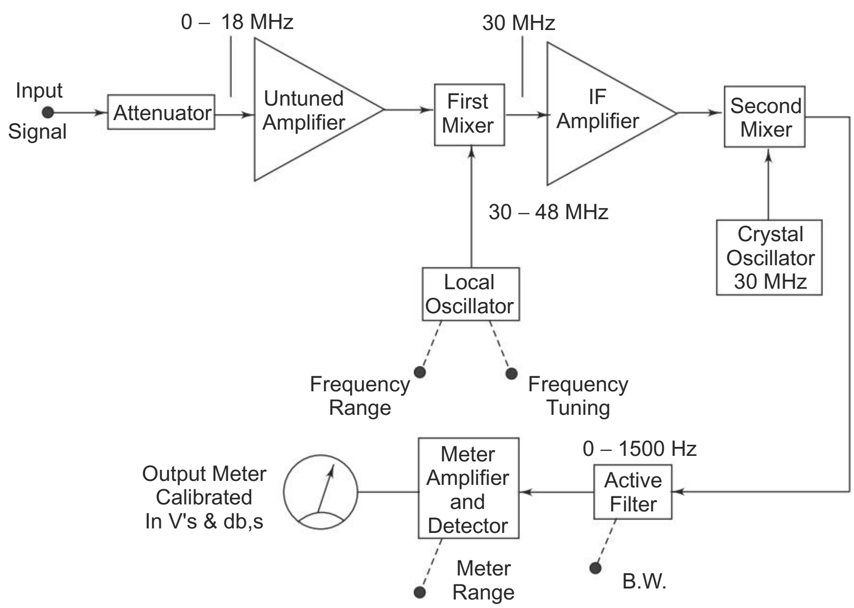


Fig. 9.4 RF heterodyne wave analyzer

The first mixer stage produces an output which is the difference of the local oscillator frequency and the input signal, to produce an IF signal of 30 MHz.

This IF frequency is uniformly amplified by the IF amplifier. This amplified IF signal is fed to the second mixer stage, where it is again heterodyned to produce a difference frequency or IF of zero frequency.

The selected component is then passed to the meter amplifier and detector circuit through an active filter having a controlled band-width. The meter detector output can then be read off on a db-calibrated scale, or may be applied to a secondary device such as a recorder.

This wave analyzer is operated in the RF range of 10 kHz – 18 MHz, with 18 overlapping bands selected by the frequency range control of the local oscillator. The bandwidth, which is controlled by the active filter, can be selected at 200 Hz, 1 kHz and 3 kHz.

HARMONIC DISTORTION ANALYZER 9.5

9.5.1 Fundamental Suppression Type

A distortion analyzer measures the total harmonic power present in the test wave rather than the distortion caused by each component. The simplest method is to suppress the fundamental frequency by means of a high pass filter whose cut off frequency is a little above the fundamental frequency. This high pass allows only the harmonics to pass and the total harmonic distortion can then be measured. Other types of harmonic distortion analyzers based on fundamental suppression are as follows.

1. Employing a Resonance Bridge The bridge shown in Fig. 9.5 is balanced for the fundamental frequency, i.e. L and C are tuned to the fundamental frequency. The bridge is unbalanced for the harmonics, i.e. only harmonic power will be available at the output terminal and can be measured. If the fundamental frequency is changed, the bridge must be balanced again. If L and C are fixed components, then this method is suitable only when the test wave has a fixed frequency. Indicators can be thermocouples or square law VTVMs. This indicates the rms value of all harmonics. When a continuous adjustment of the fundamental frequency is desired, a Wien bridge arrangement is used as shown in Fig. 9.6.

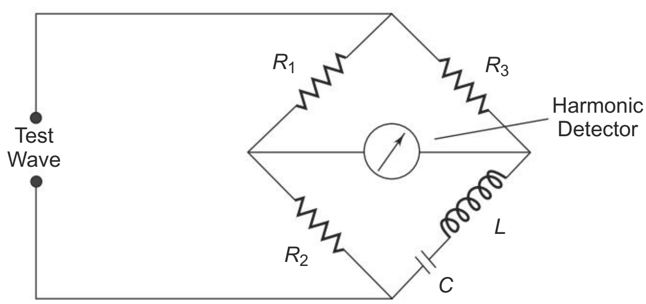


Fig. 9.5 Resonance bridge

2. Wien's Bridge Method The bridge is balanced for the fundamental frequency. The fundamental energy is dissipated in the bridge circuit elements. Only the harmonic components reach the output terminals. The harmonic distortion output can then be measured with a meter. For balance at the fundamental frequency, $C_1 = C_2 = C$, $R_1 = R_2 = R$, $R_3 = 2R_4$.

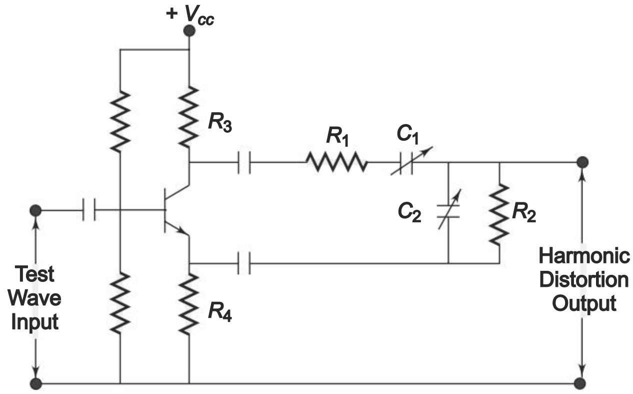


Fig. 9.6 Wien's bridge method

3. Bridged T-Network Method Referring to Fig. 9.7 the, L and C's are tuned to the fundamental frequency, and R is adjusted to bypass fundamental frequency. The tank circuit being tuned to the fundamental frequency, the fundamental energy will circulate in the tank and is bypassed by the resistance. Only harmonic components will reach the output terminals and the distorted output can be measured by the meter. The Q of the resonant circuit must be at least 3–5.

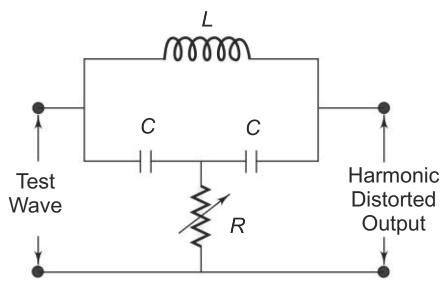


Fig. 9.7 Bridged T-network method

One way of using a bridge T-network is given in Fig. 9.8.

The switch S is first connected to point A so that the attenuator is excluded and the bridge T-network is adjusted for full suppression of the fundamental frequency, i.e. minimum output. Minimum output indicates that the bridged T-network is tuned to the fundamental frequency and that the fundamental frequency is fully suppressed.

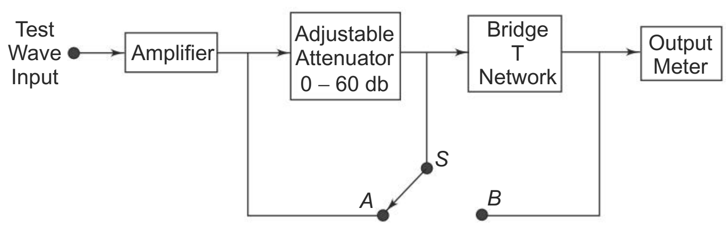


Fig. 9.8 Harmonic distortion analyzer using bridged T-network

The switch is next connected to terminal B, i.e. the bridged T-network is excluded. Attenuation is adjusted until the same reading is obtained on the meter. The attenuator reading indicates the total rms distortion. Distortion measurement can also be obtained by means of a wave analyzer, knowing the amplitude and the frequency of each component, the harmonic distortion can be calculated. However, distortion meters based on fundamental suppression are simpler to design and less expensive than wave analyzers. The disadvantage is that they give only the total distortion and not the amplitude of individual distortion components.

SPECTRUM ANALYZER 9.6

The most common way of observing signals is to display them on an oscilloscope, with time as the X-axis (i.e. amplitude of the signal versus time). This is the time domain. It is also useful to display signals in the frequency domain. The instrument providing this frequency domain view is the spectrum analyzer.

A spectrum analyzer provides a calibrated graphical display on its CRT, with frequency on the horizontal axis and amplitude (voltage) on the vertical axis.

Displayed as vertical lines against these coordinates are sinusoidal components of which the input signal is composed. The height represents the absolute magnitude, and the horizontal location represents the frequency.

These instruments provide a display of the frequency spectrum over a given frequency band. Spectrum analyzers use either a parallel filter bank or a swept frequency technique.

In a parallel filter bank analyzer, the frequency range is covered by a series of filters whose central frequencies and bandwidth are so selected that they overlap each other, as shown in Fig. 9.9(a).

Typically, an audio analyzer will have 32 of these filters, each covering one third of an octave.

For wide band narrow resolution analysis, particularly at RF or microwave signals, the swept technique is preferred.

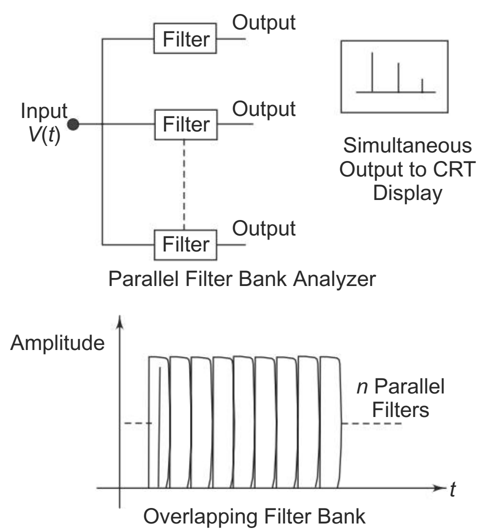


Fig. 9.9 (a) Spectrum analyzer (Parallel filter bank analyzer)

9.6.1 Basic Spectrum Analyzer Using Swept Receiver Design

Referring to the block diagram of Fig. 9.9(b), the sawtooth generator provides the sawtooth voltage which drives the horizontal axis element of the scope and this sawtooth voltage is the frequency controlled element of the voltage tuned oscillator. As the oscillator sweeps from f_{min} to f_{max} of its frequency band at a linear recurring rate, it beats with the frequency component of the input signal and produce an IF, whenever a frequency component is met during its sweep. The frequency component and voltage tuned oscillator frequency beats together to produce a difference frequency, i.e. IF. The IF corresponding to the component is amplified and detected if necessary, and then applied to the vertical plates of the CRO, producing a display of amplitude versus frequency.

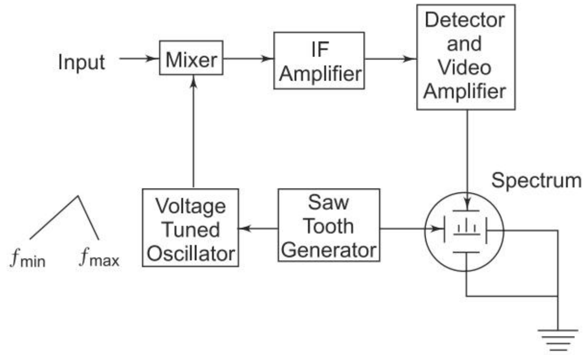


Fig. 9.9 (b) Spectrum analyzer

The spectrum produced if the input wave is a single toned A.M. is given in Figs 9.10, 9.11, and 9.12.

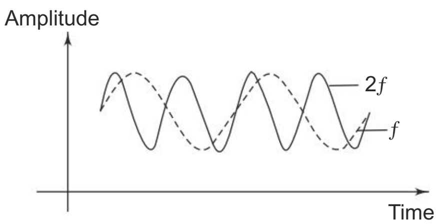


Fig. 9.10 Test wave seen on ordinary CRO

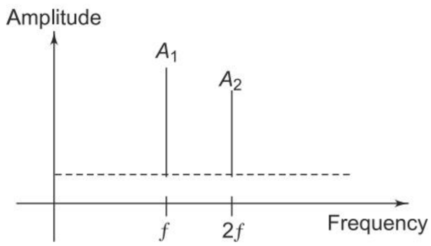


Fig. 9.11 Display on the spectrum CRO

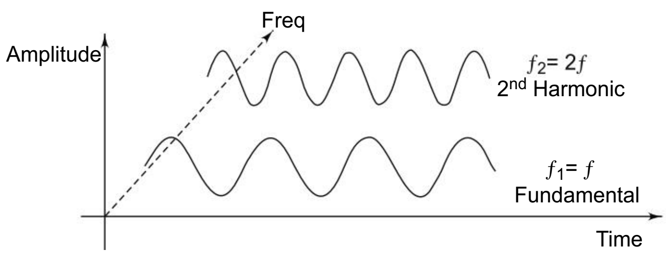


Fig. 9.12 Test waveform as seen on X-axis (Time) and Z-axis (Frequency)

One of the principal applications of spectrum analyzers has been in the study of the RF spectrum produced in microwave instruments. In a microwave instrument, the horizontal axis can display as a wide a range as 2 – 3 GHz for a broad survey and as narrow as 30 kHz, for a highly magnified view of any small portion of the spectrum. Signals at microwave frequency separated by only a few kHz can be seen individually.

The frequency range covered by this instrument is from 1 MHz to 40 GHz. The basic block diagram (Fig. 9.13) is of a spectrum analyzer covering the range 500 kHz to 1 GHz, which is representative of a superheterodyne type.

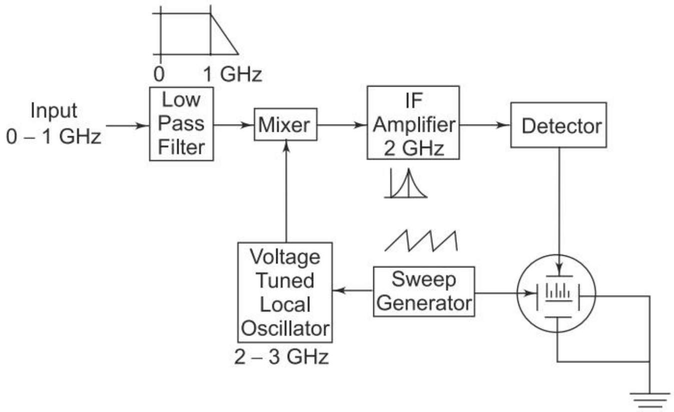


Fig. 9.13 RF spectrum analyzer

The input signal is fed into a mixer which is driven by a local oscillator. This oscillator is linearly tunable electrically over the range 2 – 3 GHz. The mixer provides two signals at its output that are proportional in amplitude to the input signal but of frequencies which are the sum and difference of the input signal and local oscillator frequency.

The IF amplifier is tuned to a narrow band around 2 GHz, since the local oscillator is tuned over the range of 2 – 3 GHz, only inputs that are separated from the local oscillator frequency by 2 GHz will be converted to IF frequency band, pass through the IF frequency amplifier, get rectified and produce a vertical deflection on the CRT.

From this, it is observed that as the sawtooth signal sweeps, the local oscillator also sweeps linearly from 2 – 3 GHz. The tuning of the spectrum analyzer is a swept receiver, which sweeps linearly from 0 to 1 GHz. The sawtooth scanning signal is also applied to the horizontal plates of the CRT to form the frequency axis. (The spectrum analyzer is also sensitive to signals from 4 – 5 GHz referred to as the image frequency of the superheterodyne. A low pass filter with a cutoff frequency above 1 GHz at the input suppresses these spurious signals.) Spectrum analyzers are widely used in radars, oceanography, and bio-medical fields.

DIGITAL FOURIER ANALYZER 9.7

The basic principle of a digital fourier analyzer is shown in Fig. 9.14. The digital fourier analyzer converts the analogue waveform over time period T into N samples.

The discrete spectral response $S_x\,(k\,\Delta f)$; $k = 1, 2, ..., N$ which is equivalent to simultaneously obtaining the output from N filters having a bandwidth given by $\Delta f = 1/T$, is obtained by applying a Discrete Fourier Transform (DFT) to the sampled version of the signal. The spectral response is thus given by

$$S_x(k\ \Delta\ f) = \frac{T}{N}\sum_{n=1}^{N} x(n \cdot \Delta t)\ \exp\left(\frac{-j\,2\,\Pi\,k\,n}{N}\right)$$

where $k = 1, 2, 3, ..., N$.

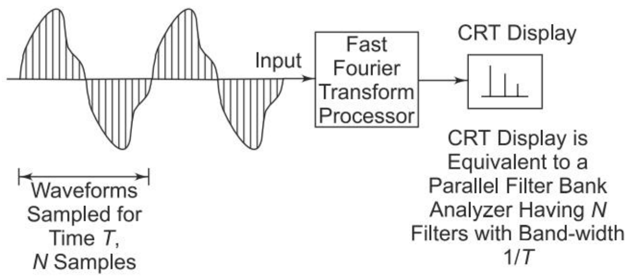


Fig. 9.14 Basic of a digital fourier analyzer

$S_x\,(k\,\Delta\,f)$ is a complex quantity, which is obtained by operating on all the sample $x\,(n \cdot \Delta\,t)$; $n = 1, 2, 3, ..., N$ by the complex factor $\exp\,[-j\,[(2\,\Pi\,kn)/N]]$.

The discrete inverse transform is given by

$$x(n \cdot \Delta t) = \frac{N}{T}\sum_{n=1}^{N} S_x(k \cdot \Delta f)\ \exp\left(\frac{-j\,2\,\Pi\,k\,n}{N}\right)$$

where $n = 1, 2, ..., N$.

Since $S_x\,(k \cdot \Delta\,f)$; $k = 1, 2, ..., N$ is a complex quantity, the DFT provides both amplitude and phase information at a particular point in the spectrum.

The discrete transforms are usually implemented by means of the Fast Fourier Transform (FFT), which is particularly suitable for implementation in a digital computer, since N is constrained to the power of 2, i.e. $2^{10} = 1024$.

A digital signal analyzer block diagram is shown in Fig. 9.15. This digital signal analyzer employs an FFT algorithm.

The block diagram is divided into three sections, namely the input section, the control section and the display section.

The input section consists of two identical channels. The input signal is applied to the input amplifier, where it is conditioned and passed through two or more anti-aliasing filters. The cut-off frequencies of these filters are selected with respect to the sampling frequency being used. The 30 kHz filter is used with a sampling rate of 102.4 kHz and the 300 kHz filter with a sampling rate of 1.024 MHz.

To convert the signal into digital form, a 12 bit ADC is used. The output from the ADC is connected to a multiplier and a digital filter.

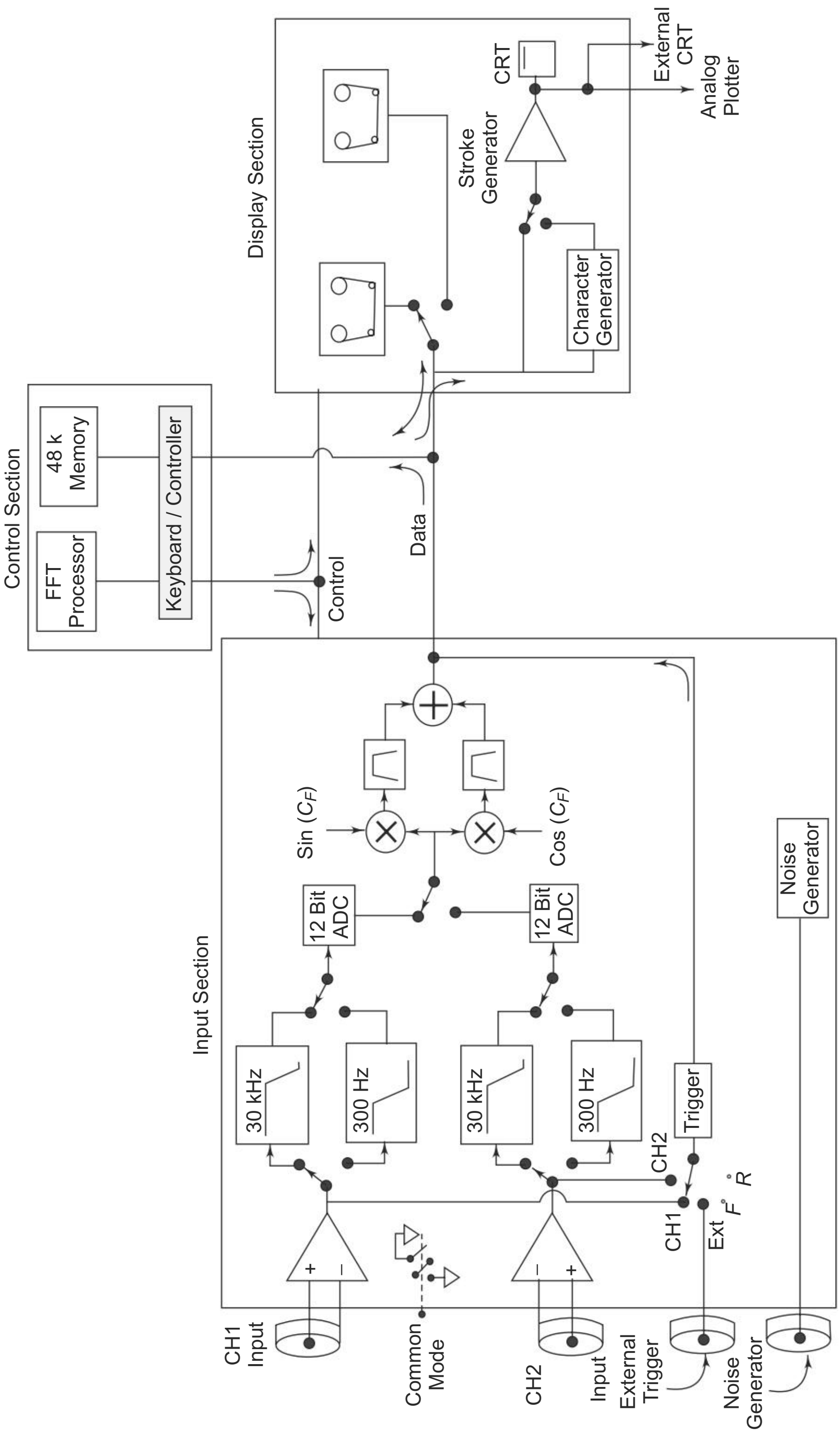


Fig. 9.15 Block diagram of a digital signal analyzer

Depending on the mode of the analyzer to be used, either in Base-band mode (in which the spectrum is displayed from a dc to an upper frequency within the bandwidth of the analyzer) or in the band selectable mode (which allows the full resolution of the analyzer to be focussed in a narrow frequency band), the signal is multiplied either by a sine or cosine function.

The processing section of the analyzer provides FFT processing on the input signal (linear or logarithm).

For one channel this can provide the real (magnitude) and imaginary (phase) of the linear spectrum $S_x(f)$ of a time domain signal

$$S_x(f) = F(x(t))$$

where $F(x(t))$ is the Fourier transform of $x(t)$. The autospectrum $G_{xx}(f)$ which contains no phase information is obtained from $S_x(f)$ as

$$G_{xx}(f) = S_x(f)\, S_x(f)^*$$

where $S_x(f)^*$ indicates the complex conjugate of $S_x(f)$.

The Power Spectral Density (PSD) is obtained by normalising the function $G_{xx}(f)$ to a bandwidth of 1 Hz, which represents the power in a bandwidth of 1 Hz centered around the frequency f.

The Inverse Fourier Transform of $G_{xx}(f)$ is given by

$$R_{xx}(\tau) = F^{-1}(G_{xx}(f))$$
$$R_{xx}(\tau) = F^{-1}(S_x(f)\, S_x(f))^*$$

writing the above equation in terms of the time domain characteristics of the signal $x(t)$, its autocorrelation function is defined as

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T x(t)\, x(t+\tau)\, dt$$

By the use of two channels, the combined properties of the two signals can be obtained. The cross-power spectrum of the two signals $x(t)$ and $y(t)$ can be computed as

$$G_{yx}(f) = S_y(t)\, S_x(t)^*$$

where $S_y(t)$ is the linear spectrum of $y(t)$ and $S_x(t)^*$ is the complex conjugate spectrum of $x(t)$.

If $x(t)$ represents the input to a system and $y(t)$ the output of the system, then its transfer function $H(f)$, which contains both amplitude and phase information can be obtained by computing

$$H(f) = \frac{\overline{G_{yx}(f)}}{\overline{G_{xx}(f)}}$$

where the bars indicate the time averaged values.

The input signal used for such measurements is often the internal random noise generator.

PRACTICAL FFT SPECTRUM ANALYSIS USING A WAVEFORM PROCESSING SOFTWARE (SS-36) 9.8

The Waveform Processing Software (SS-36) enables one to analyze waveforms very easily. The waveform is captured on a Digital Storage Oscilloscope (DSO) and is transferred to a PC via an RS 232 Serial Interface, which may be in-built in some DSOs.

Further, the software is programmed to work with almost any configuration of PC. The software automatically senses the system and configures itself for it (i.e. whether you are using a monochrome or colour monitor, whether you have a co-processor or not etc.).

Suppose a voltage and current waveform in a circuit, is captured on a DSO.

If the voltage is multiplied with the current, power is obtained and if voltage is divided by current, impedance is obtained. If this is done for all 1000 values of the voltage and the current captured on the DSO and plotted, a power and impedance curve is obtained. This process is performed by the software itself. The software can also perform multiplication, division, addition, and subtraction of two simple waveforms, as shown in Fig. 9.16. Figures 9.16(d) and (e) shows one waveform can be added and subtracted by another waveform using FFT analysis.

FFT spectrum analysis is also a very powerful tool. The conventional CRO gives a display of voltage versus time, but by using FFT, the amplitudes of the various frequencies that constitute the waveform can be obtained. The voltage versus frequency plot then shows the spectral content of the waveform. For instance, if a mains voltage is captured and analyzed, its harmonic content can be obtained.

The frequency range of FFT spectrum analysis is from 0.002 Hz to 10 MHz. The minimum step or frequency resolution possible is 0.002 Hz. The display can be linear or logarithmic.

Another example of waveform processing is of an amplitude modulated signal (100 kHz modulated by 8 kHz). This waveform, along with its fourier analysis, is illustrated in Fig. 9.17.

Figure 9.17(2a) shows a AM 100 kHz waveform modulated by 8 kHz, captured on a digital storage oscilloscope and on a PC.

Figure 9.17(2b) and (2c) shows its FFT analysis (carrier and sidebands).

Figure 9.17(1a) illustrates the captured waveform of an electrical or mechanical impulse.

Figure 9.17(1b and 1c) shows the FFT spectral analysis.

Figure 9.18 Shows a pictorial set up of FFT spectrum analysis.

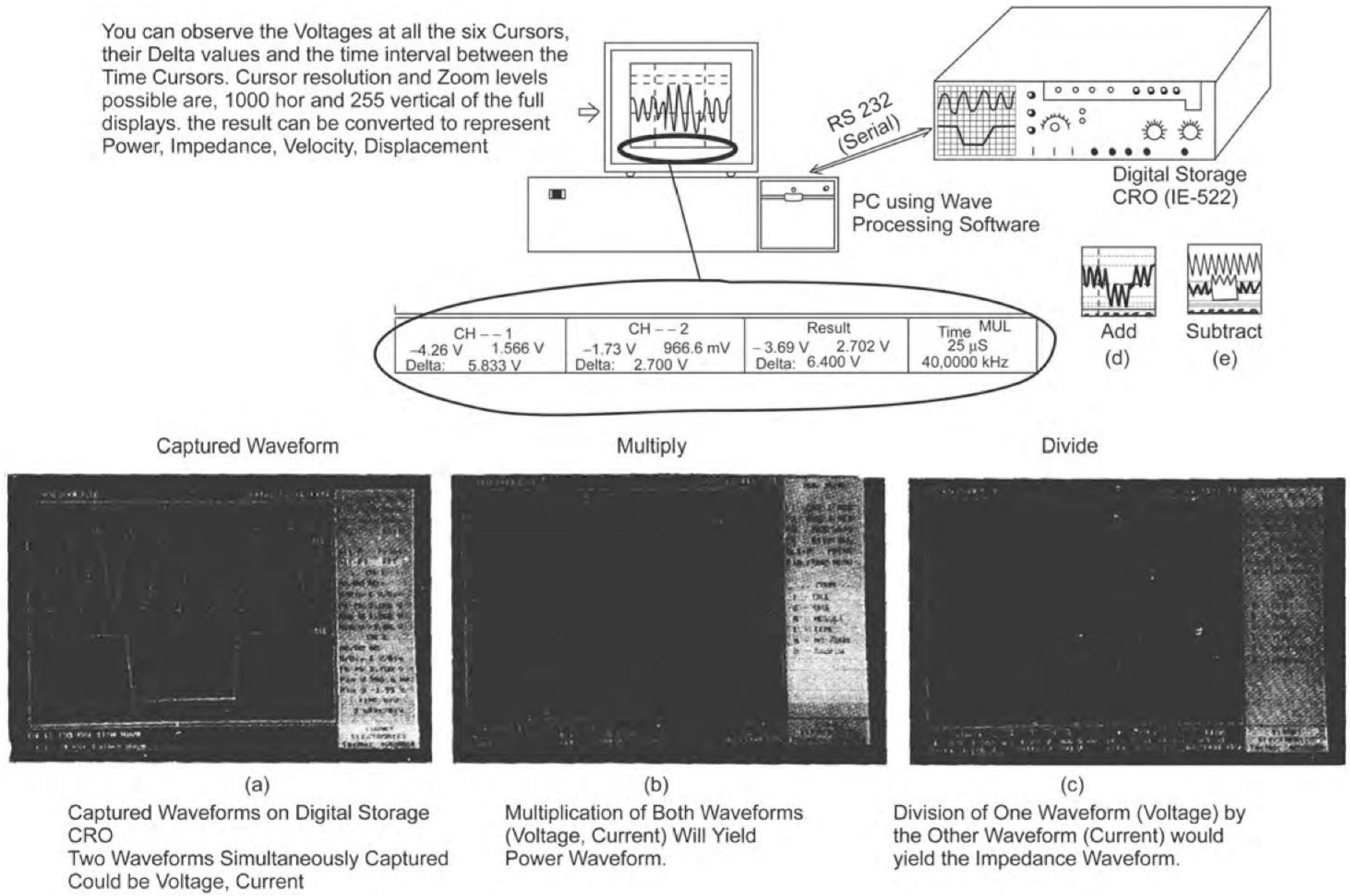
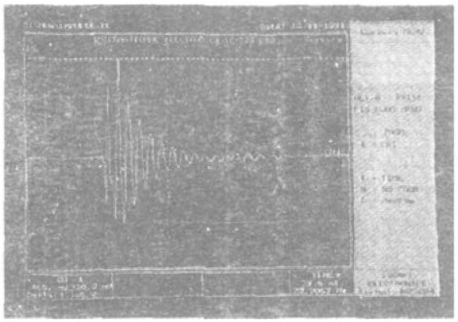
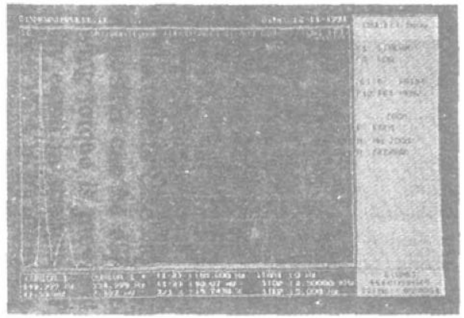


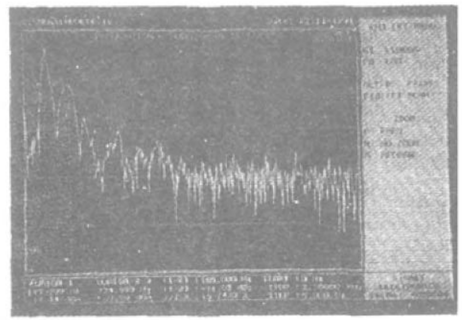
Fig. 9.16 Software waveform processing using digital storage CRO and a PC [Courtesy: International Electronics Ltd] [Marketed: Signatics Elec-



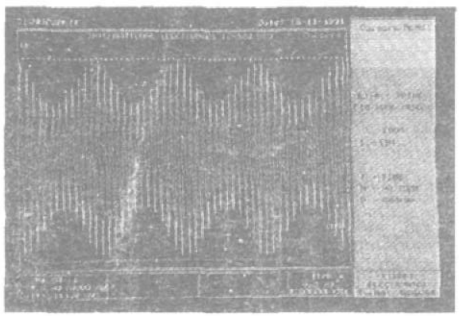
1(a) Impulse Captured on DSO can be Electrical, Mechanical



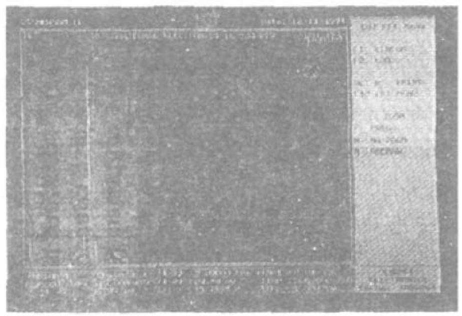
1(b) Spectral Display in LINEAR Mode, (B) in LOG Mode



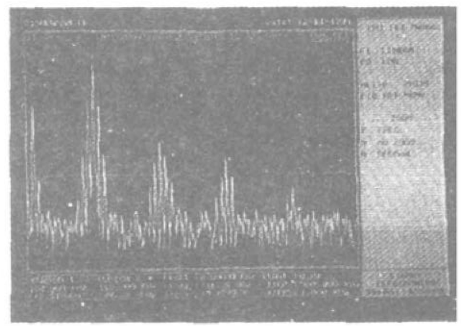
1(c) FFT Analysis of Impulse-Cursor Measure Amplitude & Freq. (T)



2(a) AM Modulated 100 kHz by 8 kHz

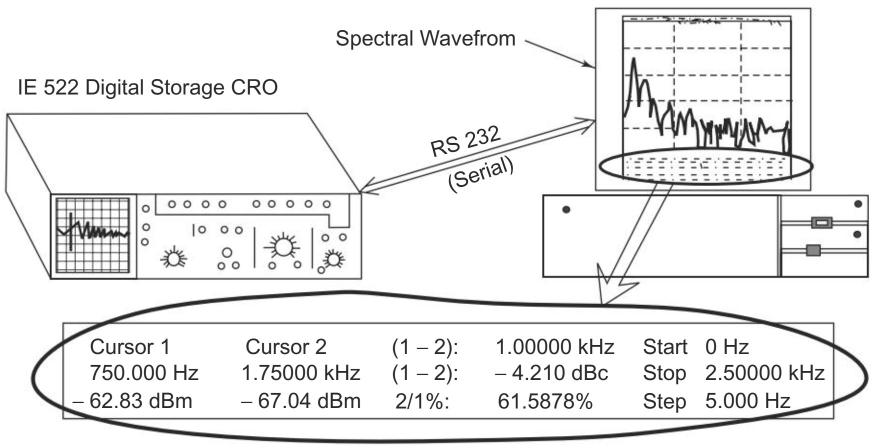


2(b) FFT Analysis Shows the Carrier and the Sidelobes



2(c) Cursors gives Ampl, Freq, % Modulation, 2nd Harmonics

Fig. 9.17 Spectral analyses using 1. Impulse (electrical/mechanical) 2. AM modulated 100 kHz by 8 kHz. Courtesy: Manfactured by International Electronics Ltd Marketd by Signetics Electronics Ltd. Bombay



Cursor 1 and Cursor 2 Amplitudes / Frequencies / Freq Difference / Amplitude Diff / Harmonic Content Ratio etc. are displayed and will be computed dynamically as the cursors are moved. The "START" and "STOP" frequency denote the min and max frequency range of the display. The STOP frequency can range from 1 Hz to 10 MHz. Thus the minimum STEP or frequency resolution possible is 0.002 Hz. Display can be in Lin/Log. Horizontal Zoom level is 500.

Fig. 9.18 Set-up of FFT spectrum analysis

Review Questions

1. Define a wave analyzer. List different types of wave analyzers.
2. Explain with a diagram the operation of a basic wave analyzer.
3. Explain with a diagram the operation of a frequency selective wave analyzer.
4. Explain heterodyning. State the working principle of a heterodyne wave analyzer.
5. Describe with a diagram the operation of a heterodyne wave analyzer.
6. Explain with help of a block diagram the operation of an AF wave analyzer.
7. Describe the term real time analysers.
8. Explain with a diagram the working principle of a spectrum analyzer.
9. Explain with help of a block diagram the operation of a spectrum analyzer. State applications of a spectrum analyzer.
10. Differentiate between AF wave analyzer and RF wave analyzer.
11. Differentiate between wave analyzer and spectrum analyzer.
12. Define distortion. Define harmonics and the term 'total harmonic distortion'.
13. Explain the meaning of distortion factor. Explain how these distortion factors can be measured.
14. Why is it necessary to measure distortion? List various methods used for measurement of harmonic distortion.
15. Define a distortion analyzer. State the working principle of a distortion analyzer.
16. State different types of harmonic distortion analyzer.
17. Explain with a diagram, the suppression method of a harmonic distortion analyzer.
18. Explain with a diagram, the Wien bridge method of a harmonic distortion analyzer.
19. Describe with a diagram the operation of a harmonic distortion analyzer using a bridged-T network.

20. Explain the procedure of measurement of a harmonic distortion analyzer using a bridged-T type.
21. Describe the operation of a distortion analyzer using resonance to suppress the fundamental frequency
22. Compare resonance bridge to a Wien bridge harmonic distortion analyzer.
23. Compare Wien Bridge to a bridged-T network type harmonic distortion analyzer.
24. State applications of a distortion factor meter.
25. Explain the front panel controls and applications of a distortion factor meter in trouble shooting.
26. Describe the causes of harmonic distortions.
27. Explain the basic principle of a digital Fourier analyzer.
28. Describe with diagram the operation of a digital Fourier analyzer.
29. Explain in brief the operation of a practical FFT spectrum analyzer.

Multiple Choice Questions

1. Wave analyzers are used in the frequency range of
   (a) VHF (b) UHF
   (c) lower RF (d) higher RF
2. Wave analyzers are used to measure the
   (a) amplitude and phase
   (b) phase and frequency
   (c) amplitude and frequency
   (d) frequency band
3. Wave analyzers are also called a
   (a) phase meters
   (b) frequency selective voltmeter
   (c) distortion analyzer
   (d) spectrum analyzer
4. A heterodyne wave analyzer operates on the principle
   (a) mixing (b) amplification
   (c) addition (d) subtraction
5. A wave analyzer consists of
   (a) RC circuit (b) LC circuit
   (c) oscillator (d) rectifier
6. The bandwidth of a wave analyzer is
   (a) wide (b) narrow (c) medium
7. A spectrum analyzer works in
   (a) time domain (b) amplitude
   (c) frequency domain (d) phase
8. A spectrum analyzer uses at the output a
   (a) frequency meter (b) TVM
   (c) rectifier (d) circuit
9. The frequency axis in a spectrum analyzer is the
   (a) X-axis (b) Y-axis (c) Z-axis
10. A spectrum analyzer is used to display
    (a) frequency band spectrum
    (b) amplitude
    (c) time (d) phase
11. A distortion is defined as
    (a) unwanted frequency
    (b) unwanted amplitude
    (c) change in shape of the waveform
    (d) unwanted signal
12. A distortion analyzer measures the total
    (a) average power (b) RMS power
    (c) peak power (d) dc power

Further Reading

1. Terman and Petit, *Electronic Measurements*, McGraw-Hill, 1952.
2. Jones, *Instrument Technology*, Vol. 4, *Instrumentation Systems*, B.E. Notlingk Butterworth, 1987.
3. *Handbook of Electronic Measurement*, Volumes I & II, Polytechnic Inst. of Brooklyn, 1956 (Microwave Research Inst.).
4. Oliver Cage, *Electronic Measurements and Instrumentation*, McGraw-Hill, 1975.

Measuring Instruments

Chapter 10

INTRODUCTION 10.1

Meters and measuring instruments play an important role in all phases of electronics. Meters helps us to determine how an electronic circuit, is performing.

The most fundamental electrical measurements are those of voltage, current and impedance. The instruments used in measuring these quantities form the building blocks of more complex equipment used in power, frequency, attenuation, and other special measurements.

A measuring device converts a primary indication into some form of energy that can be easily displayed on a scale.

OUTPUT POWER METERS 10.2

The output power Wattmeter is designed to directly measure the output power in an arbitrary load. The instrument provides a set of resistive loads to be selected for power measurements. In addition to power, the output meter can be used to measure impedance and frequency response characteristics. A simple circuit is illustrated in Fig. 10.1.

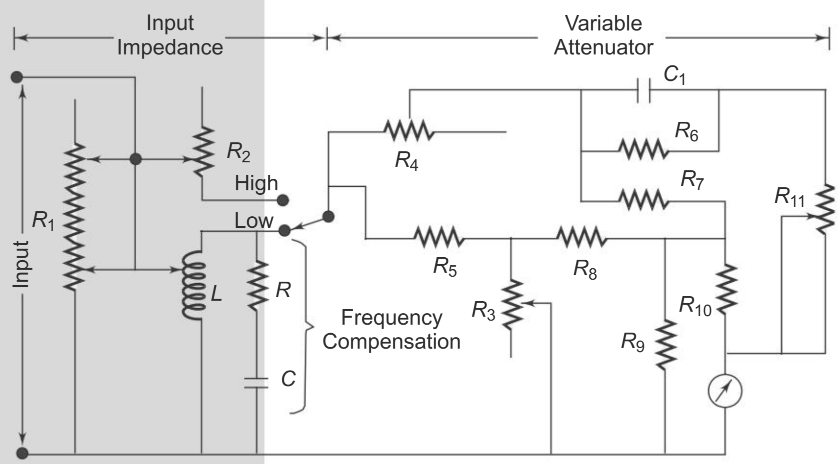


Fig. 10.1 Output-meter circuit

The input impedance network consist of two tapped resistances and a coil. The input impedance can be varied in steps from 2.5 Ω to 40 kΩ. At low impedance values, the coil shunts a portion of R_1 – an increase in resistance results in fewer turns of the coil. This arrangement keeps the meter reading proportional to the energy dissipated by the resistor. At high impedance values, the coil is replaced by another tapped resistance R_2.

The RC frequency compensator is in parallel with the coil. At low frequency, the capacitive reactance X_c is high; it decreases as frequency increases. This compensates for the losses of the coil at low frequencies. The frequency range is generally between 20 Hz and 20 kHz. R_3 is the control in a variable *T*-network. This amounts to a variable meter shunt, which is used to extend the range of the meter. The lowest meter range is normally 5 mW, but it can be extended in decade steps of 50 mW, 500 mW, etc.

The remaining circuit is a combination of a calibration and frequency compensation network. A meter of this type may have a midscale accuracy of ± 2% at 1 KHz. Over the frequency range of 20 Hz – 15 KHz, the accuracy may be within ± 2% of the 1 kHz value.

The input meter is subjected to an waveform error when the input is other than sinusoidal. (Practical instruments for measuring the power output of oscillators, amplifiers, transformers, transducers and low frequency lines use an input impedance of a tapped transformer with 48 impedance setting). It can be used to measure output impedance by adjusting the maximum power. It can also be used to check the frequency response characteristics of audio frequency devices.

FIELD STRENGTH METER 10.3

The field strength meter is used to measure the radiation intensity from a transmitting antenna at a given location. With its own small antenna, it is essentially a simple receiver with an indicator.

The wavemeter circuit with a rectifier-meter indicator, as shown in Fig. 10.2(a), is often equipped with a small whip antenna, and is called as a Field strength meter. (Although it is possible to obtain an indication with this setup when the whip antenna of the meter is positioned fairly close to the transmitting antenna, the sensitivity is generally not high enough for use with ordinary low-powered transmitters or test radiations.)

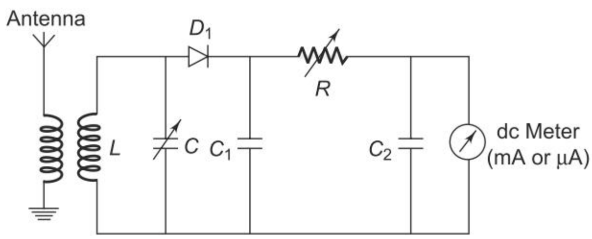


Fig. 10.2 (a) Field strength meter

The field strength measurement should be made at a distance of several wavelengths from the transmitting antenna, to avoid misleading readings when the pickup is obtained from a combination of radiation field with the induction field close to the transmitter. To enable the wavemeter combination to act as a field strength meter, greater sensitivity can be easily obtained with the addition of a transistor dc amplifier, as shown in Fig. 10.2(b). (The coil L is held near an oscillating circuit to provide loose coupling, and the capacitor C is tuned to resonance. When the dial of the variable capacitor is calibrated in terms of frequency, the unit becomes very useful as a rough frequency meter. Since the wavemeter is tuned to resonance, it will absorb the greatest amount of energy, and cause a detectable change in an indicating meter of the transmitter.)

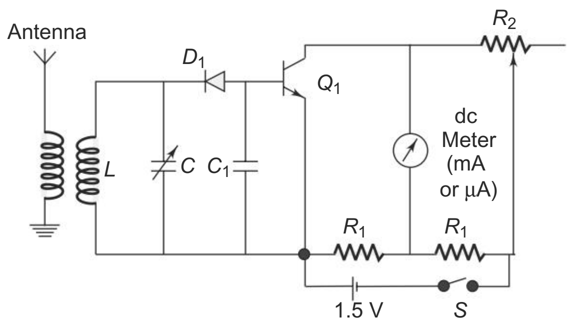


Fig. 10.2 (b) Field strength meter (transistor)

The transistor provides ample current gain, so that satisfactory sensitivity is obtained. The transistor is connected in a common emitter configuration. With no signal being received, the quiescent current is balanced out by the back up current, through the variable resistor R_2. This zero balance should be checked at intervals, since the quiescent current is sensitive to temperature changes.

The collector current through the meter provides an indication of the strength of the RF wave being picked up. This current is not strictly proportional to the field strength, because of the combined non-linearities of the semiconductor diode and transistor. However, the response is satisfactory for the relative comparison of field strength.

STROBOSCOPE 10.4

The stroboscopic principle uses a high intensity light which flashes at precise intervals. This light may be directed upon a rotating or vibrating object. The stroboscopic effect is apparent when the rotational or vibratory speed is in a proper ratio with the frequency of the light flashes.

Most stroboscopes consist of an oscillator, a reed, and a flasher, as illustrated in Fig. 10.3. The oscillator provides trigger pulses to the flasher mechanism to control the flashing rate. The oscillator is generally an externally triggered multivibrator. The vibrating reed serves as a reference for accurately calibrating the stroboscope. The reed is driven from the ac lines and vibrates at 7200 times per minute. This steady rate is used to standardise the calibration scale over a narrow range. The flasher produces the illumination for the measurements.

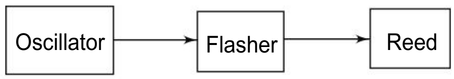


Fig. 10.3 Basic stroboscope block diagram

The flasher tube is fired by a capacitor discharge, which is, in turn, controlled by trigger pulses from the oscillator. The tube is filled with a suitable inert gas which produces light when it is ionised. The tube life ranges from 200 to 1000 hours, depending upon the operating conditions.

When the frequency of movement exactly matches the stroboscope frequency, the moving object is viewed clearly only once during each revolution. This causes the moving object to appear as a single stationary image. A stationary image also appears when the speed of rotation is some exact multiple of the stroboscope frequency. The highest scale reading that produces the single still image is the fundamental frequency.

Multiple still images appear when the stroboscope frequency is some multiple of the rotation frequency. In this case the light flashes more than once during each rotation of the object. (The radial line at the end of the shaft may appear as several equally spaced lines. If lamp frequency is twice the rotational frequency, two images are produced, 180° apart. If the lamp frequency is three times the rotational frequency, three images appear, each at a spacing of 120°.)

Moving images are obtained when the light frequency and rotational frequency are not synchronised. When the image appears to rotate in a direction opposite to that of actual rotation, the rotation frequency is less than the flasher frequency. When it appears to rotate in the same direction as the actual rotation, the rotation frequency is higher than the flasher frequency.

A stroboscope may be used to check motor or generator speeds ranging from 60 to 1,000,000 rpm. The stroboscope is highly versatile, uses no power from the circuit being measured and when calibrated, has an accuracy as close as 0.1%. (Some scopes, use the line frequency for calibration. The flash lamp and reflector assembly rotates 360° for maximum flexibility. The case may be mounted on a tripod. The flash rate is 110 to 150,000 flashes per minute, enabling measuring speeds of up to 1,000,000 rpm. The light output varies with the flash rate, from 3 μs to 0.5 μs.)

PHASE METER 10.5

Figure 10.4 shows a phase sensitive detector (or phase meter) for comparing an ac signal with a reference signal.

The detector produces a rectified output, which is fed to a dc meter, to illustrate clearly that the output of the phase sensitive detector swings the zero centre pointer in one direction for an in-phase error voltage and in the opposite direction for an out-phase condition. Thus, the function of this dual rectifier circuit is to deflect the zero centre galvanometer (or dc voltmeter) not only to indicate the value of the signal voltage V_s (that is, a measure of the error of imbalance), but also the direction of this error, and the phase polarity of the error compared to a reference voltage. Phase polarity implies that the detector distinguishes only between in phase and 180° out of phase conditions, without regard for other phase angles.

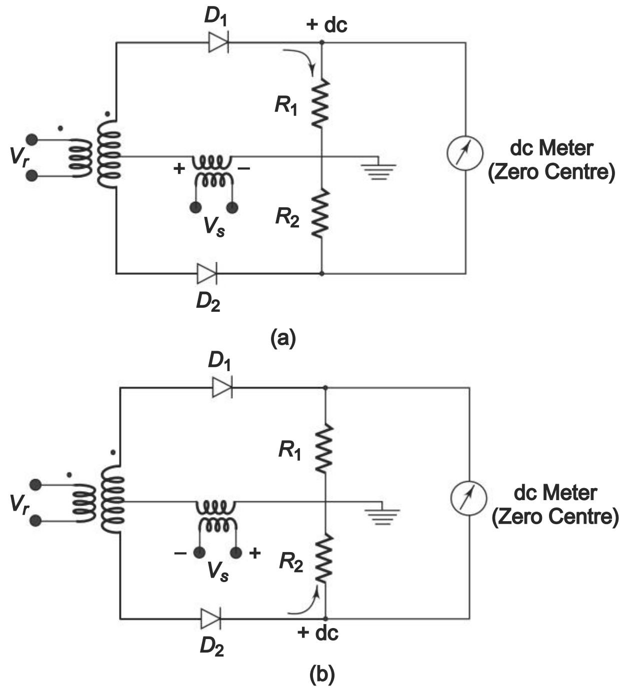


Fig. 10.4 Phase sensitive detector (a) Positive half (b) Negative half

The circuits of Figs 10.4(a) and (b) follows the action for a signal voltage V_s, which is in phase with the reference voltage V_r, starting with an initial condition when the input signal V_s is zero.

In Fig. 10.4(a), for the first half cycle the instantaneous polarity of the reference voltage V_r causes the rectified current to flow through the conduction rectifier D_1, producing a positive voltage to ground across R_1 and a tendency for the meter to deflect to the right.

On the second half cycle, Fig. 10.4(b), the instantaneous polarity of the reference voltage V_r, causes an equal rectified current to flow through diode D_2, producing an equal tendency for the meter to deflect to the left. Since these two equal and opposite tendencies are averaged over the full cycle, the galvanometer reads zero over the full cycle, with input $V_s = 0$.

When an input signal V_s is applied, it either aids or opposes the reference voltage, depending upon whether it is in phase or out of phase with it. If V_s is in phase with V_r, the signal voltage will aid the instantaneous ac voltage in the upper half of the transformer secondary, producing a larger current through D_1

and a larger dc output voltage on the first half. D_2 does not conduct unless V_s is greater than V_r, so that the voltage across R_2 is the rectified result of $V_s - V_r$, and that across R_1 is $V_s + V_r$.

On the other half cycle, the signal voltage is in the opposite direction. Diode D_1 will not conduct in the upper half and the signal voltage will oppose the instantaneous ac voltage, to produce a smaller dc voltage across R_2. The galvanometer therefore deflects to the right in proportion to the magnitude of the in-phase input signal V_s. Similarly, if V_s is 180° out of phase with V_r, the voltages add on the lower half on the transformer secondary, and the galvanometer deflects to the left in proportion to the magnitude of that input signal.

VECTOR IMPEDANCE METER (DIRECT READING) 10.6

If some knowledge of the reactive and resistive factors is needed, in addition to obtaining a direct reading of the magnitude of the impedance (Z), a test method for determining the vector impedance may be employed.

This method determines Z in polar form, that is, it gives the magnitude $|Z|$ and the phase angle (θ) of the impedance being tested, rather than the individual resistance and reactance in rectangular form, ($R + j\,X$). The test circuit is shown in Fig. 10.5. Two resistors of equal value R is used. The voltage drop across R_{AB} and R_{BC}, that is E_{AB} and E_{BC} will be equal (each value is equal to half the supply voltage, E_{AC}). Since the same current I_1 flows through the variable standard resistor R_{st}, the unknown impedance in series. The magnitude of Z_x can be determined by the equal deflection method by obtaining equal voltage drops across R_{st} and Z_x, i.e. E_{AD} and E_{DC}, and reading the calibrated standard resistor R_{st}, required to produce this condition.

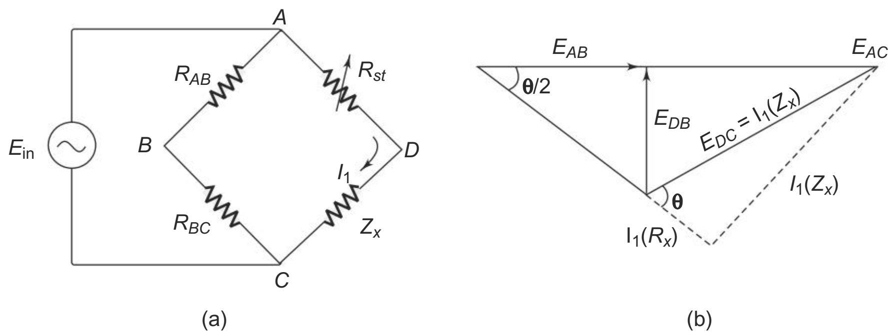


Fig. 10.5 (a) Vector impedance method (b) Vector diagram

The phase angle θ of the impedance Z_x can be obtained from the reading of the voltage at points B and D, that is, E_{DB}. The deflection of the meter will be found to vary with the Q of the unknown impedance Z_x. The VTVM ac voltage reading will vary from 0 V, when the phase angle of 0° ($Q = 0$) to the maximum voltage, with an angle of 90° (Q = infinite). The angle between the voltages E_{AB} and E_{AD} is half the phase angle θ, since E_{AD} is made equal to E_{DC}.

$$\frac{\theta}{2} = \tan\frac{E_{DB}}{E_{AB}}$$

Since E_{AB} is known to be half the known input voltage E_{in}, the voltmeter reading of E_{DB} can be interpreted in terms of $\theta/2$, and hence the phase angle θ of the unknown Z_x can be determined.

While this method for obtaining both Z and θ is approximate because of the crowding caused by the non-linear relation, it is useful for obtaining a first approximation. A commercial vector impedance meter is used for greater accuracy.

10.6.1 Commercial Vector Impedance Meter

A commercial instrument that measures impedance directly in the polar form, giving the magnitude of Z in ohms, at a phase angle θ, requires only one balancing control for both values.

It measures any combination of R, L and C, and includes not only pure resistive, capacitive or inductive elements but also complex impedances. Since the determination of magnitude and angle requires only one balance control, the awkward condition of sliding balance, frequently encountered when measuring low Q reactors with conventional bridge circuits, which necessitates so much successive adjustments, is avoided.

Measurements of impedances ranging from 0.5 – 100,000 Ω can be made over the frequency range from 30 Hz to 40 kHz, when supplied by an external oscillator. Internally generated frequencies of 60 Hz, 400 Hz or 1 kHz are available. At these internal frequencies and external frequencies up to 20 kHz, the readings have an accuracy of ± 1% for the magnitude of Z and ± 2% for θ.

The fundamental circuit, which is basic for both Z and phase angle measurement, is shown in Figs 10.6(a) and (b).

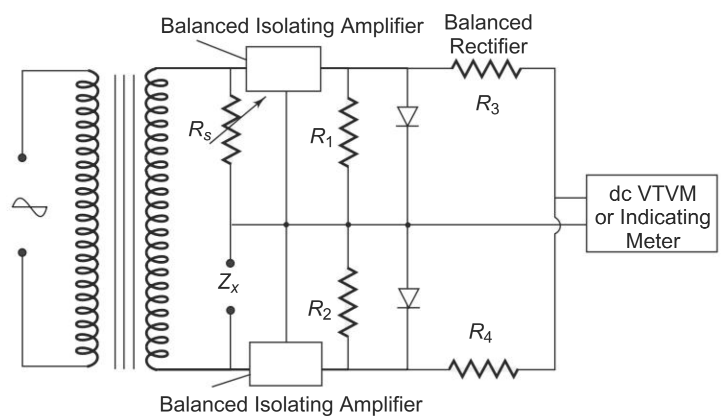


Fig. 10.6 (a) Commercial vector impedance meter (Magnitude of impedance measurement)

In both parts, the measurement makes use of the equal deflection method, by comparing the voltage drop across the unknown Z to the drop across a standard resistance, with the same current in both.

In the impedance measuring circuit of Fig. 10.6(a), Z_x is the unknown impedance and the variable resistance R_s is the standard resistance, which is varied by the calibrating impedance dial. The dial is adjusted until the voltage drops across Z_x and R_s are equal. Each voltage drop is amplified in the two sections of the balanced amplifier and applied to each section of a dual rectifier. The algebraic sum of the rectified outputs will then be zero, as indicated by the null reading of the dc VTVM, regardless of the phase angle of Z_x since rectified voltage depends only on the magnitude $|Z|$ of the unknown Z. This unknown Z, in ohms, is read directly on the dial of the variable standard R_s.

The circuit shown in Fig. 10.6(b) is used for the measurement of phase angle, after the Z balance has been obtained.

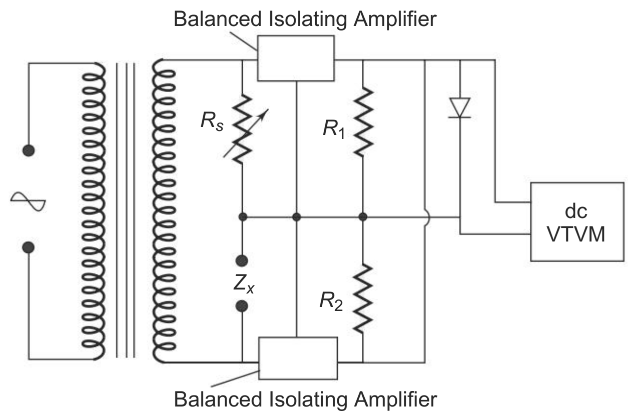


Fig. 10.6 (b) Vector impedance meter circuit for measurement of phase angle

With the switch is in the calibration position, the injection voltage is calibrated by adjusting it for full scale deflection on the indicating meter or the VTVM. The function switch is then set to the phase position. In this position, the function switch of the instrument parallels the output of the balanced amplifier, before rectification. The sum of ac output voltages from the amplifiers is now a function of the vector difference between the ac voltages impressed on the amplifiers.

The rectified voltage resulting from this vector difference is shown on the dc VTVM, as a measure of the phase angle between the voltage across Z_x and R_s, which are equal in magnitude but different in phase. Thus the meter is able to indicate direct reading values for the phase angle. If required, this angle can be converted to measure the corresponding values for dissipation factor D and quality factor Q.

Where it is necessary to determine the phase angle to a high degree of accuracy, a phase meter is usually employed, e.g. in servos and precise control applications.

Q METER 10.7

The overall efficiency of coils and capacitors intended for RF applications is best evaluated using the Q value. The Q meter is an instrument designed to measure some electrical properties of coils and capacitors. The principle of the Q meter is based on series resonance; the voltage drop across the coil or capacitor is Q times the applied voltage (where Q is the ratio of reactance to resistance, X_L/R). If a fixed voltage is applied to the circuit, a voltmeter across the capacitor can be calibrated to read Q directly.

At resonance $X_L = X_C$ and $E_L = I X_L$, $E_C = I X_C$, $E = I R$

where E — applied voltage, E_C — capacitor voltage
E_L — inductive voltage, X_L — inductive reactance
X_C — capacitive reactance, R — coil resistance
I — circuit current

Therefore
$$Q = \frac{X_L}{R} = \frac{X_C}{R} = \frac{E_C}{E}$$

From the above equation, if E is kept constant the voltage across the capacitor can be measured by a voltmeter calibrated to read directly in terms of Q.

A practical Q meter circuit is shown in Fig. 10.7.

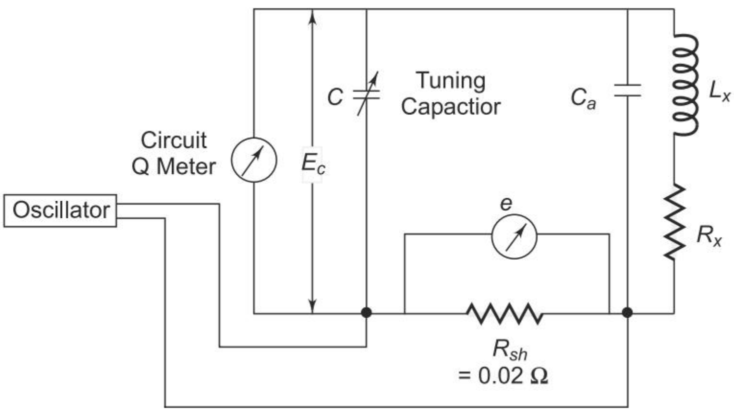


Fig. 10.7 Circuit diagram of a Q-meter

The wide range oscillator, with frequency range from 50 kHz to 50 MHz, delivers current to a resistance R_{sh} having a value of 0.02 Ω. This shunt resistance introduces almost no resistance into the tank circuit and therefore represents a voltage source of a magnitude e with a small internal resistance. The voltage across the shunt is measured with a thermocouple meter. The voltage across the capacitor is measured by an electronic voltmeter corresponding to E_c and calibrated directly to read Q.

The oscillator energy is coupled to the tank circuit. The circuit is tuned to resonance by varying C until the electronic voltmeter reads the maximum value.

The resonance output voltage E, corresponding to E_c, is $E = Q \times e$, that is, $Q = E/e$. Since e is known, the electronic voltmeter can be calibrated to read Q directly.

The inductance of the coil can be determined by connecting it to the test terminals of the instrument. The circuit is tuned to resonance by varying either the capacitance or the oscillator frequency. If the capacitance is varied, the oscillator frequency is adjusted to a given frequency and resonance is obtained. If the capacitance is pre-setted to a desired value, the oscillator frequency is varied until resonance occurs. The Q reading on the output meter must be multiplied by the index setting or the "Multiply Q by" switch to obtain the actual Q value. The inductance of the coil can be calculated from known values of the coil frequency and resonating capacitor (C).

$$X_L = X_C,\ f = \frac{1}{2\pi\sqrt{LC}} \quad \text{or} \quad L = \frac{1}{(2\pi f)^2 C} \tag{10.1}$$

The Q indicated is not the actual Q, because the losses of the resonating capacitor, voltmeter and inserted resistance are all included in the measuring circuit. The actual Q of the measured coil is somewhat greater than the indicated Q. This difference is negligible except where the resistance of the coil is relatively small compared to the inserted resistance R_{sh}.

10.7.1 Factors that May Cause Error

1. At high frequencies the electronic voltmeter may suffer from losses due to the transit time effect. The effect of R_{sh} is to introduce an additional resistance in the tank circuit, as shown in Fig. 10.8.

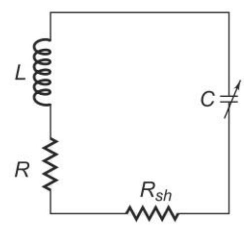


Fig. 10.8 Effect of R_{sh} on Q

$$Q_{\text{act}} = \frac{\omega L}{R} \quad \text{and} \quad Q_{obs} = \frac{\omega L}{R + R_{sh}}$$

$$\therefore \quad \frac{Q_{\text{act}}}{Q_{\text{obs}}} = \frac{R + R_{sh}}{R} = 1 + \frac{R_{sh}}{R}$$

$$\therefore \quad Q_{\text{act}} = Q_{\text{obs}}\left(1 + \frac{R_{sh}}{R}\right)$$

where Q_{act} = actual Q
Q_{obs} = observed Q

To make the Q_{obs} value as close as possible to Q_{act}, R_{sh} should be made as small as possible. An R_{sh} value of 0.02 – 0.04 Ω introduces negligible error.

2. Another source of error, and probably the most important one, is the distributed capacitance or self capacitance of the measuring circuit. The presence of distributed or stray capacitances modifies the actual Q and

the inductance of the coil. At the resonant frequency, at which the self capacitance and inductance of the coil are equal, the circuit impedance is purely resistive—this characteristic can be used to measure the distributed capacitance.

One of the simplest methods of determining the distributed capacitance (C_s) of a coil involves the plotting of a graph of $1/f^2$ against C in picofarads.

The frequency of the oscillator in the Q meter is varied and the corresponding value of C for resonance is noted. $1/f^2$ is plotted against C in picofarads, as shown in Fig. 10.9(a). The straight line produced to intercept the X-axis gives the value of C_s, from the formula given on the next page. The value of the unknown inductance can also be determined from the equation.

$$L = \frac{\text{slope}}{4\pi^2}, \text{ therefore slope} = 4\pi^2 L$$

and
$$f = \frac{1}{2\pi\sqrt{L(C + C_s)}}$$

Therefore
$$\frac{1}{f^2} = 4\pi^2 L(C + C_s)$$

If
$$\frac{1}{f^2} = 0, \text{ then } C = -C_s.$$

Another method of determining the stray or distributed capacitance (C_s) of a coil involves making two measurements at different frequencies. The capacitor C of the Q meter is calibrated to indicate the capacitance value. The test coil is connected to the Q meter terminals, as shown in Fig. 10.9(b).

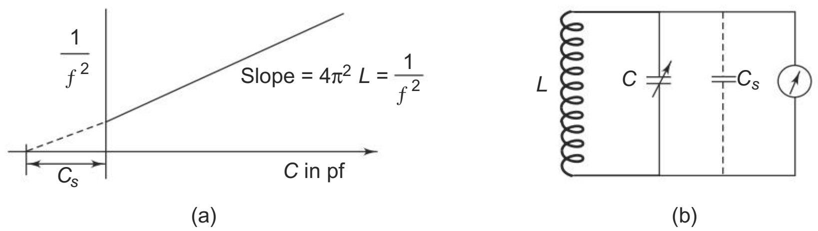


Fig. 10.9 Measurement of stray capacitance

The tuning capacitor is set to a high value position (to its maximum) and the circuit is resonated by varying the oscillator frequency. Suppose the meter indicates resonance and the oscillator frequency is found to be f_1 Hz and the capacitor value to be C_1.

The oscillator frequency, of the Q-meter is now increased to twice the original frequency, that is, $f_2 = 2f_1$, and the capacitor is varied until resonance occurs at C_2. The resonant frequency of an LC circuit is given by

$$f = \frac{1}{2\pi\sqrt{LC}}$$

Therefore, for the initial resonance condition, the total capacitance of the circuit is $(C_1 + C_s)$ and the resonant frequency equals

$$f_1 = \frac{1}{2\pi\sqrt{L(C_1 + C_s)}}$$

After the oscillator and the tuning capacitor are varied for the new value of resonance, the capacitance is $(C_2 + C_s)$, therefore

$$f_2 = \frac{1}{2\pi\sqrt{L(C_2 + C_s)}}$$

But $f_2 = 2f_1$.
Therefore

$$\frac{1}{2\pi\sqrt{L(C_2 + C_s)}} = \frac{2}{2\pi\sqrt{L(C_1 + C_s)}}$$

$$C_1 + C_s = 4(C_2 + C_s)$$

$$C_1 + C_s = 4C_2 + 4C_s$$

$$C_1 = 4C_2 + 3C_s$$

$$3C_s = C_1 - 4C_2$$

$$C_s = \frac{C_1 - 4C_2}{3} \tag{10.2}$$

The distributed capacitance can be calculated using the equation above.

Example 10.1 *The self-capacitance of the coil is measured by using the method outlined in the previous section. The first measurement is at f_1 = 2MHz and the value of capacitance C_1 = 600 pf. The second measurement is at f_2 = $2f_1$ = 4 MHz and the value of capacitance C_2 = 100 pf. Find the value of distributed capacitance and the value of L.*

Solution Using the equation

$$C_s = \frac{C_1 - 4C_2}{3} = \frac{600\text{ pf} - 4(100\text{ pf})}{3} = \frac{200\text{ pf}}{3} = 66.67\text{ pf}$$

Therefore $C_s = 66.67$ pf

The resonant frequency is given by

Step 1:
$$f_1 = \frac{1}{2\pi\sqrt{L(C_1 + C_s)}}$$

Therefore

Step 2:
$$L = \frac{1}{4 \times \pi^2 \times f_1^2 \times (C_1 + C_s)}$$

$$L = \frac{1}{4 \times 9.8696 \times 4 \times 10^{12} \times (600\ \text{pf} + 66.67\ \text{pf})}$$

$$L = \frac{1}{4 \times 9.8696 \times 4 \times (666.67)} = \frac{1}{16 \times 9.8696 \times 666.67}$$

$$L = 9.498\ \mu\text{H}$$

Example 10.2 *The self capacitance of a coil is measured by using the method outlined in the previous section. The first measurement is at $f_1 = 1$ MHz and $C_1 = 500$ pf. The second measurement is at $f_2 = 2$ MHz and $C_2 = 110$ pf. Find the distributed capacitance. Also calculate the value of L.*

Solution The distributed capacitance is given by the equation

$$C_s = \frac{C_1 - 4C_2}{3}$$

$$C_s = \frac{500\ \text{pf} - 4(110\ \text{pf})}{3}$$

$$C_s = \frac{500\ \text{pf} - 440\ \text{pf}}{3} = \frac{60\ \text{pf}}{3}$$

$$C_s = 20\ \text{pf}$$

The resonant frequency is given by the formula

$$f_1 = \frac{1}{2\pi\sqrt{L(C_1 + C_s)}}$$

$$L = \frac{1}{4\pi^2 f_1^2 (C_1 + C_s)}$$

$$L = \frac{1}{4 \times (3.14159)^2 \times (1 \times 10^6)^2 \times (500 \times 10^{-12} + 20 \times 10^{-12})}$$

$$L = \frac{1}{4 \times 9.8696 \times 10^{12} \times (520 \times 10^{-12})}$$

$$L = \frac{1}{4 \times 9.8696 \times 520}$$

$$L = \frac{1}{20528.777} = 48.712 \times 10^{-6}\ \text{H}$$

$$L = 48.712\ \mu\text{H}$$

Example 10.3 *Calculate the value of the self capacitance when the following measurements are performed.*

$f_1 = 2$ MHz and $C_1 = 500$ pf; $f_2 = 6$ MHz and $C_2 = 50$ pf

Solution Given that $f_2 = 3f_1$

$$\frac{1}{2\pi\sqrt{L(C_2+C_s)}} = \frac{3}{2\pi\sqrt{L(C_1+C_s)}}$$

$$C_1 + C_s = 9\,(C_2 + C_s)$$

$$C_1 + C_s = 9C_2 + 9C_s$$

$$C_1 - 9C_2 = 9C_s - C_s$$

$$C_1 - 9C_2 = 8C_s$$

Therefore
$$C_s = \frac{C_1 - 9C_2}{8}$$

$$C_s = \frac{500\text{ pf} - 9(50\text{ pf})}{8}$$

$$C_s = \frac{500\text{pf} - 450\text{ pf}}{8}$$

$$C_s = \frac{50\text{ pf}}{8} = 6.25\text{ pf}$$

10.7.2 Impedance Measurement Using Q Meter

An unknown impedance can be measured using a Q meter, either by series or shunt substitution method. If the impedance to be measured is small, the former is used and if it is large the latter method is used.

In the Q meter method of measurement of Z, the unknown impedance Z_x is determined by individually determining its components R_x and L_x. The technique utilises an LC tank of a Q meter, L being an externally connected standard coil.

Figure 10.10(a) shows the method of series substitution while Fig. 10.10(b) shows the shunt substitution method.

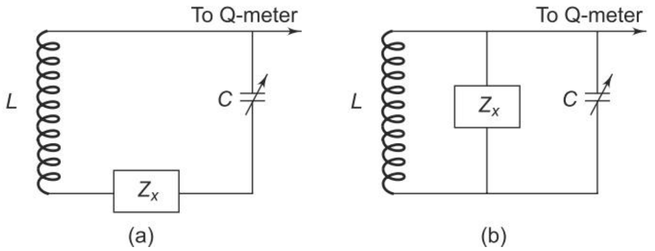


Fig. 10.10 (a) Series substitution (b) Shunt substitution

Referring to Fig. 10.10(a), the unknown impedance is shorted or otherwise not connected and the tuned circuit is adjusted for resonance at the oscillator frequency. The value of Q and C are noted. The unknown impedance is then connected, the capacitor is varied for resonance, and new values Q' and C' are noted.

From part 1, we have $\omega L = 1/\omega C$ (10.3)

From part 2, we have $\omega L + X_x = \dfrac{1}{\omega C'}$ (10.4)

Subtracting Eq. (10.4) from Eq. (10.3), we have

$$X_x = \frac{1}{\omega C'} - \frac{1}{\omega C} = \frac{1}{\omega C}\left(\frac{C - C'}{C'}\right) = \frac{1}{\omega}\left(\frac{C - C'}{CC'}\right)$$

Since $R' = R + R_x$, $R_x = R' - R$, where R is the resistance of the auxillary coil.

$$R_x = R' - R = \frac{\omega L}{Q'} - \frac{\omega L}{Q} = \omega L\left(\frac{Q - Q'}{QQ'}\right)$$

The unknown impedance Z_x can be calculated from the equation

$$Z_x = R_x + j\,X_x$$

A positive value of X_x indicates inductive reactance and a negative value indicates capacitive reactance.

If Z_x is considerably greater than X_L, the unknown impedance is shunted across the coil and the capacitor, as shown in Fig. 10.10(b).

Y_x represents the shunt admittance of the unknown impedance. It consists of two shunt elements, conductance G_x and susceptance B_x. In this method, Y_x is disconnected and the capacitor C is tuned to the resonant value. At the oscillator frequency, the values of Q and C are noted. With Y_x connected, the capacitor is tuned again for resonance at the oscillator frequency and the new values Q' and C' are noted.

Hence $\quad Y_x = G_x + jB_x$

and $\quad B_x = \omega C - \omega C'$

also $\quad G_x = \dfrac{1}{\omega L}\left(\dfrac{Q - Q'}{QQ'}\right)$

Therefore $\quad Y_x = \dfrac{Q - Q'}{\omega LQQ'} + j\omega(C - C')$

The accuracy with which the reactance can be determined by the method of substitution is quite high. Error may mainly be because

(i) C' cannot be accurately determined since the resonance curve may be flat due to additional resistance, and

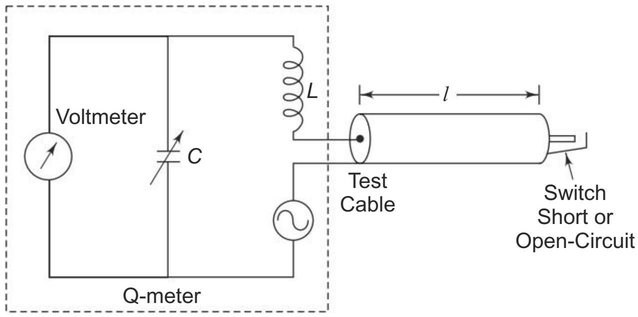
(ii) The stray inductance associated with the tuning capacitor causes errors at VHF.

The accuracy with which the resistance component of the unknown impedance is obtained is poor. If the losses in the unknown impedance are too small to introduce any change in the Q, the substitution method is quite satisfactory.

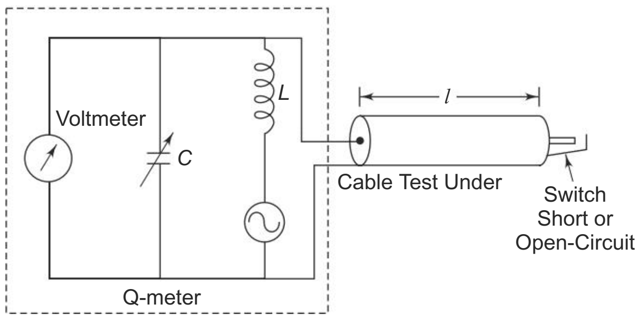
The substitution method can also be used for measuring the losses of the coil. It is not satisfactory for measuring the losses of an air-dielectric capacitor, since they are too small to be detected by this method.

10.7.3 Measurement of Characteristic Impedance (Z_0) of a Transmission Line Using Q Meter

Figure 10.11(a) shows a series substitution method for determining the characteristic impedance of a transmission line and Fig. 10.11(b) shows a shunt or parallel method of substitution for the same purpose.



(a)



(b)

Fig. 10.11 (a) Series substitution method of measurement of Z_0 (b) Shunt substitution Method of measurement of Z_0

In Fig. 10.11(a), the transmission line or cable under test is tuned for series resonance. Since the input impedance is low, the method of series substitution can be used to determine $Z_0 = R_0 + jX_0$ for the transmission line, as explained in the previous section.

The reactance/unit length of the line is the total reactance divided by the length l.

Series resonance occurs when the line is short-circuited and the line length is an even multiple of a $\lambda/4$ and when open-circuited an odd multiple of $\lambda/4$.

Parallel resonance occurs when the line is short-circuited and the length is an odd multiple of $\lambda/4$, or open-circuited it is an even multiple of $\lambda/4$.

10.7.4 Measurement of Q by Susceptance Method

The coil under test is connected in series with a calibrated low loss variable capacitor. With the use of the meter as an indicator, the circuit is tuned for resonance to the oscillator frequency, by tuning the variable capacitor to a value C_r, as shown in Fig. 10.12(a). The capacitor is then detuned to a value C_b on the low capacitance side of resonance at which the meter reading falls to 70.7% of the resonant voltage. Next, the capacitor is set on the higher capacitance side of resonance to a value C_a, where the voltmeter deflection again drops to 70.7% of the resonant voltage, as shown in Fig. 10.12(b).

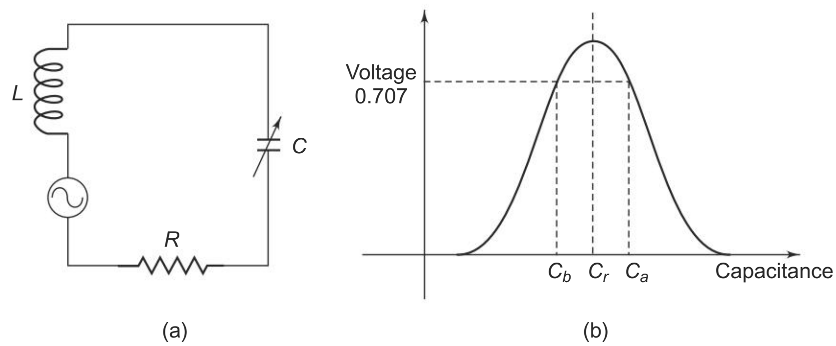


Fig. 10.12 (a) Susceptance method of Q measurement (b) Response curve

The points C_a, C_b and C_r are closer together when coil Q is high (sharp tuning) and far apart when Q is low (broad tuning). The value of Q can be then determined as follows.

C_a and C_b are the value of capacitances at the half power point and C_r is the value of the capacitance at resonance.

$\therefore \quad X_{Ca} = 1/\omega C_a$ and $X_{Cb} = 1/\omega C_b$

At half power points

$$\omega L - 1/\omega\, C_a = R \text{ and } 1/\omega\, C_b - \omega L = R$$

Adding the two equations, we have

$$\frac{1}{\omega C_b} - \frac{1}{\omega C_a} = 2R \qquad \frac{\omega C_a - \omega C_b}{\omega^2\, C_a\, C_b} = 2R$$

i.e. $$\frac{C_a - C_b}{\omega\, C_a\, C_b} = 2R \quad \text{But} \quad C_r^2 = C_a\, C_b$$

$\therefore$ $$\frac{C_a - C_b}{\omega\, C_r R C_r} = 2 \quad \text{But} \quad Q = \frac{1}{\omega\, C_r\, R}$$

Therefore $$\frac{Q(C_a - C_b)}{C_r} = 2$$

$$Q\,(C_a - C_b) = 2C_r$$

Therefore $$Q = \frac{2\,C_r}{C_a - C_b} \tag{10.5}$$

In this method, the coil under test is connected in series with a calibrated low loss capacitor. The frequency of the signal generator is kept at a suitable value and the output across the capacitor is measured by an electronic voltmeter. This method requires less expensive components than the Q meter.

LCR BRIDGE 10.8

10.8.1 Basic LCR Bridge (Skeleton Type)

A simple bridge for the measurement of resistance, capacitance and inductance may be constructed with four resistance decades in one arm, and binding post terminals to which external resistors or capacitors may be connected, to complete the other arms. Such a skeleton arrangement is useful in the laboratory, since it permits the operator to set up a number of different bridge circuits simply by plugging standards and unknown units into the proper terminals.

The schematic circuit diagram of a skeleton type bridge and its accessories for the measurement of R, L and C is shown in Figs 10.13(a) and 10.13(b).

In Fig. 10.13(a), R_b is self contained. The other arms are completed by connecting the unknown and standard component to terminals 1-2, 3-4, 7-8, 9-10, and 11-12, a null detector to 5-6, and a generator to 13-14. This bridge can be used for both ac and dc measurements.

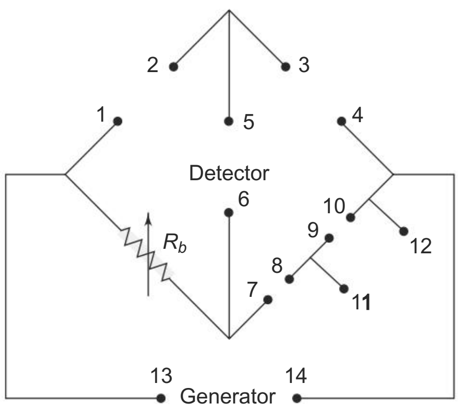


Fig. 10.13 (a) Skeleton type LCR bridge

By proper arrangement of the bridge arms, the Wheatstone's bridge shown in Fig. 10.14(a) may be set up for resistance measurement (ac and dc), a comparison circuit shown in Fig. 10.14(b) for measurement of C, and a Maxwell's circuit shown in Fig. 10.14(c) for the measurement of inductance L.

The skeleton bridge permits resistance measurements from 0.001 Ω to 11.11 MΩ, capacitance measurements from 1 pf (if stray capacitance permits) to 1111 μF and inductance measurements from 1 μH (if stray inductance and capacitance permit) to 111.1 H.

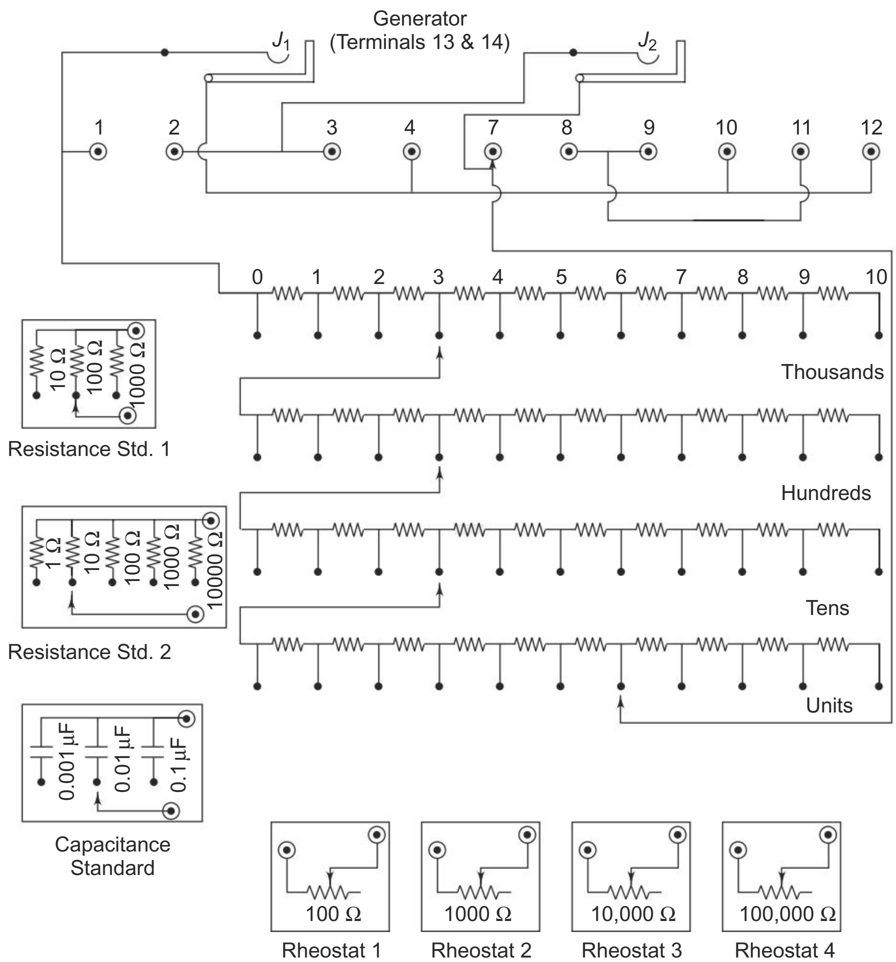


Fig. 10.13 (b) Complete circuit of the skeleton bridge and accessories

The bridge accessories are shown in Fig. 10.13(b). There are two switched resistance standard boxes (one 10 – 100 – 1000 Ω precision resistance and another 1 – 10 – 100 – 1000 – 10000 Ω precision resistance), one switched capacitor standard box containing 0.001 – 0.01 – 0.1 μF, a close tolerance capacitor, and four mounted dial calibrated rheostats having values of 100 – 1 k – 10 k – 100 kΩ.

For dc measurements 1.5 – 7.5 V batteries are used and connected to terminals 13 – 14. A centre zero 50 – 0 – 50 or 100 – 0 – 100 dc microammeter is connected to terminals 5 – 6.

For ac measurements of *R*, *L* and *C* connect a signal source or AF oscillator tuned to a frequency of 1 kHz to terminals 13 – 14 and a null detector (headphone, AC VTVM, CRO, etc) to terminal 5 – 6. The mode of operation for the measurement of *R*, *C* and *L* are as follows.

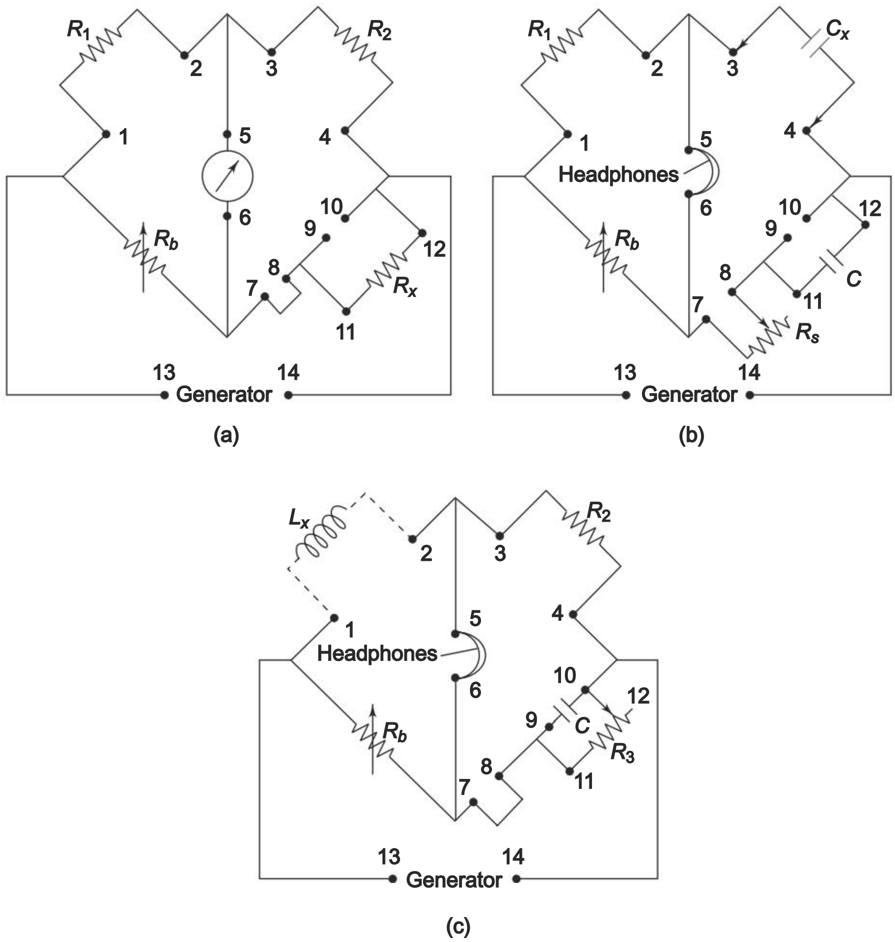


Fig. 10.14 (a) Wheatstone's bridge for resistance measurement (b) Capacitance comparison bridge for capacitance measurement (c) Maxwell's inductance bridge

For Resistance Measurements

1. Connect resistance standard 1 (set to 1000 Ω) to terminals 1–2.
2. Connect resistance standard 2 (set to 1 Ω) to terminals 3–4.
3. Terminals 11 – 12 are connected to the unknown resistance.
4. Connect a heavy wire jumper to terminals 7–8.
5. Terminals 9–10 are kept open circuit.
6. Connect generator to 13–14, detection is done by using a centre zero dc microammeter if the generator used is dc, or if an ac source is used the detector must be an ac type null detector.
7. Adjust the decade switches for null. If null is not obtained with the resistance standard settings given in steps 1 and 2, try other settings and repeat the decade adjustments.

8. At null, determine the unknown resistance R_x from the readings of the decades (R_b), resistance standard 1 and resistance 2 using the following equation.

$$R_x = \frac{R_2 R_b}{R_1} \tag{10.6}$$

The possible ranges are given in Table 10.1.

Table 10.1

Resistance 1 (Ω)	*Resistance 2 (Ω)*	*Range (Ω)*
1000	1	0.001– 11.11
1000	10	.01 – 111.11
1000	100	.1 – 1111
1000	1000	1.0 – 11110
1000	10000	10.0 – 111100
100	10000	100.0 – 1111000
10	10000	1000.0 – 11110000

For Capacitance Measurement

1. Connect resistance standard 1 (set to 1 kΩ) to terminals 1 – 2.
2. Connect capacitor standard (set to 0.001 μF) to terminals 11 – 12.
3. Connect the unknown capacitance to terminals 3 – 4.
4. Keep terminals 9 – 10 open-circuit.
5. Connect the 100 Ω rheostat to terminals 7 – 8.
6. Connect an ac generator to terminals 13 – 14 and a headphone or ac null detector to terminals 5 – 6.
7. Adjust the decade switches for null. If null is not obtained with the standard settings given in steps 1 and 2, try other standard resistance and capacitor settings.
8. At null determination the value of the unknown capacitor is determined from the readings of the decades R_b, resistance standard 1, and capacitor standard (C) by using the following equation.

$$C_x = \frac{C R_b}{R_1} \mu\text{F} \tag{10.7}$$

where R_1 and R_b are in ohms and C is in microfarads.

The possible ranges are given in Table 10.2.

Table 10.2

Resistance 1 (Ω)	*Capacitor standard (μ F)*	*Ranges (Ω)*
1000	0.001	1 – 11.110
1000	0.01	10 – 0.1111
100	0.01	100 – 1.1111
1000	1.00	0.001 – 11.11
100	1.00	0.01 – 111.1
10	1.00	0.1 – 1111

The power factor of the unknown capacitor can be determined from the rheostat dial reading (R_s), and from the value of the unknown capacitor by the following equation.

$$\text{Power factor (PF)} = 0.000628 \times f \times R_s \times C_x\ \% \qquad (10.8)$$

For Inductive Measurements

1. Connect the capacitor standard (set to 0.01 μF) to terminals 9 – 10.
2. Connect the resistance standard 2 (set to 100 Ω) to terminal 3 – 4.
3. Connect the 100 Ω rheostat to terminals 11 – 12.
4. Connect a heavy jumper to terminals 7 – 8.
5. Connect the unknown inductances to terminals 1 – 2.
6. Connect an ac generator to terminals 13 – 14 and also connect an ac detector to terminals 5 – 6.
7. Adjust the decade switches for null. Adjust the rheostat for a sharper null.
8. If null is not obtained with the standard *R* and *C* settings given in steps 1 and 2, try other settings.
9. At complete null, determine the unknown impedance (L_x) from the readings of the decades R_b, resistance standard 2 (R_2) and capacitor standard (C) by means of the following equation.

$$L_x = C\, R_b\, R_2\, H \qquad (10.9)$$

where R_2 and R_b are in Ω's and C in F's.

The possible ranges are given in Table 10.3.

Table 10.3

Resistance standard 2	*Capacitor standard (F)*	*Range*
100 Ω	0.01	1 μH – 11.11 mH
1000 Ω	0.01	10 μH – 111.1 mH
10 kΩ	0.01	100 μH – 1.111 H
10 kΩ	1.0	10 mH – 111.1 H

The complete circuit of an LCR bridge is shown in Fig. 10.15 (kit type bridge).

A wide range of resistance (ac and dc), inductance and capacitance measurements can be done using a Kit type impedance bridge.

This bridge measures inductance (L) from 1 μH – 100 H, capacitance (C) from 10 μμF – 100 μF, resistance (R) from 0.01 Ωs – 10 MΩ, dissipation factor (D) from 0.001 – 1 and Q from 1 – 1000. Resistance, capacitance and inductance units are read directly on the same dial scale, graduated 0 – 1, and are multiplied by settings of a multiplier switch.

Referring to Fig. 10.15, six inductance ranges are provided

(i) 10 – 100 μH	(ii) 50 μH – 10 mH
(iii) 0.5 – 100 mH	(iv) 5 mH – 1 H
(v) 50 mH – 10 H	(vi) 0.5 – 100 H

These ranges include all common inductances of coils of all types employed in electronics.

Also eight resistance ranges are provided.

(i) 0.01 – 1 Ω	(ii) 0.05 – 10 Ω
(iii) 0.5 – 100 Ω	(iv) 5 – 1000 Ω
(v) 50 – 10,000 Ω	(vi) 500 – 100,000 Ω
(vii) 5,000 – 1 MΩ	(viii) 50,000 – 10 MΩ

There are six capacitance ranges provided:

(i) 10 – 1000 μμF	(ii) 50 μμF – 0.01 μF
(iii) 0.0005 – 0.1 μF	(iv) 0.005 – 1 μF
(v) 0.05 – 10 μF	(vi) 0.5 – 100 μF

The two pole, eight position Multiplier Switch sets the bridge to the desired R, C or L range. This switch cuts the various precision resistors, R_1 to R_9 in or out of the circuit. The dial settings of the multiplier switch show the various R, C and L factors by which the settings of the main control dial must be multiplied to obtain the correct value of the component under test.

The function of the Detector Switch is to connect an appropriate null detector across the bridge output terminals. When this switch is thrown to its external position, the two terminals labelled external detector are connected across the bridge output, and an external null detector may be connected to these terminals. (Satisfactory external detectors are high resistance headphones, ac VTVMs, oscilloscopes, sensitive centre zero dc galvanometers, etc.). The galvanometer is used only in resistance measurements when the internal 6 V battery (or a higher voltage external battery) is used to power the bridge.

When the detector switch is thrown to its galvanometer position, the self contained centre zero (100 – 0 – 100) dc microammeter is connected, as a dc null detector across the bridge output. When the detector switch is thrown to its shunted galvanometer position, the microammeter is connected across the bridge output, but in parallel with the 100 Ω resistor R_{10}. This resistor decreases the microammeter sensitivity and acts to prevent meter damage when an unknown resistance is first checked.

Two self contained bridge power (signal) sources are employed. One is a General Radio 1000 Hz oscillator (an electromechanical type of oscillator) which supplies the signal voltage for inductance, capacitance and resistance measurements at ac. The second source is a 6 V battery used only to power the bridge when dc measurements of resistance are made with a microammeter (or an external galvanometer) as a null detector.

The function of the Generator Switch is to connect an appropriate power (signal) source across the input points of the bridge circuit. When this switch is in its 1 kHz position, the output of the self contained 1000 cycles per second oscillator is connected to the bridge input, and the 6 V battery is switched in automatically to drive the oscillator. When the switch is in its dc position, the self contained 6 V battery is switched across the bridge input circuit. When the Generator Switch is in its External position, the bridge input points are connected to the two terminals labelled external generator. A suitable external generator may now be connected to these terminals. When an ac type null detector is used, a satisfactory generator is an audio oscillator. When a self contained microammeter or external galvanometer is used as the bridge detector, a battery would be employed as the external generator.

The main control is a 10,000 Ω wire-wound rheostat having a logarithmic taper. The dial attached to this component is graduated 0 – 10 and is read directly in μμf, μF, μH, mH, H, Ω and MΩ, depending upon the setting of the multiplier switch. Thus the reading "4" on this dial is read as 400 mH if the bridge is set up for inductance measurement, and the multiplier switch is in position D, as shown in Fig. 10.15. But it would be read as 40,000 Ω if the bridge were set up for resistance measurement and the multiplier switch were in position F. Two separate pairs of unknown terminals are provided. Resistors under test are connected to the pair labelled R, and capacitors and coil to the pair labelled C-L.

The Selector Switch performs two functions. The first is to set up the bridge automatically for either resistance, capacitance or inductance measurements. For resistance, the circuit is a standard Wheatstone bridge, while for capacitance it is a conventional four arm bridge with capacitances in two legs and a resistance in the ratio arm. The Maxwell bridge circuit is employed for measuring inductors with Q factors of 10 or less, and the Hay bridge circuit for inductors with a Q factor higher than 10. The second function of the Selector Switch is to select the proper rheostat for reading the dissipation factor of capacitors or the Q of coils. When this switch is in its C – D position, the bridge is set up for measuring capacitance, and rheostat D is selected for dissipation factor readings from 0.001 – 0.1 (corresponding to a capacitor power factor ranging from 0.1 – 10%). When the switch is in its CDQ position, the bridge is set for capacitance measurements, and the rheostat DQ is selected for dissipation factor readings from 0.01 – 1. When the switch is in its LQ position, the bridge is set for inductance measurements, and the rheostat Q is selected for Q readings from 10 – 1000. When the switch is in its LDQ position, the bridge is set up for inductance measurement, and the rheostat DQ is selected for Q readings from 1 – 10. When the Selector Switch is in its R position, the bridge is set up for resistance measurement only, and the rheostats D, Q and DQ are automatically switched out of the circuit.

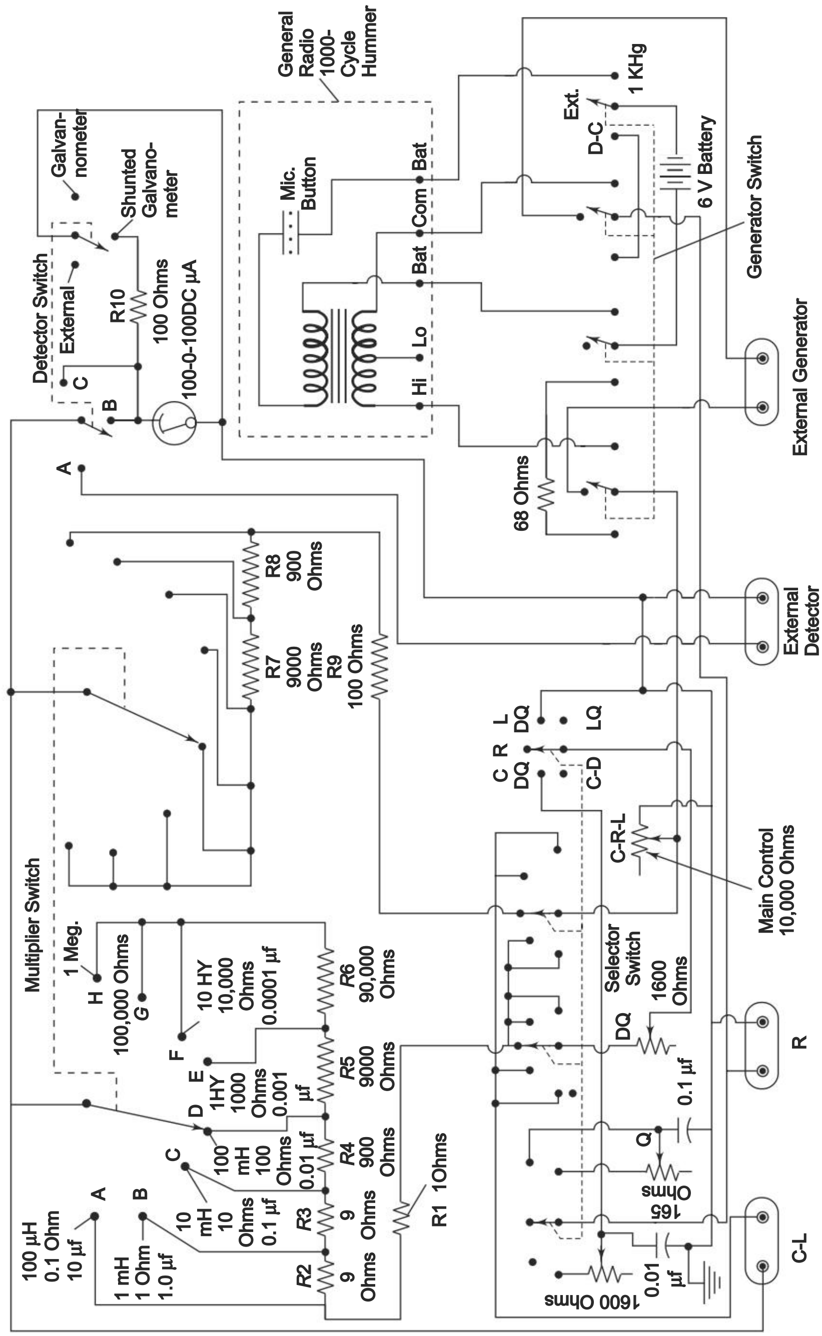


Fig. 10-15 Kit type impedence bridge

Rheostat Q has a logarithmic taper while rheostats D and DQ have a linear taper.

The reactive standards are self contained 0.01 and 0.1 μF precision capacitors. Both capacitance and inductance measurements are made against these capacitors as standards.

Initial Calibration The settings of the Main Control dial pointer must agree (when referred to the printed scale under this pointer) with the settings of the Main Control rheostat. It is the purpose of the initial calibration to align the scale and rheostat.

1. Set the Detector Switch to its shunted galvanometer position.
2. Set the Generator Switch to its D – D position.
3. Set the D, Q, and DQ rheostats to their zero or minimum resistance readings.
4. Set the Multiplier Switch to its 1000 Ω position (position D).
5. Set the Selector Switch to its R position.
6. Connect an accurately known 10,000 Ω resistor to the pair of terminals labelled R.
7. Adjust the Main Control rheostat for null, as indicated by zero reading of the internal microammeter.
8. Set the Detector Switch to its galvanometer position, and read just the main control for a sharper null indication.
9. The Main Control could read exactly 10 (which is an indication of 10,000 Ω. The value of the resistor connected to the R terminals). If it fails to do so, loosen the set screw of the Main Control knob without disturbing the setting of the rheostat, set the pointer exactly to 10, and retighten the set screw.

Since precision resistors are used in positions R_1 to R_9 and accurate capacitors are used for the 0.01 and 0.1 μF standards, this one-point calibration automatically calibrates all ranges of the bridge for ac and dc. If a highly precise, point by point calibration is required on the various ranges, this may be accomplished with a number of precision resistors, capacitors and inductors. Each component is connected to the bridge set up in the proper manner, a null adjustment is made, and the value of the component recorded against the setting of the main control. The direct readings of the D, Q and DQ rheostat dials are accurate only for a 1000 cycles per second bridge signal frequency.

RX METERS 10.9

An RX meter is used to measure the separate resistive and reactive components of a parallel Z network. It is especially useful in the design of RF tank circuits but can be used with any parallel "Z". The basic circuit of a typical RX meter is shown in Fig. 10.16.

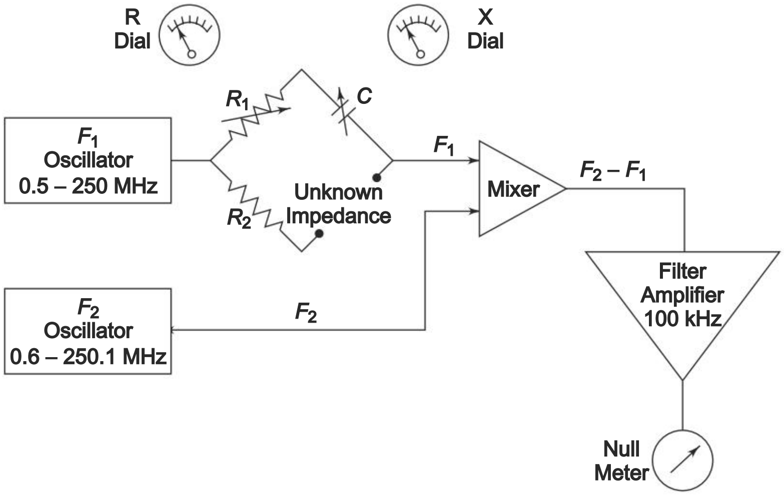


Fig. 10.16 Basic circuit of R-X meter

As shown in Fig. 10.16, there are two variable frequency oscillators that track each other at frequencies 100 kHz apart. The output of a 0.5 – 250 MHz oscillator, F_1 is fed into a bridge. When the impedance network to be measured is connected across one arm of the bridge, the equivalent parallel resistance and reactance (capacitive or inductive) unbalances the bridge and the resulting voltage is fed to the mixer. The output of the 0.6 – 250.1 MHz oscillator F_2, tracking 100 kHz above F_1, is also fed to the mixer.

This results in a 100 kHz difference frequency proportional in level to the bridge unbalance. The difference frequency signal is amplified by a filter amplifier combination and is applied to a null meter. When the bridge resistive and reactive controls are nulled, their respective dials accurately indicate the parallel impedance components of the network under test. For example, if balance is achieved with 50 Ω of resistance and 300 Ω of reactance, the network under test has the same values.

AUTOMATIC BRIDGES 10.10

The bridges discussed so far require that the controls be adjusted for balance after each capacitor (or other devices being tested) is connected to the bridge. In effect, they are manual. In recent years a number of automatic bridges have been developed. These bridges provide an automatic readout without adjustment of balance controls.

In some cases, the automatic bridges also provide a Binary Coded Digital (BCD) readout to external equipment. Automatic bridges are similar in operation to digital meters. To understand their operation, it is necessary to understand logic and digital circuit methods.

A typical automatic bridge circuit is shown in Fig. 10.17. The circuit is transformer-ratio-arm bridge for the automatic measurement of capacity. The circuit is in balance when the currents through the standard capacitor and the unknown capacitor are equal, so that the current in the phase detector is zero. The range is chosen automatically by relays that select decade taps on the ratio transformer. The phase detector determines whether the current passing through the unknown arm, of the bridge is higher or lower than that through the standard arm and produces an error signal that indicates whether more or less voltage is required on the standard capacitor to reach a balance.

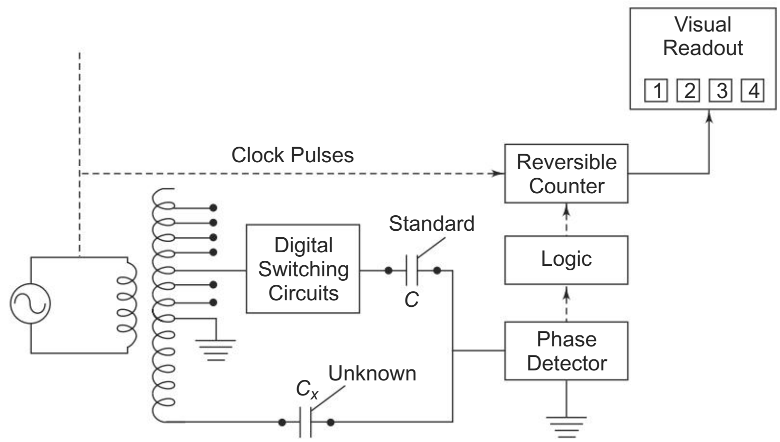


Fig. 10.17 Typical automatic bridge circuit

This information is used by a decade counter that controls the voltage on the standard capacitor through electronic switching circuits. The counter then counts in the direction which minimises the error signal, until a balance is reached. At balance, the value of the unknown is displayed on an in-line digital readout that indicates capacitance. This information is also presented in BCD form for use with printers and other data handling equipments.

TRANSISTOR TESTER 10.11

The term transistor tester (or analyser) used in this text is for instruments giving quantative measurements of transistor parameters. The tester should be able to provide direct readings for atleast two important measurements, such as:

1. A value for the forward gain in the common emitter configuration (h_{FE} for ac gain or h_{FE} for dc gain), that is β gain.
2. A value for collector to base leakage current, with emitter open I_{cbo}.

The latter measurement of C to B reverse current is generally regarded as most significant to test for the ageing of a transistor. It is an comparatively difficult measurement because of the small currents involved (in µA), and because of its extreme temperature sensitivity.

A typical service type transistor and diode tester examines transistors for the following characteristics.

1. Short circuits from C - E or base.
2. Direct measurement of I_{cbo}.
3. Direct reading of dc- β gain measurements (h_{FE})

10.11.1 Short-Circuit Test for C-E Breakdown

In the arrangement for the short-circuit test given in Fig. 10.18, the emitter and base terminals of the transistor under test are tied together and a reverse voltage of 4.5 V is applied between the collector and the two leads that are connected together. Figure 10.18(a) shows an arrangement for a PNP transistor and Fig. 10.18 (b) shows an arrangement for an NPN transistor.

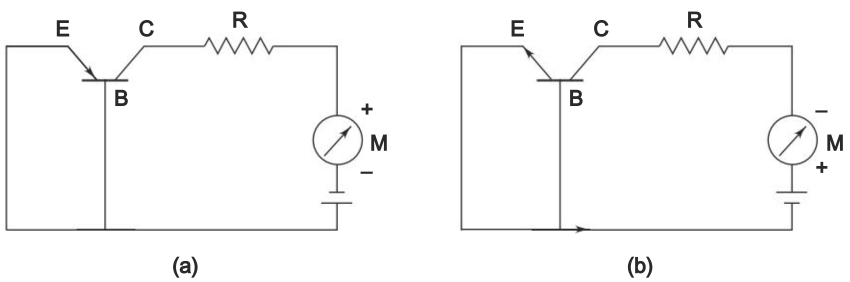


Fig. 10.18 Short circuit test

If there is a breakdown (usually from collector to emitter), the indicating meter tends to read full scale deflection. In that case, no further tests are performed on that transistor, thus avoiding possible damage to the meter circuit.

If the reading in the short position is less than the maximum allowable amount indicated on the chart, the next test for reverse current I_{cbo} is set up.

10.11.2 Direct Measurement of Collector-Leakage Current

Collector-leakage current is a function of the temperature and resistivity of the material of the transistor. Excessive leakage generally occurs when the conductor surface is contaminated. Other causes of this condition are overheating, or other types of damage. This is reverse current from the collector to the base, with the emitter open, denoted I_{cbo}. An excessive I_{cbo} indicates a faulty transistor. The tester circuit for this test is illustrated in Fig. 10.19(a). The collector-base junction is reverse biased and the emitter is open. The ammeter (in the micro range) indicates the reverse current.

Current from the collector to the emitter with an open base is denoted I_{ceo}. Figure 10.19(b) illustrates the circuit for this test. I_{ceo} should be expected to be much larger than I_{cbo}.

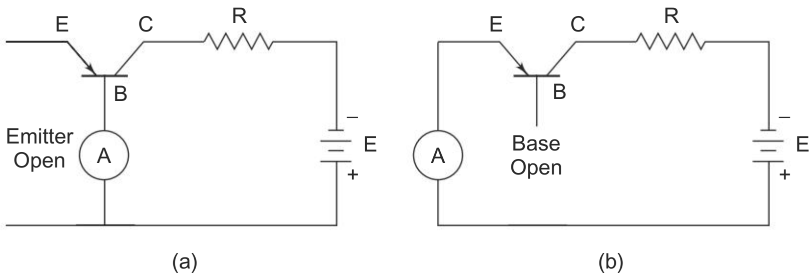


Fig. 10.19 (a) I_{cbo} test circuit (b) Tester circuit for I_{ceo} test

10.11.3 Direct Current Gain Test

Direct current gain is a measure of the effectiveness of the base in controlling collector current. It is a useful parameter for transistors used in low frequency power, amplification, switching, and control circuits. This test is useful only if the transistor is used in a common-emitter configuration.

Figure 10.20 illustrates the tester circuit used for test. R_1 is adjusted for a null on the voltmeter, at this point the dc gain is equal both to I_c/I_b, and R_x/R_2. R_1 usually has a calibrated dial to provide a direct reading of dc gain.

As a transistor ages, there is a tendency for the dc gain to decrease. This causes the amplification to decrease, which leads to distortion of the signal.

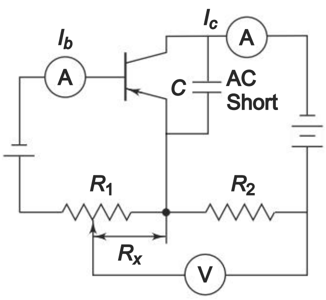


Fig. 10.20 Tester circuit for (gain) direct current test

10.11.4 Alternating Current Gain Test

Alternating current gain is expressed in two different ways, depending on the type of configuration the transistor is used in. For a common-base configuration, the amplification factor is alpha (α), which is the ratio of the change in the collector current to the change in the emitter current, with the collector voltage held constant. The collector has an ac short to the base during this test.

$$\alpha = \frac{\Delta I_c}{\Delta I_e}$$

If the transistor is in a common-emitter configuration, the amplification factor is beta β, which is the ratio of the change in collector current to the change in base current when the collector voltage is held constant. The collector has an ac short to the emitter during this test.

$$\beta = \frac{\Delta I_c}{\Delta I_b}$$

There is a definite relationship between alpha and beta, so that either may be calculated when the other is known.

$$\alpha = \frac{\beta}{1+\beta} \quad \text{and} \quad \beta = \frac{\alpha}{1-\alpha}$$

The test circuit for measuring beta is similar to that for measuring dc gain (see Fig. 10.20). The primary difference is that the beta measurement requires an ac signal at the transistor base. The proper values of either beta or alpha are given on the manufacturer's data sheet. The measured values should match these fairly closely, if the transistor is good.

10.11.5 Four Terminal Parameter Test (Hybrid Parameters)

A transistor may be considered as a four terminal network in order to determine the relationships between input and outputs. These relationships are referred to as hybrid (*h*) parameters, which are referred to in data sheets and on test instrument.

Hybrid parameters are very useful in determining the quality of a transistor. The four terminal network is shown in Fig. 10.21(a).

With this arrangement, there are two currents and two voltages to consider. If the two currents are considered as dependent variables, the resulting parameters are short-circuited parameters, and they are measured in mhos. When the two voltages are considered as dependent variables, the resulting parameters are open-circuit parameters, and they are measured in ohms. Hybrid (*h*) parameters are obtained by using one current and one voltage as dependent variables. The designations for the four *h* parameters are as follows.

h_i – input impedance with output shorted
h_r – reverse voltage ratio with input open
h_f – forward current gain with output shorted
h_o – output admittance with input open

The unit of measure for h_i is ohms, and for h_o is mhos. There are no units for h_f and h_r, since they are ratios.

The *h* parameters can be applied to any of the three basic amplifier configurations. An additional subscript letter is generally used to designate the type of configuration. Subscript *b* indicates common-base, *e* designates common-emitter and *c* denotes a common-collector.

The *h* parameter designations for a common emitter are h_{ie}, h_{re}, h_{fe}, and h_{oe}. Alpha for a common base circuit is equal to h_{fb} and beta in a common emitter is equal to h_{fe}.

The tester circuit for obtaining *h* parameters is illustrated in Fig. 10.21(b). G_1 is a calibrated current generator and G_2 is a calibrated voltage generator. The ac meter is used to indirectly measure current. The switches are four ganged sections of a five position rotary switch.

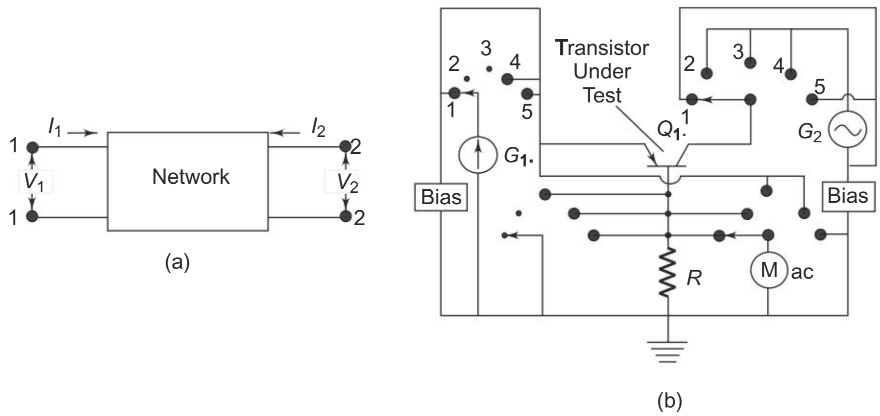


Fig. 10.21 (a) Basic four terminal network (b) Tester circuits for measuring hybrid parameters

Position 1 of the switch connects the calibrated current generator to the emitter of the transistor under test. The ac meter indicates $I + h_f$, which is the ratio of the ac base current (i_b) to the ac emitter current (i_e).

Position 2 of the switch connects the calibrated voltage generator to the collector, and the meter indicates the output admittance, h_o. The emitter is open during this measurement.

Position 3 measures h_r, position 4 measures h_i, and position 5 measures alpha.

10.11.6 Transistor Tester for Polarity

This tester checks the transistor for polarity (PNP or NPN). An audible signal gives an indication of gain. This tester is illustrated in Fig. 10.22.

The tester can also be used as a GO/NO GO tester to match unmarked devices.

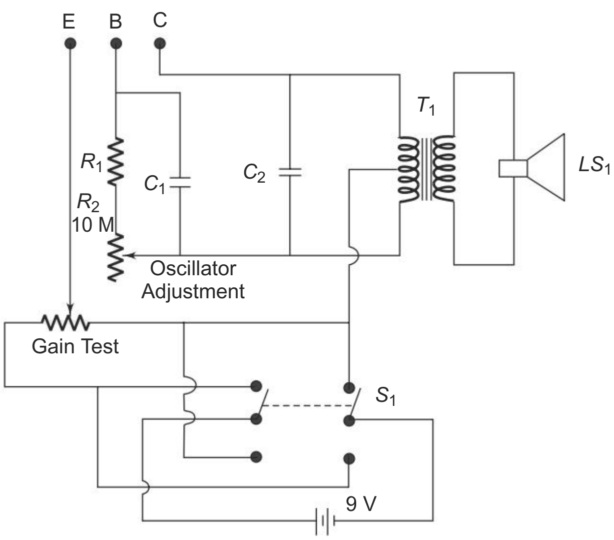


Fig. 10.22 Transistor tester for NPN/PNP

MEGGER 10.12

Another common method of measuring resistances above 50 M Ω is the Megger (megaohmmeter) shown in Fig. 10.23(a). This instrument is used to measure very high resistances, such as those found in cable insulations, between motor windings, in transformer windings, etc.

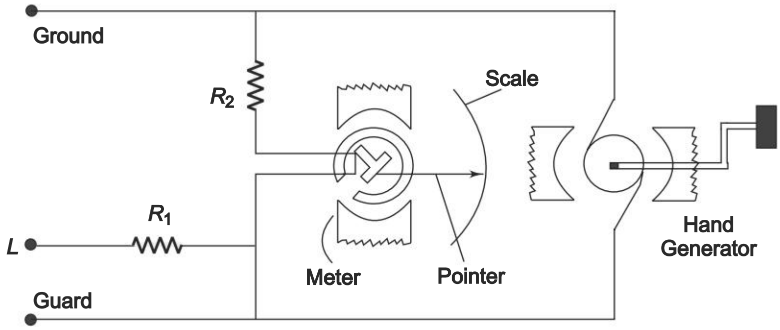


Fig. 10.23 (a) Basic megger circuit

Normal (shop) VOMs do not provide accurate indications above 10 MΩ because of the low voltage used in the ohmmeter circuit. Some laboratory test meters have a built-in ohmmeter with a high voltage power supply. The high voltage permits accurate high resistance measurement, but such meters are usually not portable. The Megger is essentially a portable ohmmeter with a built-in high voltage source. The Megger, shown in Fig. 10.23(b), has two main elements, a magnet-type dc generator to supply current for making measurements, and an ohmmeter which measures the resistance value. The generator armature is turned by a hand crank usually through step up gears, to produce an output voltage of 500 V. When the crank is turned, the gears turn the generator at high speed to generate an output voltage that may be 100, 500, 1000, 2500 or 5000 V, depending on the model.

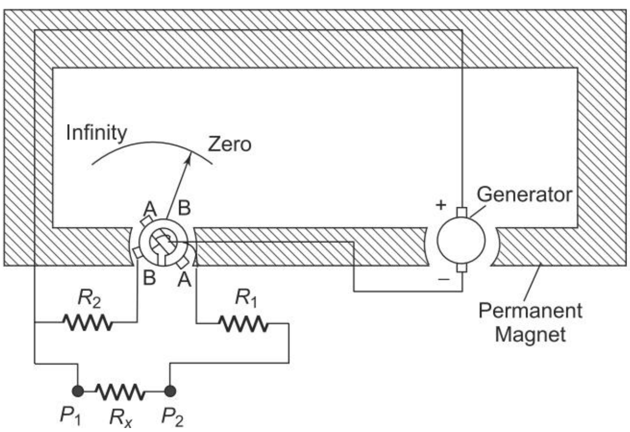


Fig. 10.23 (b) Megaohmmeter Circuit

The meter used differs slightly from the standard D'Arsonval movement, in that it has two windings. One windings is in series with the reference R_2 across the output of the generator, and is wound in such a way as to move the pointer towards the high resistance end of the scale when the generator is in operation. The other winding coil (A) and resistance R_1 are connected in series between the negative pole of the generator and the line terminal. This winding is so wound that when current flows through it from the generator, it tends to move the pointer towards the zero end of the scale. The two coils are mounted on the same shaft but at right angles to each other. Current is fed to both the coils by means of flexible connections that do not hinder the rotation of the element.

Coil A is the current coil with one terminal connected to the negative output and the other connected in series with R_1 to the test lead P_2. Test lead P_1 is connected to the generator positive output. When the unknown resistance R_x is connected across P_1 and P_2, current flows from the generator through coil A, resistances R_1 and R_x. The value of R_1 is so chosen so as to ensure that even if the line terminals are short-circuited, the current coil A is not damaged.

Coil B is the voltage coil and is connected across the generator output through the resistance R_2. If the test leads are left open, no current flows in coil A and coil B alone moves the pointer. Coil B take a position on the opposite side in the core, and the pointer indicates infinity or open.

When an extremely high resistance appears across the terminals, such as in an open circuit, the pointer reads infinity. On the other hand, when a resistance of relatively low value appears across the test points, such as when the cable insulation is wet, the current through the series winding causes the pointer to move towards zero (resistance short-circuited). However, the pointer stops at a point on the scale determined by the current through the series resistor, which in turn is governed by the value of the resistance being measured.

When an unknown resistance R_x is connected across the test leads, the current flows in coil A. The corresponding torque developed moves the pointer away from the infinity position, into a field of gradually increasing strength, until the torque fields between coils A and B are equal. Variations in the speed of the hand-cranked generator do not affect the Megaohmmeter readings, since charges in generator voltage affects both coils in the same manner.

Meggers may read resistances of several hundred or even thousands of megaohms. They have the advantage, as compared to an ordinary ohmmeter, of applying a high voltage to the circuit under test, and this voltage causes a current if any electrical leakage exists.

ANALOG pH METER 10.13

pH is defined as the negative logarithm of the active hydrogen ion. It is a measure of the acidity or alkalinity of an aqueous solution. The pH scale runs from 0 – 14, with pH 7, the neutral point, being the pH of pure water at 25°C. The lower the pH value, the more acidic the solution. Increasing pH values above 7.0, indicate increasing alkalinity.

Usually the pH is measured by immersing a special glass electrode and reference electrode into the solution. There are two types of methods used to measure the pH, the colorimetric method and the electrical method.

The colorimetric method is based on the assumption that if an indicator has the same colour in two solutions, then the pH of both solutions is the same. However, in practice this assumption does not hold good always, since the colour developed depends not only on the pH but also on other factors.

The electrical method is the most popular and is based upon a measurement of the electrode potential. The principle of this method is that when an electrode is immersed in the solution, a potential arises at the electrode solution boundary known as the electrode potential. This electrode potential, at a given temperature, depends upon the concentrations of ions of the electrolyte which exist in the solution. The electrode potential (in volts) of a metal immersed in a solution with ions of the same metal can be expressed by the following relation.

$$E = E_o + \frac{0.0001982\,(273 + t)}{n}\ \log_{10} a$$

where E_o = potential of the electrode, when its active-ion concentration in the solution is equal to unity.

t = temperature in degree centigrade

n = valency of the ion

a = active concentration of the ions of the metal in gram-equivalents per litre

In practice, only the potential difference can be measured and so the pH always has two elements; a measuring element, the potential of which depends on the concentration of the hydrogen ions, and a comparison element, the potential of which must remain constant. Two such elements connected electrically form a galvanic system, and by measuring the emf of this system we can drive the active concentration of the hydrogen ions in the solution under investigation.

The various electrodes used for pH measurements are as follows.

1. Hydrogen electrode
2. Calomel electrode
3. Quinhydrone and antimony electrodes
4. Glass electrodes

We now give some of the electrical methods used for measuring pH.

10.13.1 pH Measurement Using Hydrogen Electrode

The hydrogen electrode comprises a platinum plate covered with platinum black kept in a hydrogen element, where gaseous hydrogen at atmospheric pressure acts directly on this plate.

It consists of a hollow tube which is opened at the bottom end and immersed in the solution whose pH value is to be measured. The platinum plate is inserted into this hollow tube and joined to the platinum wire which is welded to the glass tube. This platinum wire is joined to a copper lead, as shown in Fig. 10.24.

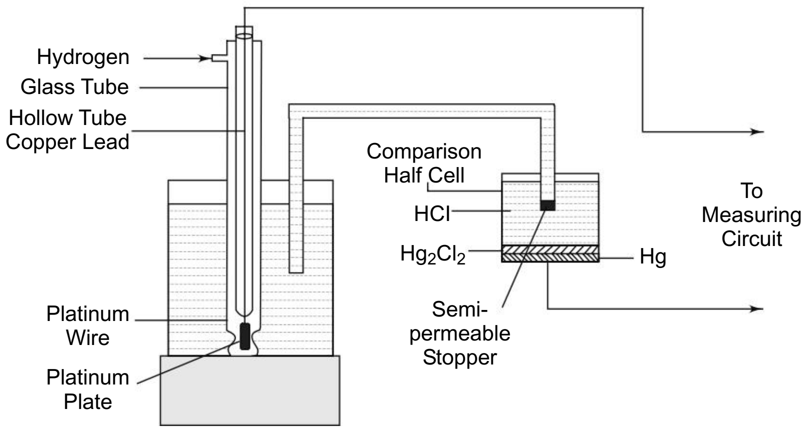


Fig. 10.24 pH measurement using hydrogen electrode

The hydrogen is fed to the outer tube from above and passes out through the lower apertures placed half way up the platinum plate. The hydrogen electrode is joined to the comparison half cell by an electrolytic solution. The cell is filled with an inert salt (for example, HCl) and closed by a semi-permeable stopper. This means of connecting the elements lowers the diffusion potential (up to 1– 2 mV), which arises at the boundary of the two solutions and introduces an error in the result of the measurement.

The electric potential developed on a hydrogen electrode dipped in a solution is caused by the concentration of hydrogen ions in the solution, or the pH of the solution. By measuring this electric potential, the pH of the solution can be determined. There is no method by which the absolute value of the potential developed on the hydrogen electrode can be determined. It is always the potential difference between two objects or the potential difference between two points on the same object which is determined. Therefore it is possible only to determine the relative potential of the hydrogen electrode compared to the known potential of some reference electrode.

10.13.2 pH Measurement Using a Thermocompensator

A thermocompensator is a temperature sensitive element (usually thermistor) installed in the solution with the electrodes. The glass electrode potential is influenced by changes in the temperature of the sample solution, and the thermocompensator corrects for this change. It should be noted that the thermocompensator corrects for the voltage/temperature relationship of the glass electrode and does not correct for the actual changes in the pH of the solution with temperature variations. The use of a matched pair of thermistors can provide a linear temperature compensation response of 90% in less than one minute.

10.13.3 Differential Input pH Amplifier

Differential input techniques use integrated circuit operational amplifiers in conjunction with a difference amplifier. This device is so small that it can be mounted on a single circuit board and sealed on top of the electrode station (provided the temperature to which it is subjected is less than 70°C; otherwise it is mounted in separate junction box suitably located at the site).

As the pH glass electrode is a high impedance device with output at extremely low current levels, a high impedance precision amplifier with high stability is required for the accurate and reliable measurement of this signal. The very high resistance of the glass electrode demands that the measuring techniques does not require current when the reading is made. If any current is taken from the circuit, the portion of the emf used to force the current through the resistance subtracted from that originally available, and all values obtained are fictitiously low.

In a pH measurement system, any extraneous leakage current to the ground find a return path through the normally low resistance reference electrode. With a conventional single input amplifier, this extraneous voltage developed across the reference electrode adds to the measured pH potential, resulting in an error in the indicated pH. If the resistance of the reference electrode is relatively low, this error is usually insignificant. However, if this resistance increases due to coating or junction clogging, the extraneous voltage across it increases and can cause a significant error in the indicated pH.

In the differential input pH amplifier, the glass pH measuring electrode and the reference electrode are each inputted to separate high impedance constant gain integrated circuit operational amplifiers A_1 and A_2, with negligible current flowing through the electrodes. This arrangement is shown in Fig. 10.25.

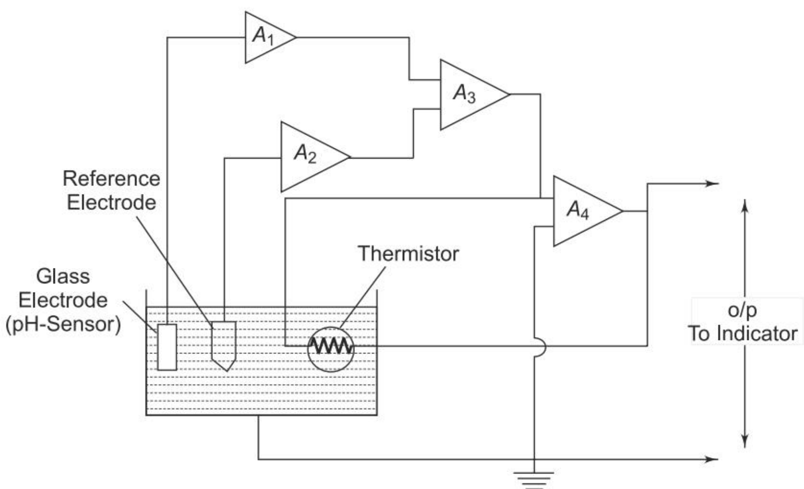


Fig. 10.25 Differential input pH meter

The glass electrode amplifier A_1 measures the pH potential as well as any potential through the solution to ground. The reference electrode amplifier A_2 measures the potential through the solution to ground. Amplifiers A_1 and A_2 are outputted to a difference amplifier A_3 which cancels out the solution potential and produces a signal equal only to the pH potential.

This method has the following advantages.

1. Virtually eliminates drift and noise
2. Has no stringent installation and maintenance requirements
3. Allows the pH indicator to be kept virtually any distance from the electrodes
4. Eliminates the need for costly coaxial cables between the measuring point and the pH indicator

10.13.4 Instruments for the Measurement of emf of Electrodes for pH Measurement

Indicating and automatic recording type instruments are used in industries for measuring the pH of different solutions. The pH is measured by measuring the difference of potential between the electrodes of the half cells. In order to eliminate distortions while measuring the emf of a sensor (e.g. the main sensing device of a pH meter, which consists of the appropriate electrode), the magnitude of the input resistance of the measuring instruments must be atleast two orders higher than the resistance of the electrodes. For example, if the resistance of glass electrodes is $1 - 2 \times 10^8\ \Omega$, the resistance of the instrument must be about $1 - 2 \times 10^{10}\ \Omega$. The dc voltage is transformed into ac voltage by frequency transformation circuits using either vibration converters or dynamic condensers.

TELEMETRY 10.14

Telemetry may be defined as measurement at a distance. According to ASA, *telemetering is the indicating, recording or integrating of a quantity at a distance by electrical means*.

There are several methods of classifying telemetering systems that are used.

IEEE bases its classification on the characteristics of the electrical signal. They are

(a) voltage (b) current (c) position (d) frequency, and (e) pulse.

Telemetering systems can also be classified as

(i) analog, and (ii) digital.

Telemetering can further be classified

(a) short-distance type, and (b) long-distance type.

The classification can also be based on whether the user has control over a transmission channel or not.

All of the IEEE's classifications can be used for shorter distance telemetring, but only the frequency and pulse types are suitable for long-distance telemetring.

The voltage, current, position and frequency can be used for analog telemetry, while only pulse type can be used for digital telemetring.

Voltage, current, position telemetring requires a physical connection between the transmitter and the receiver. The physical connection is called as a channel, which consists of one or two or more wires depending upon the system.

In case of RF, the telemetry channel is not a physical link

10.14.1 General Telemetring System

A general telemetring system is as shown in Fig 10.26. The primary detector and the final stage of the telemetring system have the same function as in a general measurement system.

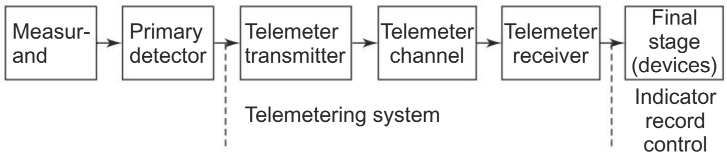


Fig 10.26 General telemetring system

The function of the telemeter transmitter is to convert output of a primary detector into a related quantity, which can be transmitted over the channel.

The function of the telemeter receiver, at the remote location, is to convert the transmitted signal into a related suitable quantity.

10.14.2 Electrical Telemetring System

The electrical telemetring system consists of a transmitter, which converts the input measured into an electrical signal, which is transmitted through a telemetring channel and is received at the other end by a receiver located at a distance in a remote location. This signal is then converted into a suitable usable form by the receiver and then can be indicated, recorded by the final stage device, which is calibrated in terms of the measurand.

The electrical telemetring systems can be broadly classified as

(i) dc systems, and (ii) ac systems.

10.14.3 DC Telemetring System

The dc telemetring system is categorized as

(i) Voltage Telemetring system
(ii) Current Telemetring system
(iii) Position Telemetring system

In a dc telemetring system the signal is transmitted through a channel, which utilizes direct transmission through cables in order to convey the information.

(i) Voltage Telemetring System

A voltage telemetring system transmits the measured variable as a function of an ac or dc voltage.

A simple voltage telemetring system is shown in Fig 10.27.

As seen from the diagram, a slide wire potentiometer is connected to a battery. The variable sliding contact of the potentiometer is moved by the pressure sensitive bourdon tube, which expands as the pressure is applied.

The telemetring channel consists of a pair of wires, which are connected to a voltage-measuring device such as a null balance dc indicator or recorder.

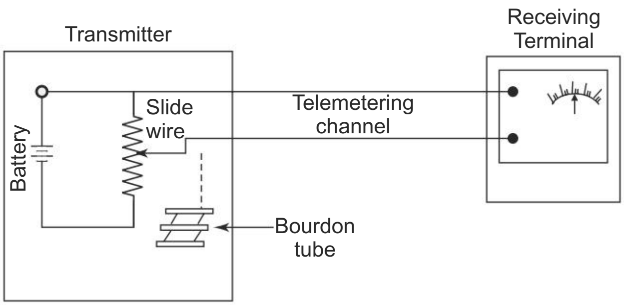


Fig. 10.27 Simple voltage telemetring system

As the measured pressure changes, the bourdon actuates the sliding contact, which in turn changes the voltage. The dc null balance potentiometer measures the voltage and positions the pointer on a scale calibrated to measure pressure.

The advantage of using a null balance dc potentiometer is, it reduces the current carried by the telemetring channel to a minimum.

The primary elements used by most of the systems produce a voltage signal. The primary elements can be thermocouples, tachometers, etc.

The application of voltage systems in the industry is limited to distances of up to about 300 metres.

Most of the receivers often use self-balancing potentiometers for such systems.

A voltage telemetring system requires high quality circuits than the current type. The signal-to-noise ratio must be comparatively high.

Since the power level is small for the voltage telemetring system, the transmission channel must be well protected from noise and interferences, which can be of the same order of the signal.

(ii) Current Telemetring System

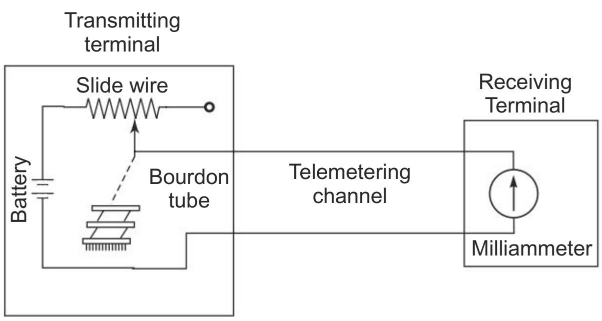


Fig 10.28 Basic current telemetering system

The basic version of a current telemetring system is shown in Fig 10.28.

This is similar to the voltage telemetring system except that the slide wire potentiometer is now connected in series with the battery.

The sliding contact of the potentiometer is connected to the pressure sensitive bourdon tube, which moves as the bourdon tube expands.

In this case, the telemetring channel consists of a pair of wires, which are connected to a current measuring device.

As the pressure changes, the bourdon tube expands, which in turn changes the position of the sliding contact on the (potentiometer) slide wire. This changes the current flow in the circuit. This current is measured by a milliammeter, whose scale has been calibrated to read pressure directly.

The most commonly used current telemetring systems are the (a) motion, and (b) force balance types which are improved version of the basic type.

(a) Motion Balance System In a motion balance system, the potentiometer or the slide wire is replaced by a position detector, such as LVDT as shown in Fig. 10.29.

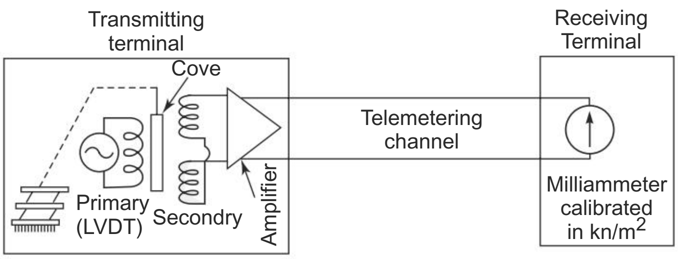


Fig 10.29 Motion current telemetering system

In this system, the bourdon tube is connected to the core of the LVDT. When pressure is applied to the bourdon tube, the bourdon tube expands causing a displacement. This moves the core of the LVDT, thereby producing an output voltage, which is then amplified and rectified. This voltage produces a dc current of the order of 4 mA to 20 mA in the telemetry channel and is measured by the milliammeter.

The scale of the milliammeter is directly calibrated in terms of pressure that is being measured.

(b) Force Balance Systems A force balance systems is as shown in Fig. 10.30, a part of the current output is fed back to oppose the motion (movement) of the input variable.

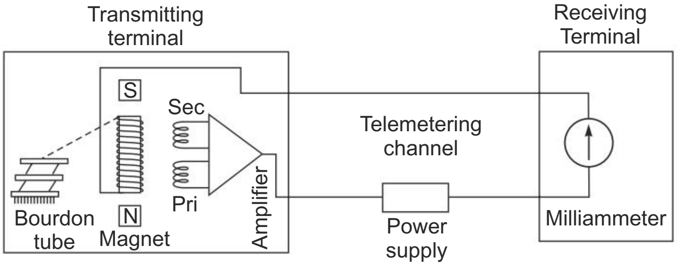


Fig 10.30 Force balance current telemetering system

The bourdon tube operates the system which rotates the feedback force coil.

This in turns changes the flux linkage between the primary and the secondary coils.

This change in flux linkages varies the amplitude of the amplifier. The output signal is then fed back to the feedback force coil, which in turn produces a force opposing the bourdon input.

A force balance system increases the accuracy, as smaller motions are required which results in better linearity.

Review Questions

1. State the working principle of an output power meter.
2. Explain with a diagram the working of an output power meter.
3. How is field strength measured? Explain the basic principle of a field strength meter.
4. Explain with a diagram the working of a basic field strength meter.
5. Explain with a diagram the working of a field strength meter using transistors.
6. State the basic principle on which the stroboscope operates.
7. Explain with a diagram the operation of a stroboscope.
8. Explain how speed of a motor can be measured using a stroboscope.
9. Explain the principle of a phase meter or a phase-sensitive detector.
10. Explain with a diagram the working of a phase-sensitive detector.
11. State the principle used in a basic vector impedance meter.
12. Explain with a diagram the working of a vector impedance meter.
13. Explain with a diagram how impedance can be measured using a commercial vector impedance meter. Explain with a diagram how phase angle can be measured using a commercial vector impedance meter.
14. Define Q-factor and resonance. Explain the working principle of a Q-meter.
15. Describe with a diagram the operation of a Q-meter. List the factors that causes error in a Q-meter.
16. Explain how Q-meter can be used to measure the following:
    (a) dc resistance of a coil
    (b) stray capacitance
    (c) impedance of a circuit
    (d) characteristics impedance of a transmission line.
17. Explain the operation of a Q-meter for measurement of stray capacitance.
18. Explain the operation of a Q-meter for measurement of HF resistance. Explain the operation of a Q-meter for measurement of low impedance value.
19. Explain the operation of a Q-meter for measurement of high impedance value
20. How can Q be measured using susceptance method?
21. What is a *LCR* bridge? How can L, C and R be measured using a skeleton *LCR* bridge?
22. State the features of a kit type *LCR* bridge.
23. Explain in brief the working of a kit type *LCR* bridge.
24. State the principle of R-X meters. Explain with a basic diagram the operation of an R-X meter.
25. Explain with a diagram the operation of an automatic bridge.

26. What is a transistor tester?
27. Explain with diagram how a transistor tester can be used for the measurement of the following:
    (i) Faulty transistor (ii) I_{cbo}
    (ii) I_{ceo} (iv) Beta gain
28. What is a megger? Explain with a diagram the working of a basic megger.
29. Explain with a diagram the working of a megaohmmeter circuit.
30. What do you understand by pH? Define pH.
31. How can pH be measured? State the different methods of pH measurement.
32. Explain with diagram pH measurement using hydrogen electrode.
33. Explain with diagram pH measurement using a thermocompensator.
34. Explain with diagram the operation of a differential input pH meter.
35. State the advantages of differential input pH meter.
36. Why is the necessity of using a thermocompensator for pH measurement?

Multiple Choice Questions

1. The output power meter is designed to measure the output power
   (a) inversely (b) directly
   (c) squared (d) indirectly
2. The field strength meter is used to measure
   (a) voltage
   (b) frequency
   (c) current
   (d) radiation intensity
3. The stroboscopic principle uses a
   (a) flashing light (b) inductance
   (c) capacitance (d) resistance
4. The stroboscope is used to measure
   (a) voltage (b) frequency
   (c) current (d) speed
5. The phase meter is measured by comparing
   (a) ac signal with reference signal
   (b) ac signal with dc signal
   (c) dc signal with reference signal
   (d) dc signal with ac signal
6. A vector impedance meter determines impedance in
   (a) magnitude form
   (b) polar form
   (c) exponential form
   (d) rectangular form
7. Q-factor is defined as
   (a) Reactance / Resistance
   (b) Resistance / Reactance
   (c) Resistance / Impedance
   (d) Impedance / Resistance
8. A Q meter is used to measure
   (a) voltage (b) inductance
   (c) capacitance (d) resistance
9. The shunt resistance R_{sh} used in Q meter has a value of
   (a) 0.02 Ω to 0.04 Ω
   (b) 2 Ω to 4 Ω
   (c) 0.2 Ω to 0.4 Ω
   (d) 2 kΩ to 4 kΩ.
10. The applied voltage to the resonant tank circuit has a
    (a) very low value 0.02 V
    (b) high value 4 V
    (c) medium value 0.04 V
    (d) very high value 400 V
11. A megger is used to measure
    (a) conductance (b) reactance
    (c) impedance (d) resistance
12. The pH scale runs from
    (a) 0–14 (b) 10–14
    (c) 7–14 (d) 0–7
13. The neutral point of pH is
    (a) 0 (b) 14
    (c) 7 (d) 10
14. A thermocompensator is a
    (a) thermocouple (b) RTD
    (c) thermometer (d) thermsistor

Practice Problems

1. Determine the distributed stray capacitances for the following data
First measurement f_1 = 4 MHz and C_1 = 3.3 kpf
Second measurement $f_2 = 3f_1$ = 12 MHz and C_2 = 1000 pf
Also calculate the value of inductance.
2. The distributed capacitance was found to be 20 pf by use of a Q meter. The first resonance occurred at C_1 = 300 pf and f_1 was half the second resonance frequency. Determine the value of C_2 and f_2 at the second resonance (given L = 40 μH).

Further Reading

1. *Handbook of Electronic Measurements*, Vols. I & II, Polytechnic Institute of Brooklyn, 1956. (Microwave Research Institute)
2. John D. Lenk, *Handbook of Electronic Meters, Theory and Applications*, Prentice-Hall, 1980.
3. Rugus, D. Turner, *Basic Electronics Test Instruments*, Rinehart Books, 1953.
4. Vestor Robinson, *Handbook of Electronic Instrumentation, Testing and Troubleshooting*, D.B. Taraporevala Sons & Co, 1979.
5. Miles, Retter Sander (jr), *Electronic Meters Techniques and Troubleshooting*, Reston Publishing Co, 1977.
6. John D. Lenk, *Handbook of Electronic Test Equipment, Prentice-Hall, 1971.*

Chapter

Bridges

11

INTRODUCTION 11.1

A bridge circuit in its simplest form consists of a network of four resistance arms forming a closed circuit, with a dc source of current applied to two opposite junctions and a current detector connected to the other two junctions, as shown in Fig. 11.1.

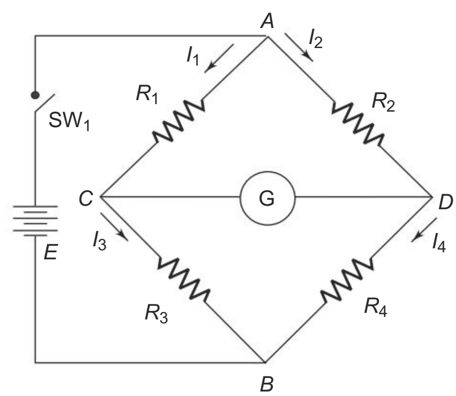


Fig. 11.1 Wheatstone's bridge

Bridge circuits are extensively used for measuring component values such as *R*, *L* and *C*. Since the bridge circuit merely compares the value of an unknown component with that of an accurately known component (a standard), its measurement accuracy can be very high. This is because the readout of this comparison is based on the null indication at bridge balance, and is essentially independent of the characteristics of the null detector. The measurement accuracy is therefore directly related to the accuracy of the bridge component and not to that of the null indicator used.

The basic dc bridge is used for accurate measurement of resistance and is called Wheatstone's bridge.

WHEATSTONE'S BRIDGE (MEASUREMENT OF RESISTANCE) 11.2

Wheatstone's bridge is the most accurate method available for measuring resistances and is popular for laboratory use. The circuit diagram of a typical Wheatstone bridge is given in Fig. 11.1. The source of emf and switch is connected to points *A* and *B*, while a sensitive current indicating meter, the galvanometer, is connected to points *C* and *D*. The galvanometer is a sensitive microammeter, with a zero center scale. When there is no current through the meter, the galvanometer pointer rests at 0, i.e. mid scale. Current in one direction causes the pointer to deflect on one side and current in the opposite direction to the other side.

When SW_1 is closed, current flows and divides into the two arms at point A, i.e. I_1 and I_2. The bridge is balanced when there is no current through the galvanometer, or when the potential difference at points C and D is equal, i.e. the potential across the galvanometer is zero.

To obtain the bridge balance equation, we have from the Fig. 11.1.

$$I_1 R_1 = I_2 R_2 \tag{11.1}$$

For the galvanometer current to be zero, the following conditions should be satisfied.

$$I_1 = I_3 = \frac{E}{R_1 + R_3} \tag{11.2}$$

$$I_2 = I_4 = \frac{E}{R_2 + R_4} \tag{11.3}$$

Substituting in Eq. (11.1)

$$\frac{E \times R_1}{R_1 + R_3} = \frac{E \times R_2}{R_2 + R_4}$$

$$R_1 \times (R_2 + R_4) = (R_1 + R_3) \times R_2$$

$$R_1 R_2 + R_1 R_4 = R_1 R_2 + R_3 R_2$$

$$R_4 = \frac{R_2 R_3}{R_1}$$

This is the equation for the bridge to be balanced.

In a practical Wheatstone's bridge, at least one of the resistance is made adjustable, to permit balancing. When the bridge is balanced, the unknown resistance (normally connected at R_4) may be determined from the setting of the adjustable resistor, which is called a standard resistor because it is a precision device having very small tolerance.

Hence $$R_x = \frac{R_2 R_3}{R_1} \tag{11.4}$$

Example 11.1 *Figure 11.1 consists of the following parameters. $R_1 = 10$ k, $R_2 = 15$ k and $R_3 = 40$ k. Find the unknown resistance R_x.*

Solution From the equation for bridge balance we have

$$R_1 R_4 = R_2 R_3, \text{ i.e. } R_1 R_x = R_2 R_3$$

Therefore $$R_x = \frac{R_2 R_3}{R_1} = \frac{15\text{ k} \times 40\text{ k}}{10\text{ k}} = 60\text{ k}\Omega$$

11.2.1 Sensitivity of a Wheatstone Bridge

When the bridge is in an unbalanced condition, current flows through the galvanometer, causing a deflection of its pointer. The amount of deflection is a function of the sensitivity of the galvanometer. Sensitivity can be thought of as deflection per unit current. A more sensitive galvanometer deflects by a greater amount for the same current. Deflection may be expressed in linear or angular units of measure, and sensitivity can be expressed in units of S = mm/µA or degree/µA or radians/µA.

Therefore it follows that the total deflection D is $D = S \times I$, where S is defined above and I is the current in microamperes.

11.2.2 Unbalanced Wheatstone's Bridge

To determine the amount of deflection that would result for a particular degree of unbalance, general circuit analysis can be applied, but we shall use Thevenin's theorem.

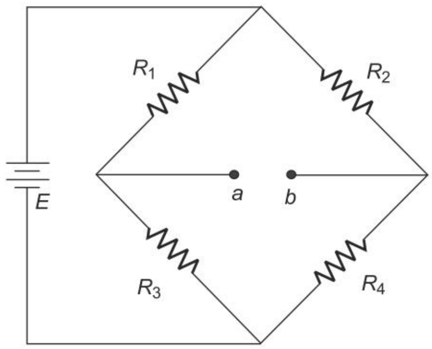


Fig. 11.2 Unbalanced wheatstone's bridge

Since we are interested in determining the current through the galvanometer, we wish to find the Thévenin's equivalent, as seen by the galvanometer.

Thévenin's equivalent voltage is found by disconnecting the galvanometer from the bridge circuit, as shown in Fig. 11.2, and determining the open-circuit voltage between terminals a and b.

Applying the voltage divider equation, the voltage at point a can be determined as follows

$$E_a = \frac{E \times R_3}{R_1 + R_3} \quad \text{and at point } b, \quad E_b = \frac{E \times R_4}{R_2 + R_4}$$

Therefore, the voltage between a and b is the difference between E_a and E_b, which represents Thévenin's equivalent voltage.

$$E_{th} = E_{ab} = E_a - E_b = \frac{E \times R_3}{R_1 + R_3} - \frac{E \times R_4}{R_2 + R_4}$$

Therefore $$E_{ab} = E\left(\frac{R_3}{R_1 + R_3} - \frac{R_4}{R_2 + R_4}\right)$$

Thévenin's equivalent resistance can be determined by replacing the voltage source E with its internal impedance or otherwise short-circuited and calculating the resistance looking into terminals a and b. Since the internal resistance is assumed to be very low, we treat it as 0 Ω. Thévenin's equivalent resistance circuit is shown in Fig. 11.3.

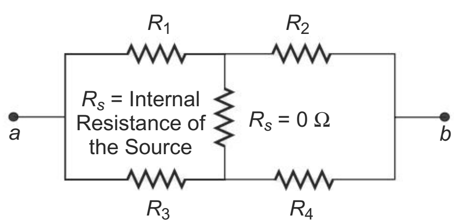


Fig. 11.3 Thévenin's resistance

The equivalent resistance of the circuit is $R_1 // R_3$ in series with $R_2 // R_4$ i.e. $R_1 // R_3 + R_2 // R_4$.

$$\therefore \qquad R_{th} = \frac{R_1 R_3}{R_1 + R_3} + \frac{R_2 R_4}{R_2 + R_4}$$

Therefore, Thévenin's equivalent circuit is given in Fig. 11.4. Thévenin's equivalent circuit for the bridge, as seen looking back at terminals a and b in Fig. 11.2, is shown in Fig. 11.4.

If a galvanometer is connected across the terminals a and b of Fig. 11.2, or its Thévenin equivalent Fig. 11.4 it will experience the same deflection at the output of the bridge. The magnitude of current is limited by both Thévenin's equivalent resistance and any resistance connected between a and b. The resistance between a and b consists only of the galvanometer resistance R_g. The deflection current in the galvanometer is therefore given by

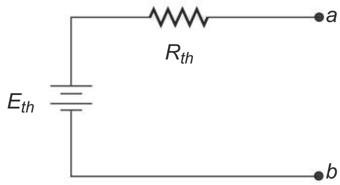


Fig. 11.4 Thévenin's equivalent

$$I_g = \frac{E_{th}}{R_{th} + R_g} \tag{11.5}$$

Example 11.2 *An unbalanced Wheatstone bridge is given in Fig. 11.5. Calculate the current through the galvanometer.*

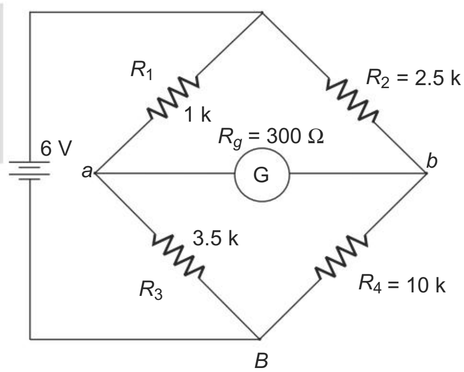


Fig. 11.5

Solution The Thévenin's equivalent voltage between a and b is the difference of voltages at these points i.e.

$$E_{th} = E_a - E_b = E_b - E_a.$$

$$\therefore \quad E_{th} = E\left(\frac{R_4}{R_2 + R_4} - \frac{R_3}{R_1 + R_3}\right)$$

$$E_{th} = 6\left(\frac{10\text{ k}}{2.5\text{ k} + 10\text{ k}} - \frac{3.5\text{ k}}{1\text{ k} + 3.5\text{ k}}\right)$$

$$E_{th} = 6\ (0.800 - 0.778)$$

$$E_{th} = 0.132\text{ V}$$

Thévenin's equivalent resistance is

$$R_{th} = \frac{R_1 R_3}{R_1 + R_3} + \frac{R_2 R_4}{R_2 + R_4}$$

$$R_{th} = \frac{1\text{ k} \times 3.5\text{ k}}{1\text{ k} + 3.5\text{ k}} + \frac{2.5\text{ k} \times 10\text{ k}}{2.5\text{ k} + 10\text{ k}}$$

$$= 0.778 \text{ k} + 2 \text{ k}$$
$$= 2.778 \text{ k}$$

The equivalent circuit connected along with the galvanometer is as shown in Fig. 11.6.

The current through the galvanometer is given by

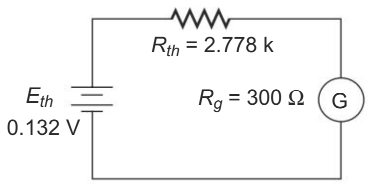


Fig. 11.6 Equivalent circuit

$$I_g = \frac{E_{th}}{R_{th} + R_g} = \frac{0.132 \text{ V}}{2.778 \text{ k} + 0.3 \text{ k}} = 42.88 \ \mu\text{A}$$

11.2.3 Slightly Unbalanced Wheatstone's Bridge

If three of the four resistor in a bridge are equal to R and the fourth differs by 5% or less, we can develop an approximate but accurate expression for Thévenin's equivalent voltage and resistance.

Consider the circuit in Fig. 11.7. The voltage at point a is

$$E_a = \frac{E \times R}{R + R} = \frac{E \times R}{2R} = \frac{E}{2}$$

The voltage at point b is

$$E_b = \frac{R + \Delta r \times E}{R + R + \Delta r} = \frac{E(R + \Delta r)}{2R + \Delta r}$$

Thévenin's equivalent voltage between a and b is the difference between these voltages.

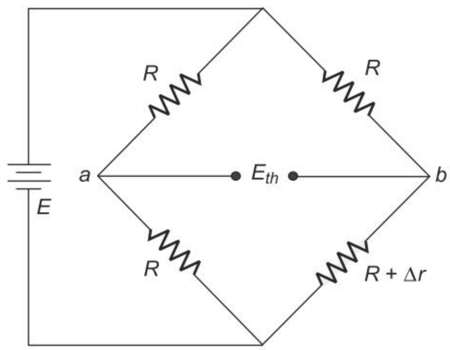


Fig. 11.7 Slightly unbalanced Wheatstone's bridge

Therefore $\quad E_{th} = E_a - E_b = E\left(\frac{(R + \Delta r)}{2R + \Delta r} - \frac{1}{2}\right)$

$$= E\left(\frac{2(R + \Delta r) - (2R + \Delta r)}{2(2R + \Delta r)}\right)$$

$$= E\left(\frac{2R + 2\Delta r - 2R - \Delta r}{4R + 2\Delta r}\right)$$

$$= E\left(\frac{\Delta r}{4R + 2\Delta r}\right)$$

If Δr is 5% of R or less, Δr in the denominator can be neglected without introducing appreciable error. Therefore, Thévenin's voltage is

$$E_{th} = \frac{E \times \Delta r}{4R} = E\left(\frac{\Delta r}{4R}\right)$$

The equivalent resistance can be calculated by replacing the voltage source with its internal impedance (for all practical purpose short-circuit). The Thévenin's equivalent resistance is given by

$$R_{th} = \frac{R \times R}{R + R} + \frac{R(R + \Delta r)}{R + R + \Delta r}$$

$$= \frac{R}{2} + \frac{R(R + \Delta r)}{2R + \Delta r}$$

Again, if Δr is small compared to R, Δr can be neglected. Therefore,

$$R_{th} = \frac{R}{2} + \frac{R}{2} = R$$

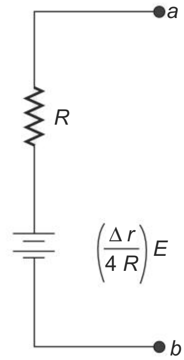


Fig. 11.8 Thévenin's equivalent of a slightly unbalanced Wheatstone's bridge

Using these approximations, the Thévenin's equivalent circuit is as shown in Fig. 11.8. These approximate equations are about 98% accurate if $\Delta r \leq 0.05\,R$.

Example 11.3 *Given a centre zero 200 – 0 – 200 µA movement having an internal resistance of 125 Ω. Calculate the current through the galvanometer given in Fig. 11.9 by the approximation method.*

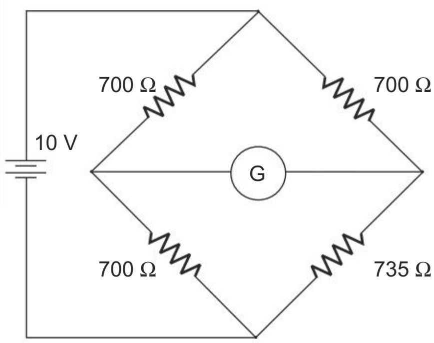


Fig. 11.9

Solution The Thévenin's equivalent voltage is

$$E_{th} = \frac{E(\Delta r)}{4R}$$

$$= \frac{10 \times 35}{4 \times 700} = 0.125 \text{ V}$$

Thévenin's equivalent resistance is

$$R_{th} = R = 700\ \Omega$$

The current through the galvanometer is

$$I_g = \frac{E_{th}}{R_{th} + R_g} = \frac{0.125 \text{ V}}{700 + 125} = \frac{0.125}{825} = 151.5\ \mu\text{A}$$

If the detector is a 200 – 0 – 200 µA galvanometer, we see that the pointer is full scale for a 5% change in resistance.

11.2.4 Application of Wheatstone's Bridge

A Wheatstone bridge may be used to measure the dc resistance of various types of wire, either for the purpose of quality control of the wire itself, or of some assembly in which it is used. For example, the resistance of motor windings, transformers, solenoids, and relay coils can be measured.

Wheatstone's bridge is also used extensively by telephone companies and others to locate cable faults. The fault may be two lines shorted together, or a single line shorted to ground.

11.2.5 Limitations of Wheatstone's Bridge

For low resistance measurement, the resistance of the leads and contacts becomes significant and introduces an error. This can be eliminated by Kelvin's Double bridge.

For high resistance measurements, the resistance presented by the bridge becomes so large that the galvanometer is insensitive to imbalance. Therefore, a power supply has to replace the battery and a dc VTVM replaces the galvanometer. In the case of high resistance measurements in mega ohms, the Wheatstones bridge cannot be used.

Another difficulty in Wheatstone's bridge is the change in resistance of the bridge arms due to the heating effect of current through the resistance. The rise in temperature causes a change in the value of the resistance, and excessive current may cause a permanent change in value.

KELVIN'S BRIDGE 11.3

When the resistance to be measured is of the order of magnitude of bridge contact and lead resistance, a modified form of Wheatstone's bridge, the Kelvin bridge is employed.

Kelvin's bridge is a modification of Wheatstone's bridge and is used to measure values of resistance below 1 Ω. In low resistance measurement, the resistance of the leads connecting the unknown resistance to the terminal of the bridge circuit may affect the measurement.

Consider the circuit in Fig. 11.10, where R_y represents the resistance of the connecting leads from R_3 to R_x (unknown resistance). The galvanometer can be connected either to point c or to point a. When it is connected to point a, the resistance R_y, of the connecting lead is added to the unknown resistance R_x, resulting in too high indication for R_x. When the connection is made to point c, R_y is added to the bridge arm R_3 and resulting measurement of R_x is lower than the actual value, because now the actual value of R_3 is higher than its nominal value by the resistance R_y. If the galvanometer is connected to point b, in between points c and a, in such a way that the ratio of the resistance from c to b and that from a to b equals the ratio of resistances R_1 and R_2, then

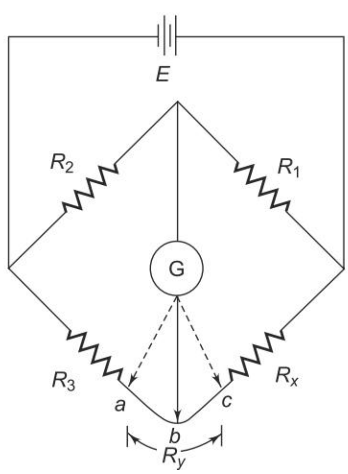


Fig. 11.10 Kelvin's bridge

$$\frac{R_{cb}}{R_{ab}} = \frac{R_1}{R_2} \tag{11.6}$$

and the usual balance equations for the bridge give the relationship

$$(R_x + R_{cb}) = \frac{R_1}{R_2}\,(R_3 + R_{ab}) \tag{11.7}$$

but $\qquad R_{ab} + R_{cb} = R_y$ and $\dfrac{R_{cb}}{R_{ab}} = \dfrac{R_1}{R_2}$

$$\frac{R_{cb}}{R_{ab}} + 1 = \frac{R_1}{R_2} + 1$$

$$\frac{R_{cb} + R_{ab}}{R_{ab}} = \frac{R_1 + R_2}{R_2}$$

i.e. $$\frac{R_y}{R_{ab}} = \frac{R_1 + R_2}{R_2}$$

Therefore $\qquad R_{ab} = \dfrac{R_2\,R_y}{R_1 + R_2}$ and as $R_{ab} + R_{cb} = R_y$

$$\therefore \qquad R_{cb} = R_y - R_{ab} = R_y - \frac{R_2\,R_y}{R_1 + R_2}$$

$$\therefore \qquad R_{cb} = \frac{R_1\,R_y + R_2\,R_y - R_2\,R_y}{R_1 + R_2} = \frac{R_1\,R_y}{R_1 + R_2}$$

Substituting for R_{ab} and R_{cb} in Eq. (11.7), we have

$$R_x + \frac{R_1\,R_y}{R_1 + R_2} = \frac{R_1}{R_2}\left(R_3 + \frac{R_2\,R_y}{R_1 + R_2}\right)$$

$$R_x + \frac{R_1\,R_y}{R_1 + R_2} = \frac{R_1\,R_3}{R_2} + \frac{R_1\,R_2\,R_y}{R_2\,(R_1 + R_2)}$$

Hence $$R_x = \frac{R_1\,R_3}{R_2} \tag{11.8}$$

Equation (11.8) is the usual Wheatstone's balance equation and it indicates that the effect of the resistance of the connecting leads from point a to point c has been eliminated by connecting the galvanometer to an intermediate position, b.

The above principle forms the basis of the construction of Kelvin's Double Bridge, popularly known as Kelvin's Bridge. It is a Double bridge because it incorporates a second set of ratio arms. Figure 11.11 shows a schematic diagram of Kelvin's double bridge.

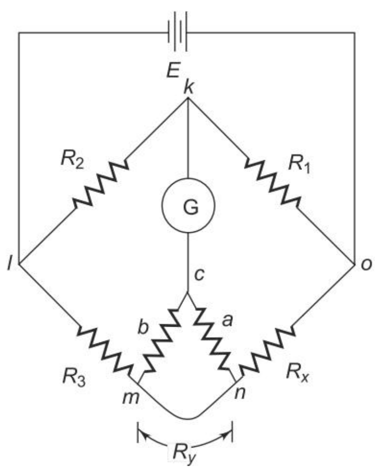


Fig. 11.11 Kelvin's double bridge

The second set of arms, a and b, connects the galvanometer to a point c at the appropriate potential between m and n connection, i.e. R_y. The ratio of the resistances of arms a and b is the same as the ratio of R_1 and R_2. The galvanometer indication is zero when the potentials at k and c are equal.

$$\therefore \quad E_{lk} = E_{lmc}$$

But
$$E_{lk} = \frac{R_2}{R_1 + R_2} \times E \tag{11.9}$$

and
$$E = I\left(R_3 + R_x + \frac{(a+b)R_y}{a+b+R_y}\right)$$

Substituting for E in Eq.(11.9),

we get
$$E_{lk} = \frac{R_2}{R_1 + R_2} \times I\left(R_3 + R_x + \frac{(a+b)R_y}{a+b+R_y}\right) \tag{11.10}$$

Similarly,
$$E_{lmc} = I\left(R_3 + \frac{b}{a+b}\left[\frac{(a+b)R_y}{a+b+R_y}\right]\right) \tag{11.11}$$

But
$$E_{lk} = E_{lmc}$$

i.e.
$$\frac{IR_2}{R_1 + R_2}\left(R_3 + R_x + \frac{(a+b)R_y}{a+b+R_y}\right) = I\left[R_3 + \frac{b}{a+b}\left\{\frac{(a+b)R_y}{a+b+R_y}\right\}\right]$$

$$\therefore \quad R_3 + R_x + \frac{(a+b)R_y}{a+b+R_y} = \frac{R_1 + R_2}{R_2}\left(R_3 + \frac{bR_y}{a+b+R_y}\right)$$

$$\therefore \quad R_3 + R_x + \frac{(a+b)R_y}{a+b+R_y} = \left(\frac{R_1}{R_2} + 1\right)\left(R_3 + \frac{bR_y}{a+b+R_y}\right)$$

$$R_x + \frac{(a+b)R_y}{a+b+R_y} + R_3 = \frac{R_1 R_3}{R_2} + R_3 + \frac{bR_1 R_y}{R_2\,(a+b+R_y)} + \frac{bR_y}{a+b+R_y}$$

$$R_x = \frac{R_1 R_3}{R_2} + \frac{bR_1 R_y}{R_2\,(a+b+R_y)} + \frac{bR_y}{a+b+R_y} - \frac{(a+b)R_y}{a+b+R_y}$$

$$R_x = \frac{R_1 R_3}{R_2} + \frac{bR_1 R_y}{R_2\,(a+b+R_y)} + \frac{bR_y - aR_y - bR_y}{a+b+R_y}$$

$$R_x = \frac{R_1 R_3}{R_2} + \frac{bR_1 R_y}{R_2\,(a+b+R_y)} - \frac{aR_y}{a+b+R_y}$$

$$R_x = \frac{R_1 R_3}{R_2} + \frac{bR_y}{(a+b+R_y)}\left(\frac{R_1}{R_2} - \frac{a}{b}\right)$$

But
$$\frac{R_1}{R_2} = \frac{a}{b}$$

Therefore, $$R_x = \frac{R_1 R_3}{R_2}$$

This is the usual equation for Kelvin's bridge. It indicates that the resistance of the connecting lead R_y, has no effect on the measurement, provided that the ratios of the resistances of the two sets of ratio arms are equal. In a typical Kelvin's bridge the range of a resistance covered is 1 – 0.00001 Ω (10 μohm) with an accuracy of ± 0.05% to ± 0.2%.

Example 11.4 *If in Fig. 11.12 the ratio of R_a to R_b is 1000 Ω, R_1 is 5 Ω and R_1 = 0.5 R_2. What is the value of R_x.*

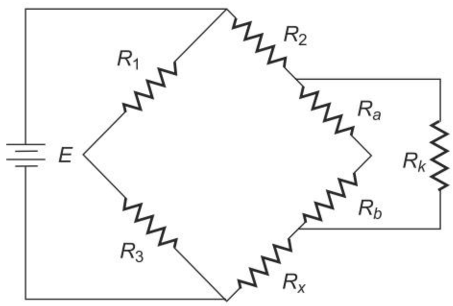


Fig. 11.12 Kelvin's bridge

Solution Resistance R_x can be calculated as follows.

$$\frac{R_x}{R_2} = \frac{R_b}{R_a}$$

Therefore,

$$\frac{R_x}{R_2} = \frac{R_b}{R_a} = \frac{1}{1000}$$

Since

$$R_1 = 0.5\ R_2,\ R_2 = 5/0.5 = 10\ \Omega.$$

Therefore $$R_x/10 = 1/1000 = 10 \times 1/1000 = 1/100 = 0.01\ \Omega.$$

PRACTICAL KELVIN'S DOUBLE BRIDGE 11.4

Figure 11.13 shows a commercial Kelvin's bridge capable of measuring resistances from 10 – 0.00001 Ω.

Contact potential drops in the circuit may cause large errors. This effect is reduced by varying a standard resistance consisting of nine steps of 0.001 Ω each, plus a calibrated manganin bar of 0.0011 Ω with a sliding contact. When both contacts are switched to select the suitable value of standard resistance, the voltage drop between the ratio arm connection points is changed, but the total resistance around the battery circuit is unchanged.

This arrangement places any contact resistance in series with the relatively high resistance value of the ratio arms, rendering the contact resistance effect negligible. The ratio R_1/R_2 is selected (as given in Fig. 11.13) such that a relatively large part of the standard resistance is used and hence R_x is determined to the largest possible number of significant figures. Therefore, measurement accuracy improves.

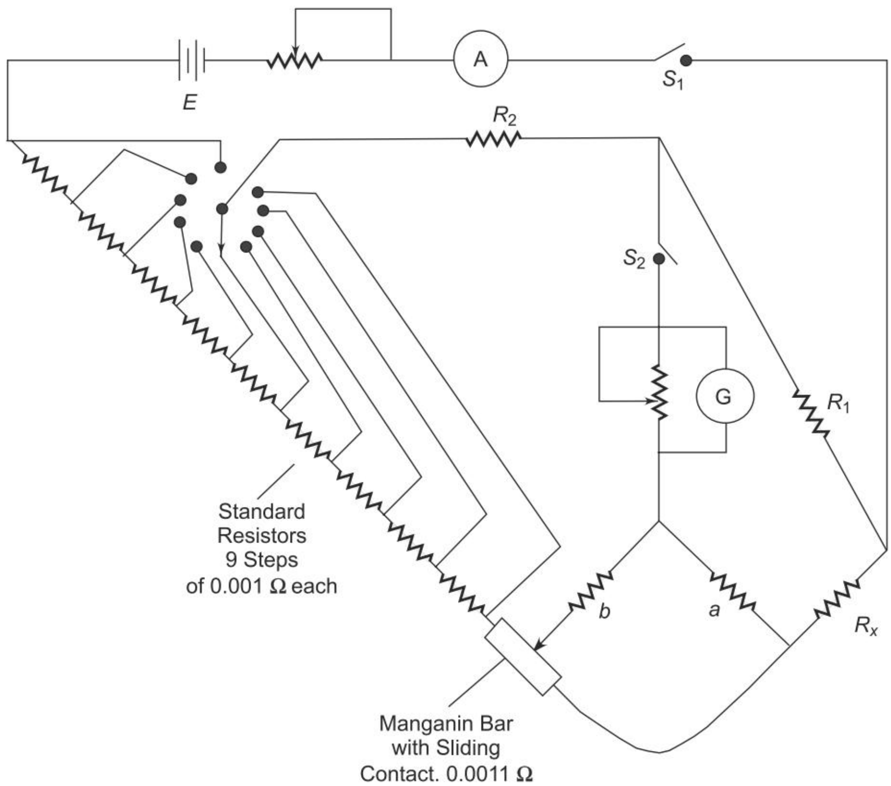


Fig. 11.13 Practical kelvin's bridge

BRIDGE CONTROLLED CIRCUITS 11.5

Whenever a bridge is unbalanced, a potential difference exists at its output terminal. The potential difference causes current to flow through the detector (say, a galvanometer) when the bridge is used as part of a measuring instrument. When the bridge is used as an error detector in a control circuit, the potential difference at the output of the bridge is called an error signal, as in Fig. 11.14.

Passive circuit elements such as strain gauges, temperature sensitive resistors (thermistors) and photo resistors, produce no output voltage. However, when used as one arm of Wheatstones bridge, a change in their sensitive parameter (heat, light, pressure) produces a change in their resistances. This causes the bridge to be unbalanced, thereby producing an output voltage or an error signal.

Resistor R_v in Fig. 11.14 may be sensitive to one of many different physical parameters, such as heat or light. If the particular parameter to which the resistor is sensitive, is of a magnitude such that the ratio R_2/R_v equals R_1/R_3, then the error signal is zero. If the physical parameters changes, R_v also changes. The bridge then becomes unbalanced and an error signal occurs. In most control applications the measured and controlled parameter is corrected, restoring R_v to the value that creates a null condition at the output of the bridge.

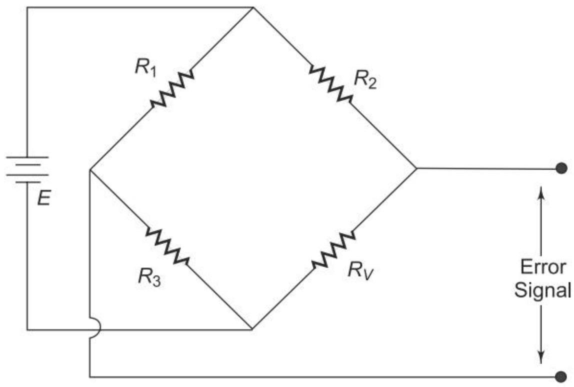


Fig. 11.14 Wheatstone's bridge error detector with resistance R_v sensitive to some physical parameters

Since R_v varies by only a small amount, the amplitude of the error signal is normally quite low. It is therefore amplified before being used for control purposes.

Example 11.5 *Resistor R_v in Fig. 11.15(a) is temperature sensitive, with a relation between resistance and temperature as shown in Fig. 11.15(b). Calculate (i) at what temperature the bridge is balanced, and (ii) The amplitude of the error signal at 60°C.*

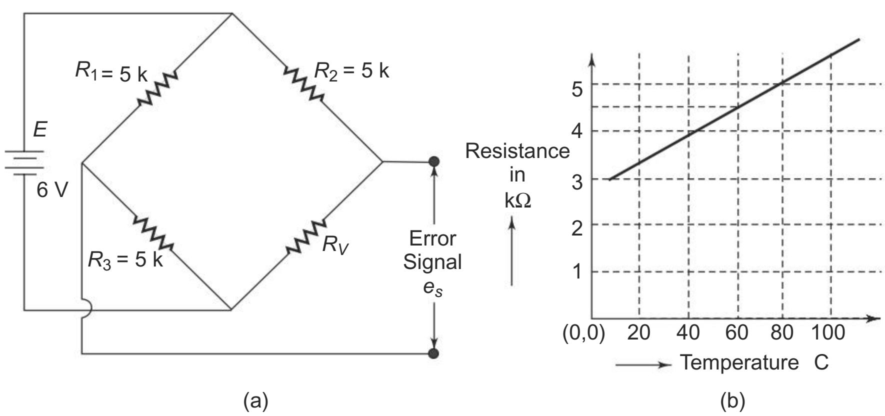


Fig. 11.15

Solution

(i) The value of R_v when the bridge is balance is calculated as

$$R_v = \frac{R_2\, R_3}{R_1} = \frac{5\text{ k} \times 5\text{ k}}{5\text{ k}} = 5\text{ k}\Omega$$

The bridge is balanced when the temperature is 80°C. This is read directly from the graph of Fig. 11.15(b).

(ii) We can also determine the resistance of R_v at 60°C directly from the graph. This values of 4.5 kΩ. Therefore the error signal is given by

$$e_s = E\left(\frac{R_3}{R_1+R_3} - \frac{R_v}{R_2+R_v}\right)$$

$$= 6\left(\frac{5\text{ k}}{5\text{ k}+5\text{ k}} - \frac{4.5\text{ k}}{5\text{ k}+4.5\text{ k}}\right)$$

$$= 6\,(0.5 - 0.4736)$$

$$= 6\,(0.0263)$$

$$= 0.158\text{ V}$$

The error signal can also be determined by using the following equation.

$$e_s = E_{th} = E\left(\frac{\Delta r}{4R}\right) = 6\left(\frac{500}{4\times 5\text{ k}}\right)$$

$$e_s = 0.150\text{ V}$$

DIGITAL READOUT BRIDGES 11.6

The tremendous increase in the use of digital circuitry has had a marked effect on electronic test instruments. The early use of digital circuits in bridges was to provide a digital readout. The actual measuring circuitry of the bridge remained the same, but operator error in observing the reading was eliminated. The block diagram for a Wheatstone bridge with digital readout is shown in Fig. 11.16. Note that a logic circuit is used to provide a signal to R_3, sense the null, and provide a digital readout representing the value of R_x.

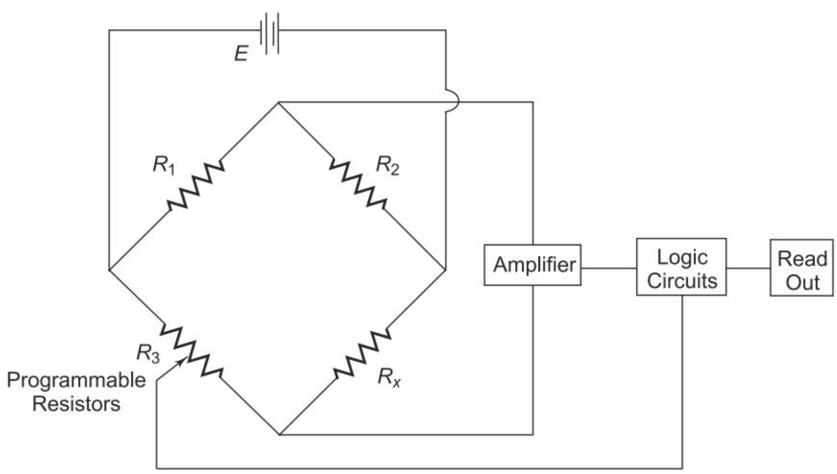


Fig. 11.16 Block diagram of wheatstone's bridge with digital readout

MICROPROCESSOR CONTROLLED BRIDGES 11.7

Digital computers have been used in conjunction with test systems, bridges, and process controllers for several years. In these applications, computers were used to give instructions and perform operations on the data measured. When microprocessors were first developed they were used in much the same way as digital computers. However, real improvements in performance occurred when the microprocessor was truly integrated into the instrument. With this accomplished, microprocessors cannot only give instructions about measurement, but also they can change the way the measurements are taken. This innovation has given rise to a whole new class of instruments, called Intelligent Instruments.

The complexity and cost of making analog measurements can be reduced using a microprocessor. This reduction of analog circuitry is important, even if additional digital circuitry must be added, because precision analog components are expensive. Also, adjusting, testing and troubleshooting analog circuits is time consuming and often expensive. Digital circuits can often replace analog circuits because various functions can be done either way.

The following are some of the ways in which microprocessors are reducing the cost and complexity of analog measurements.

1. Replacing sequential control logic with stored control programs.
2. Eliminating some auxiliary equipment by handling interfacing, programming and other system functions.
3. Providing greater flexibility in the selection of measurement circuits, thereby making it possible to measure one parameter and calculate another parameter of interest.
4. Reducing accuracy requirements by storing and applying correction factors.

Instruments in which microprocessors are an integral part can take the results of a measurement that is easiest to make in a given circuit, then calculate and display the value of some other desired parameter, which may be much more difficult to measure directly.

For example, conventional counters can measure the period of a low frequency waveform. This is then converted to frequency either manually, or using extensive circuitry. On the other hand, such calculations are done very easily by a microprocessor. Measurements of resistance and conductance, which are reciprocals of each other offer another example. Some hybrid digital/analog bridges are designed to measure conductance by measuring current. This measurement is then converted to a resistance value by rather elaborate circuitry. With a microprocessor based instrument, a resistance value is easily obtained from the conductance measurement.

Many other similar examples could be presented. However, the important thing to remember is that the microprocessor is an integral part of the measuring instrument. This results in an intelligent instrument that allows us to choose the easiest method of measurement and requires only one measurement circuit to obtain various results. Specifically, one quantity can be measured in terms of another, or several others with completely different dimensions, and the desired results calculated with the microprocessor.

(One such microprocessor-based instrument is the General Radio model 1658RLC digibridge.)

Such intelligent instruments represent a new era in impedance measuring instruments. The following are some features of these instruments.

1. Automatically measures R, inductance L, capacitance C, dissipation factor D and storage factors for inductors Q.
2. 0.1% basic accuracy
3. Series or parallel measurement mode
4. Autoranging
5. No calibration required
6. Ten bins for component sorting/binning (equivalent, binary number)
7. Three test speeds
8. Three types of display-programmed bin limits, measured values or bin number.

Most of these features are available because of the use of a microprocessor, e.g. the component sorting/binning feature is achieved by programming the microprocessor.

When using the instrument in this mode, bins are assigned a tolerance range. When a component is measured, a digital readout (bin number) indicating the proper bin for that component is displayed on the keyboard control panel.

AC BRIDGES 11.8

Impedances at AF or RF are commonly determined by means of an ac Wheatstone bridge. The diagram of an ac bridge is given in Fig. 11.17. This bridge is similar to a dc bridge, except that the bridge arms are impedances. The bridge is excited by an ac source rather than dc and the galvanometer is replaced by a detector, such as a pair of headphones, for detecting ac. When the bridge is balanced,

$$\frac{Z_1}{Z_3} = \frac{Z_2}{Z_4}$$

where Z_1, Z_2, Z_3 and Z_4 are the impedances of the arms, and are vector complex quantities that possess phase angles. It is thus necessary to adjust both the magnitude and phase angles of the impedance arms to achieve balance, i.e. the bridge must be balanced for both the reactance and the resistive component.

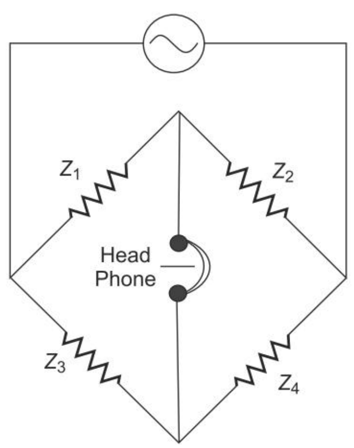


Fig. 11.17 ac Wheatstone's bridge

CAPACITANCE COMPARISON BRIDGE 11.9

Figure 11.18 shows the circuit of a capacitance comparison bridge. The ratio arms R_1, R_2 are resistive. The known standard capacitor C_3 is in series with R_3. R_3 may also include an added variable resistance needed to balance the bridge. C_x is the unknown capacitor and R_x is the small leakage resistance of the capacitor. In this case an unknown capacitor is compared with a standard capacitor and the value of the former, along with its leakage resistance, is obtained. Hence.

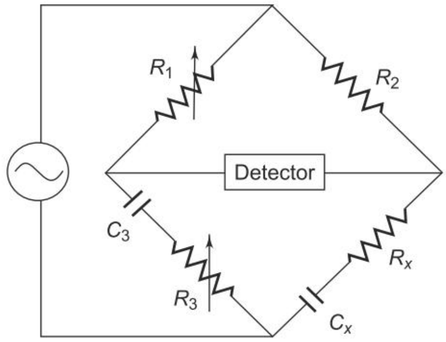


Fig. 11.18 Capacitance comparison bridge

$$Z_1 = R_1$$
$$Z_2 = R_2$$
$$Z_3 = R_3 \text{ in series with } C_3 = R_3 - j/\omega C_3$$
$$Z_x = R_x \text{ in series with } C_x = R_x - j/\omega C_x$$

The condition for balance of the bridge is

$$Z_1\, Z_x = Z_2\, Z_3$$

i.e. $$R_1\left(R_x - \frac{j}{\omega C_x}\right) = R_2\left(R_3 - \frac{j}{\omega C_3}\right)$$

$\therefore$ $$R_1\, R_x - \frac{j R_1}{\omega C_x} = R_2\, R_3 - \frac{j R_2}{\omega C_3}$$

Two complex quantities are equal when both their real and their imaginary terms are equal. Therefore,

i.e. $R_1\, R_x = R_2\, R_3$ $\qquad \therefore R_x = \dfrac{R_2\, R_3}{R_1}$ [11.12(a)]

and $\dfrac{R_1}{\omega C_x} = \dfrac{R_2}{\omega C_3}$ $\qquad C_x = \dfrac{C_3\, R_1}{R_2}$ [11.12(b)]

Since R_3 does not appear in the expression for C_x, as a variable element it is an obvious choice to eliminate any interaction between the two balance controls.

Example 11.6 (a) *A capacitance comparison bridge (similar angle bridge) is used to measure a capacitive impedance at a frequency of 2 kHz. The bridge constants at balance are $C_3 = 100\ \mu F$, $R_1 = 10\ k\Omega$, $R_2 = 50\ k\Omega$, $R_3 = 100\ k\Omega$. Find the equivalent series circuit of the unknown impedance.*

Solution Finding R_x using the equation

$$R_x = \frac{R_2 R_3}{R_1}$$

$$= \frac{100\text{ k} \times 50\text{ k}}{10\text{ k}} = 500\text{ k}\Omega$$

Then finding C_x using the equation

$$C_x = \frac{R_1}{R_2} C_3$$

$$= \frac{10\text{ k}}{50\text{ k}} \times 100 \times 10^{-6} = 20\ \mu\text{F}$$

The equivalent series circuit is shown in Fig. 11.19.

Fig. 11.19

Example 11.6 (b) *In the measurement of capacitance using capacitance bridge comparison bridge.*

R_1 in branch BC = 2000 Ω

R_2 in branch CD = 2850 Ω

R_4 in branch DA = 52 Ω in series with C_4 = 0.5µF

R_x in series with C_x in branch AB (unknowns)

f = 400 Hz.

Solution

Step 1: $$R_x = \frac{R_1 R_4}{R_2} = \frac{2000}{2850} \times 52 = 36.5\ \Omega$$

Step 2: $$C_x = \frac{R_2}{R_1} \times C_4 = \frac{2850}{2000} \times 0.5\ \mu\text{F} = 0.7125\ \mu\text{F}$$

Step 3: *L*oss angle of the capacitor (a series *RC* circuit) is defined as the angle by which current departs an exact quadrature from the applied voltage. 'δ' is the loss angle of the capacitor and is given by tan δ.

$$\tan\delta = \frac{R_x}{X_x} = \omega C_x R_x = 2\pi f C_x R_x$$

$$= 2 \times 3.14 \times 400 \times 36.5 \times 0.7125\ \mu\text{F} = 0.06533$$

Hence δ = 3º 74'

INDUCTANCE COMPARISON BRIDGE 11.10

Figure 11.20 gives a schematic diagram of an inductance comparison bridge. In this, values of the unknown inductance L_x and its internal resistance R_x are obtained by comparison with the standard inductor and resistance, i.e. L_3 and R_3.

The equation for balance condition is

$$Z_1 Z_x = Z_2 Z_3.$$

The inductive balance equation yields

$$L_x = \frac{L_3 R_2}{R_1} \qquad [11.13(a)]$$

and resistive balance equations yields

$$R_x = \frac{R_2 R_3}{R_1} \qquad [(11.13(b)]$$

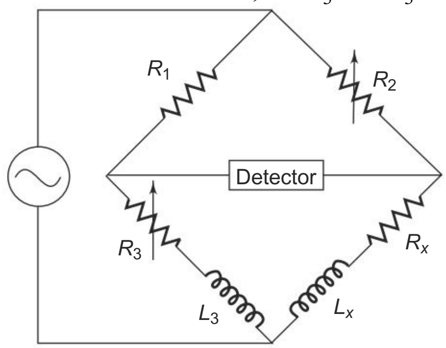


Fig. 11.20 Inductance comparison bridge

In this bridge R_2 is chosen as the inductive balance control and R_3 as the resistance balance control. (It is advisable to use a fixed resistance ratio and variable standards). Balance is obtained by alternately varying L_3 or R_3. If the Q of the unknown reactance is greater than the standard Q, it is necessary to place a variable resistance in series with the unknown reactance to obtain balance.

If the unknown inductance has a high Q, it is permissible to vary the resistance ratio when a variable standard inductor is not available.

Example 11.7 *An inductance comparison bridge is used to measure inductive impedance at a frequency of 5 KHz.The bridge constants at balance are L_3 = 10 mH, R_1 = 10 kΩ, R_2 = 40 KΩ, R_3 = 100 KΩ. Find the equivalent series circuit of the unknown impedance.*

Solution Given L_3 = 10 mH, R_1 = 10 kΩ, R_2 = 40 kΩ, R_3 = 100 kΩ. To find R_x and L_x.

From balance equation,

Step 1: $$R_x = \frac{R_2\ R_3}{R_1} = \frac{40\text{ K} \times 100\text{ K}}{10\text{ K}} = 400\text{ k}\Omega$$

Step 2: $$L_x = \frac{R_2\ L_3}{R_1} = \frac{10\text{ mH} \times 40\text{ K}}{10\text{ K}} = 40\text{ mH}$$

The equivalent series circuit is shown in Fig. 11.21.

Fig. 11.21

MAXWELL'S BRIDGE 11.11

Maxwell's bridge, shown in Fig. 11.22, measures an unknown inductance in terms of a known capacitor. The use of standard arm offers the advantage of compactness and easy shielding. The capacitor is almost a loss-less component. One arm has a resistance R_1 in parallel with C_1, and hence it is easier to write the balance equation using the admittance of arm 1 instead of the impedance.

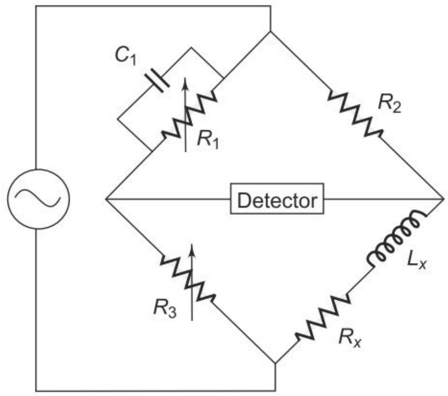


Fig. 11.22 Maxwell's bridge

The general equation for bridge balance is

$$Z_1 Z_x = Z_2 Z_3$$

i.e. $$Z_x = \frac{Z_2 Z_3}{Z_1} = Z_2 Z_3 Y_1 \tag{11.14}$$

Where $$Z_1 = R_1 \text{ in parallel with } C_1 \text{ i.e. } Y_1 = \frac{1}{Z_1}$$

$$Y_1 = \frac{1}{R_1} + j\omega C_1$$

$$Z_2 = R_2$$

$$Z_3 = R_3$$

$$Z_x = R_x \text{ in series with } L_x = R_x + j\omega L_x$$

From Eq. (11.14) we have

$$R_x + j\omega L_x = R_2 R_3 \left(\frac{1}{R_1} + j\omega C_1\right)$$

$$R_x + j\omega L_x = \frac{R_2 R_3}{R_1} + j\omega C_1 R_2 R_3$$

Equating real terms and imaginary terms we have

$$R_x = \frac{R_2 R_3}{R_1} \text{ and } L_x = C_1 R_2 R_3 \tag{11.15}$$

Also $$Q = \frac{\omega L_x}{R_x} = \frac{\omega C_1 R_2 R_3 \times R_1}{R_2 R_3} = \omega C_1 R_1$$

Maxwell's bridge is limited to the measurement of low Q values (1 – 10). The measurement is independent of the excitation frequency. The scale of the resistance can be calibrated to read inductance directly.

The Maxwell bridge using a fixed capacitor has the disadvantage that there is an interaction between the resistance and reactance balances. This can be avoided by varying the capacitances, instead of R_2 and R_3, to obtain a reactance balance. However, the bridge can be made to read directly in Q.

The bridge is particularly suited for inductances measurements, since comparison with a capacitor is more ideal than with another inductance. Commercial bridges measure from 1 – 1000 H, with ± 2% error. (If the Q is very large, R_1 becomes excessively large and it is impractical to obtain a satisfactory variable standard resistance in the range of values required).

Example 11.8 (a) *A Maxwell bridge is used to measure an inductive impedance. The bridge constants at balance are*
$C_1 = 0.01\ \mu F$, $R_1 = 470\ k\Omega$, $R_2 = 5.1\ k\Omega$, and $R_3 = 100\ k\Omega$.
Find the series equivalent of the unknown impedance.

Solution We need to find R_x and L_x.

$$R_x = \frac{R_2 R_3}{R_1} = \frac{100\text{ k} \times 5.1\text{ k}}{470\text{ k}} = 1.09\text{ k}\Omega$$

$$\begin{aligned} L_x &= R_2 R_3 C_1 \\ &= 5.1\text{ k} \times 100\text{ k} \times 0.01\ \mu\text{f} \\ &= 5.1\text{ H} \end{aligned}$$

The equivalent series circuit is shown in Fig. 11.23.

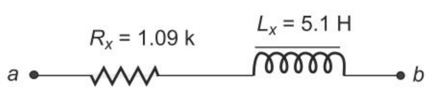


Fig. 11.23

Example 11.8 (b) *The arms of an ac Maxwell's bridge are arranged as follows:*
AB and BC are non-reactive resistors of 100 Ω each, DA a standard variable reactor L_1 of resistance 32.7 Ω and CD consists of a standard variable resistor R in series with a coil of unknown impedance Z, balance was found with L_1 = 50 mH and Z = 1.36R. Find the R and L of coil.

Solution Given: $R_1 = 32.7\ \Omega$, $L_1 = 50$ mH
$R_2 = 1.36\ \Omega$, $R_3 = 100\ \Omega$, $R_4 = 100\ \Omega$

Step 1: To find 'r' and L_2 where r is the resistance of the coil

Given that $R_4 R_1 = R_3(R_2 + r)$

$\therefore$ $32.7 \times 100 = 100\ (1.36 + r)$

$\therefore$ $100(32.7 - 1.36) = 100\ r$

$\therefore$ $r = 32.7 - 1.36$

$r = 31.34\ \Omega$

Step 2: To find L_2, $L_2 = L_1 \times \dfrac{R_4}{R_3} = 50\text{ mH} \times \dfrac{100}{100}$

$\therefore$ $L_2 = 50$ mH

HAY'S BRIDGE 11.12

The Hay bridge, shown in Fig. 11.24, differs from Maxwell's bridge by having a resistance R_1 in series with a standard capacitor C_1 instead of a parallel. For large phase angles, R_1 needs to be low; therefore, this bridge is more convenient for measuring high-Q coils. For $Q = 10$, the error is ± 1%, and for $Q = 30$, the error is ± 0.1%. Hence Hay's bridge is preferred for coils with a high Q, and Maxwell's bridge for coils with a low Q.

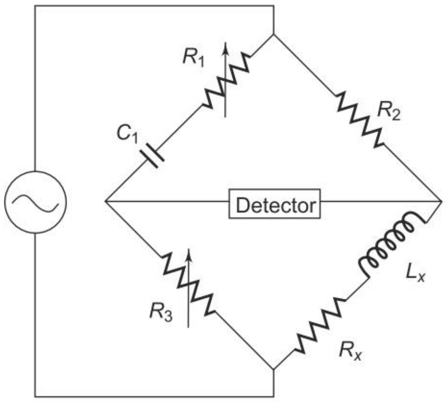


Fig. 11.24 Hay's Bridge

At balance $\quad Z_1 Z_x = Z_2 Z_3$, where

$$Z_1 = R_1 - j/\omega C_1$$
$$Z_2 = R_2$$
$$Z_3 = R_3$$
$$Z_x = R_x + j\omega L_x$$

Substituting these values in the balance equation we get

$$\left(R_1 - \frac{j}{\omega c_1}\right)(R_x + j\omega L_x) = R_2 R_3$$

$$R_1 R_x + \frac{L_x}{C_1} - \frac{j R_x}{\omega C_1} + j\omega L_x R_1 = R_2 R_3$$

Equating the real and imaginary terms we have

$$R_1 R_x + \frac{L_x}{C_1} = R_2 R_3 \tag{11.16}$$

and

$$\frac{R_x}{\omega C_1} = \omega L_x R_1 \tag{11.17}$$

Solving for L_x and R_x we have, $R_x = \omega^2 L_x C_1 R_1$.
Substituting for R_x in Eq. (11.16)

$$R_1 (\omega^2 R_1 C_1 L_x) + \frac{L_x}{C_1} = R_2 R_3$$

$$\omega^2 R_1^2 C_1 L_x + \frac{L_x}{C_1} = R_2 R_3$$

Multiplying both sides by C_1 we get

$$\omega^2 R_1^2 C_1^2 L_x + L_x = R_2 R_3 C_1$$

Therefore, $$L_x = \frac{R_2 R_3 C_1}{1 + \omega^2 R_1^2 C_1^2} \tag{11.18}$$

Substituting for L_x in Eq. (11.17)

$$R_x = \frac{\omega^2 C_1^2 R_1 R_2 R_3}{1 + \omega^2 R_1^2 C_1^2} \tag{11.19}$$

The term ω appears in the expression for both L_x and R_x. This indicates that the bridge is frequency sensitive.

The Hay bridge is also used in the measurement of incremental inductance. The inductance balance equation depends on the losses of the inductor (or Q) and also on the operating frequency.

An inconvenient feature of this bridge is that the equation giving the balance condition for inductance, contains the multiplier $1/(1 + 1/Q^2)$. The inductance balance thus depends on its Q and frequency.

Therefore, $$L_x = \frac{R_2 R_3 C_1}{1 + (1/Q)^2}$$

For a value of Q greater than 10, the term $1/Q^2$ will be smaller than 1/100 and can be therefore neglected.

Therefore $L_x = R_2 R_3 C_1$, which is the same as Maxwell's equation. But for inductors with a Q less than 10, the $1/Q^2$ term cannot be neglected. Hence this bridge is not suited for measurements of coils having Q less than 10.

A commercial bridge measure from 1 μ H – 100 H with ± 2% error.

Example 11.9 (a) *Find the series equivalent inductance and resistance of the network that causes an opposite angle (Hay bridge) to null with the following bridges arms. (See Fig. 11.25.)*

$\omega = 3000$ rad/s, $R_2 = 10\ k\Omega$,
$R_1 = 2\ k\Omega$, $C_1 = 1\ \mu F$
$R_3 = 1\ k\Omega$

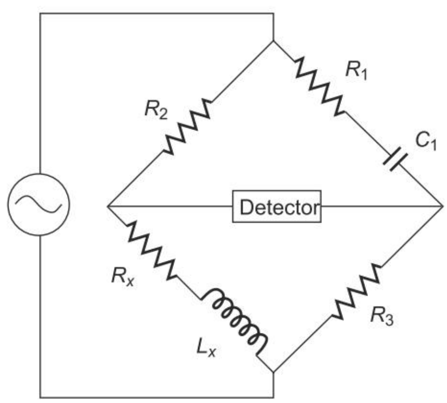


Fig. 11.25

Solution We need to find R_x and L_x.

From Eq. (11.19) we have

$$R_x = \frac{\omega^2 R_1 R_2 R_3 C_1^2}{1+\omega^2 R_1^2 C_1^2}$$

$$= \frac{(3000)^2 \times 10\text{ k} \times 2\text{ k} \times 1\text{ k} \times (1\times 10^{-6})^2}{1+(3000)^2 \times (2\text{ k})^2 \times (1\times 10^{-6})^2}$$

$$= \frac{180\times 10^3}{1+36} = \frac{180}{37} \times 10^3$$

$$= 4.86\text{ k}\Omega$$

and from Eq. (11.18) we have,

$$L_x = \frac{R_2 R_3 C_1}{1+\omega^2 R_1^2 C_1^2}$$

$$= \frac{10\text{ k} \times 1\text{ k} \times (1\times 10^{-6})}{1+(3000)^2 \times (2\text{ k})^2 \times (1\times 10^{-6})^2}$$

$$= \frac{10}{1+36} = \frac{10}{37} = 0.27 = 270\text{ mH}$$

Therefore $R_x = 4.86$ k and $L_x = 270$ mH

Example 11.9 (b) *Four arms of a Hay Bridge are arranged as follows: AD is coil of unknown impedance Z, DC is a non-inductive resistance of 1 kΩ, CB is a non-inductive resistance of 800 Ω in series with a standard capacitor of 2 μF, BA is a non-inductive resistance of 16500 Ω, if the supply frequency is 50 Hz. Calculate the value of L and R of coil When the bridge is balanced.*

Solution Given $R_2 = 1000\ \Omega$, $R_3 = 16500\ \Omega$, $R_4 = 800\ \Omega$, $C_4 = 2\ \mu\text{F}$, $f = 50$ Hz

Step 1:

$\therefore \quad \omega = 2\pi f = 2 \times 3.14 \times 50 = 314$ and $\omega^2 = (314)^2 = 98596$

Step 2:

$$L_x = \frac{R_2 R_3 C_4}{1+\omega^2 C_4^2 R_4^2}, \quad R_x = \frac{\omega^2 C_4^2 R_4 R_2 R_3}{1+\omega^2 C_4^2 R_4^2}$$

$$\therefore \quad L_1 = \frac{(1000)\times 16500 \times 2 \times 10^{-6}}{1+98596 \times (2\mu\text{F})^2 \times (800)^2} = 26.4\text{ H}$$

Step 3: $$R_x = \frac{\omega^2 C_4^2 R_4 R_2 R_3}{1+\omega^2 C_4^2 R_4^2}$$

$$R_1 = \frac{(314)^2 \times (2\mu\text{F})^2 \times 16500 \times 800 \times 1000}{1+(314)^2 \times (2\mu\text{F})^2 \times (800)^2} = 4.18\text{ k}\Omega$$

Example 11.9 (c) *Find the unknown resistance and inductance having the following bridge arms*

$C_4 = 1\ \mu F,\ R_2 = R_3 = R_4 = 1000\ \Omega,\ \omega = 314$ rad/s

Solution To find R_1 and L_1

Step 1: Given

$$L_1 = \frac{R_2 R_3 C_4}{1+\omega^2 C_4^2 R_4^2},\ R_1 = \frac{\omega^2 C_4^2 R_4 R_2 R_3}{1+\omega^2 C_4^2 R_4^2}$$

$$\therefore \quad L_1 = \frac{1000 \times 1000 \times 1 \times 10^{-6}}{1+(314)^2 \times (1\,\mu\text{F})^2 \times (1000)^2} = 0.91\ \text{H}$$

Step 2:

$$R_x = \frac{\omega^2 C_4^2 R_4 R_2 R_3}{1+\omega^2 C_4^2 R_4^2}$$

$$R_1 = \frac{(314)^2 \times (1\,\mu\text{F})^2 \times 1000 \times 1000 \times 1000}{1+(314)^2 \times (1\,\mu\text{F})^2 \times (1000)^2} \approx 89.79\ \Omega$$

SCHERING'S BRIDGE 11.13

A very important bridge used for the precision measurement of capacitors and their insulating properties is the Schering bridge. Its basic circuit arrangement is given in Fig. 11.26. The standard capacitor C_3 is a high quality mica capacitor (low-loss) for general measurements, or an air capacitor (having a very stable value and a very small electric field) for insulation measurement.

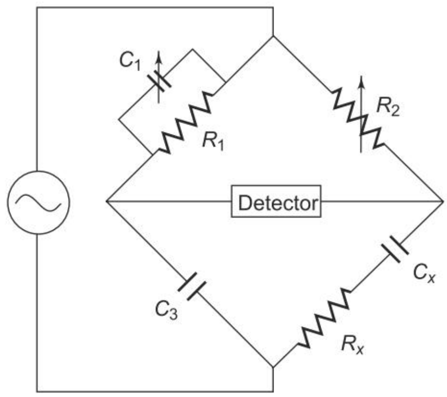


Fig. 11.26 Schering's bridge

For balance, the general equation is

$$Z_1 Z_x = Z_2 Z_3$$

$$\therefore \quad Z_x = \frac{Z_2 Z_3}{Z_1},\ Z_x = Z_2 Z_3 Y_1$$

where
$$Z_x = R_x - j/\omega C_x$$
$$Z_2 = R_2$$
$$Z_3 = -j/\omega C_3$$
$$Y_1 = 1/R_1 + j\,\omega C_1$$

as
$$Z_x = Z_2 Z_3 Y_1$$

$$\therefore \quad \left(R_x - \frac{j}{\omega C_x}\right) = R_2 \left(\frac{-j}{\omega C_3}\right) \times \left(\frac{1}{R_1} + j\omega C_1\right)$$

$$\left(R_x - \frac{j}{\omega C_x}\right) = \frac{R_2\,(-j)}{R_1\,(\omega C_3)} + \frac{R_2\,C_1}{C_3}$$

Equating the real and imaginary terms, we get

$$R_x = \frac{R_2\,C_1}{C_3} \qquad [11.20(a)]$$

and

$$C_x = \frac{R_1}{R_2}\,C_3 \qquad [11.20(b)]$$

The dial of capacitor C_1 can be calibrated directly to give the dissipation factor at a particular frequency.

The dissipation factor D of a series RC circuit is defined as the cotangent of the phase angle.

$$D = \frac{R_x}{X_x} = \omega\,C_x\,R_x$$

Also, D is the reciprocal of the quality factor Q, i.e. $D = 1/Q$. D indicates the quality of the capacitor.

Commercial units measure from 100 pf – 1 μf, with ± 2% accuracy. The dial of C_3 is graduated in terms of direct readings for C_x, if the resistance ratio is maintained at a fixed value.

This bridge is widely used for testing small capacitors at low voltages with very high precision.

The lower junction of the bridge is grounded. At the frequency normally used on this bridge, the reactances of capacitor C_3 and C_x are much higher than the resistances of R_1 and R_2. Hence, most of the voltage drops across C_3 and C_x, and very little across R_1 and R_2. Hence if the junction of R_1 and R_2 is grounded, the detector is effectively at ground potential. This reduces any stray-capacitance effect, and makes the bridge more stable.

Example 11.10 (a) *An ac bridge has the following constants (refer Fig. 11.27).*

Arm AB — capacitor of 0.5 μF in parallel with 1 kΩ resistance

Arm AD — resistance of 2 kΩ

Arm BC — capacitor of 0.5 μF

Arm CD — unknown capacitor C_x and R_x in series

Frequency — 1 kHz

Determine the unknown capacitance and dissipation factor.

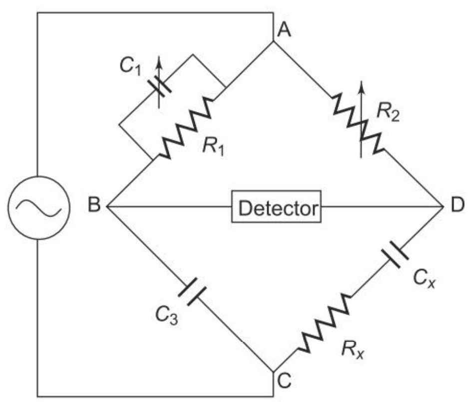


Fig. 11.27

Solution From Eqs 11.20(a) and 11.20(b), we have

$$R_x = \frac{C_1}{C_3}\, R_2 = \frac{0.5\ \mu\text{F}}{0.5\ \mu\text{F}} \times 2\ \text{k} = 2\ \text{k}\Omega$$

$$C_x = \frac{R_1}{R_2} \times C_3 = \frac{1\ \text{k}}{2\ \text{k}} \times 0.5\ \mu\text{F} = 0.25\ \mu\text{F}$$

The dissipation factor is given by

$$\begin{aligned} D &= \omega\, C_x\, R_x \\ &= 2 \times 3.142 \times 1000 \times 2\ \text{k} \times 0.25\ \mu\text{F} \\ &= 4 \times 3.142 \times 0.25 \\ &= 3.1416 \end{aligned}$$

Example 11.10 (b) *A sample of insulation was placed in arm AB of a Schering bridge, when the bridge was balanced at a frequency of 50 Hz, the other arms of the bridge were as follows*

Arm BC – a non-inductive R of 100 Ω

Arm CD – a non-inductive R of 300 Ω in parallel with a capacitor of 0.5 μF

Arm DA – a loss free capacitor of 100 pf

Determine the capacitance, equivalent series resistance and PF of the insulation in test arm AB

Solution Given $R_3 = 100\ \Omega$

$R_4 = 300\ \Omega$, $C_4 = 0.5\ \mu\text{F}$

$C_2 = 100$ pf , $f = 50$ Hz

Step 1:

$\therefore \quad \omega = 2\pi f = 2 \times 3.14 \times 50 = 314$

Step 2 : From the balance condition, we have

$$C_1 = \frac{R_4}{R_3} \times C_2 = \frac{300\ \Omega}{100\ \Omega} \times 100\ \text{pf} = 300\ \text{pf}$$

Step 3:

$$R_1 = \frac{C_4}{C_2} \times R_3 = \frac{0.5\ \mu\text{F}}{100\ \text{pf}} \times 100$$

$$= \frac{0.5 \times 10^{-6}}{100 \times 100^{-12}} \times 100 = 0.5 \times 10^{6}\ \Omega = 0.5\ \text{M}\Omega$$

Step 4: Power factor

$$\begin{aligned} &= \omega R_4 C_4 \\ &= 314 \times 300 \times 0.5 \times 10^{-6} \\ &= 314 \times 30\cancel{0} \times \frac{5}{1\cancel{0}} \times 10^{-6} \\ &= 4710 \times 10^{-6} = 0.0471 \end{aligned}$$

Example 11.10 (c) *A condenser bushing forms arm BC of a Schering Bridge and a standard capacitor of 500 pf and negligible loss forms an arm AB. Arm CD consists of a non-inductive resistance of 300 Ω. When this bridge is balanced, arm AD has a resistance and capacitor in parallel of 100 Ω and 0.1 μF respectively. The supply frequency is 50 Hz. Calculate the capacitance and dielectric loss of angle of the bushing.*

Solution Given $R_4 = 100\ \Omega$, $C_4 = 0.1\ \mu F$

$C_2 = 500$ pF, $R_3 = 300\ \Omega$

$f = 50$ Hz

Step 1:

$\therefore \quad \omega = 2\pi f = 2 \times 3.14 \times 50 = 314$

Step 2: $C_1 = \dfrac{R_4}{R_3} \times C_2 = \dfrac{100}{300} \times 500 \times 10^{-12} = 166.6$ pF

Step 3: Dielectric loss angle is given by

$$\begin{aligned}\tan\delta &= \omega C_4 R_4\\ &= 3.14 \times 0.1 \times 10^{-6} \times 100\\ &= 3.14 \times 10 \times 10^{-6}\\ \tan\delta &= 0.03140\end{aligned}$$

Hence $\delta = 1.8°$

Example 11.10 (d) *A sheet of 4.5-mm thick Bakelite is tested at 50 Hz between 12 cm in diameter. The Schering bridge uses a standard air capacitor C_2 of 105 pF capacitor, a non–reactive, R_4 of 1000/π in parallel with a variable capacitor and is obtained with C_4 = 0.5 μF and R_3 = 260 Ω. Calculate the capacitance, PF and relative permittivity of the sheet.*

Solution It is given that

d = thickness of sheet in metre = 4.5×10^{-3}

f = 50 Hz,

$\therefore \quad \omega = 2\pi f = 2 \times 3.14 \times 50 = 314$

A = area of the electrodes in m^2 = $\pi(6 \times 10^{-2})^2$

$C_2 = 105 \times 10^{-12}$, $R_4 = \dfrac{1000}{\pi}$, $C_4 = 0.5\ \mu F$, $R_3 = 260\ \Omega$

Step 1: $C_1 = \dfrac{R_4}{R_3} \times C_2 = \dfrac{1000}{\pi \times 260} \times 105\ \text{pF} = \dfrac{1000}{8164} \times 105\ \text{pf}$

$C_1 = 128.7$ pf

Step 2: PF is given by

$$\omega R_4 C_4 = 2 \times 3.14 \times 50 \times \frac{1000}{\pi} \times 0.5 \times 10^{-6}$$

$$= 2 \times \cancel{3.14} \times 50 \times \frac{1000}{\cancel{\pi}} \times 0.5 \times 10^{-6}$$

$$= 10^{-1} \times 0.5 = 0.05$$

Step 3: Given that the capacitance is given by $C_1 = K_r K_o \dfrac{A}{d}$

$\therefore$ Relative permittivity is given by

$$K_r = \frac{C_1 d}{K_o A}$$

$$= \frac{128.7\,\text{pF} \times 4.5 \times 10^{-3}}{8.854 \times 10^{-12} \times \pi(6 \times 10^{-2})^2}$$

$$= \frac{128.7 \times 4.5 \times 10^{-15}}{8.854 \times 10^{-12} \times 3.14 \times 36 \times 10^{-4}}$$

$$= \frac{579.15 \times 10^{-15}}{1000.85 \times 10^{-16}}$$

$$= 5.786$$

Example 11.10 (e) *A capacitor is tested by a Schering bridge which forms one arm AB of the bridge. The other arms are*

AD – a non-inductive resistance of 100 Ω,

DC – a non-reactive resistance of 300 Ω in parallel with a capacitor of 0.5 μF,

BC – a standard loss free capacitor of 100 pF,

The supply frequency is 50 Hz. The bridge is balanced.

Calculate the capacitor value and the power factor of the capacitor under test.

Solution Given $R_3 = 100\ \Omega$, $R_4 = 300\ \Omega$, $C_2 = 100$ pF, $C_4 = 0.5\ \mu$F

Let the desired capacitance of the capacitor to be tested be C_1 and r_1 is the resistance representing the loss.

Step 1: From the given equation for C_1 we have

$$C_1 = \frac{R_4}{R_3} \times C_2 = \frac{300}{100} \times 100\ \text{pF}$$

$$C_1 = 300\ \text{pf}$$

Step 2: $$r_1 = R_3 \times \frac{C_4}{C_2} = 100 \times \frac{0.5\ \mu\text{F}}{100\ \text{pF}} \times 500\ \text{k}\Omega$$

Step 3: The power factor can be as follows:

$$\text{PF} = \omega C_4 R_4 = 2\pi f \times C_4 R_4$$

$$= 2 \times 3.14 \times 50 \times 300 \times 0.5 \times 10^{-6}$$

$$= 0.0471$$

Example 11.10 (f) *A sample Bakelite was tested by the bridge method (Schering) at 11 kV,50 Hz. Balance was obtained at the followings values*
AB – dielectric material under test in the form of a capacitor
BC – a standard air capacitor of 100 pf
CD – capacitor of 0.6 μF in parallel with a non-reactive resistance of 300 Ω
DA – non–reactive resistance of 100 Ω
Calculate the capacitance and equivalent series resistance of the sample.

Solution The given bridge is of a Schering bridge. To find R_x and C_x
Given $R_1 = 300\ \Omega$, $R_2 = 100\ \Omega$, $C_1 = 0.6\ \mu F$, $C_3 = 100$ pf

Step 1: $$R_x = R_2 \times \frac{C_1}{C_3} = 100 \times \frac{0.6 \times 10^{-6}}{100 \times 10^{-12}} = 6\ \text{M}\Omega$$

Step 2: $$C_x = \frac{R_1}{R_2} \times C_3 = \frac{300}{100} \times 100 \times 10^{-12} = 300\ \text{pF}$$

Example 11.10(g) *An ac bridge has the following constants:*
Arm AB – capacitor of 0.1 μF in parallel with 2 kΩ resistor
Arm AD – resistance of 5 kΩ
Arm BC – capacitor of 0.25 μF
Arm AB – unknown capacitor C_x and R_x in series
f = 2 kHz
Determine the unknown capacitance and dissipation factor.

Solution From the balance equation for a Schering bridge, we have

Step 1: $$R_x = \frac{C_1}{C_2} \times R_2 = \frac{0.1\ \mu\text{F}}{0.25\ \mu\text{F}} \times 5\ \text{k}\Omega$$

$$= \frac{10}{25} \times 5\ \text{k}\Omega = 2\ \text{k}\Omega$$

Step 2: $$C_x = \frac{R_1}{R_2} \times C_3 = \frac{2\ \text{k}\Omega}{5\ \text{k}\Omega} \times 0.25\ \mu\text{F}$$

$$= \frac{2}{5} \times \frac{25\ \mu\text{F}}{100} = 0.1\ \mu\text{F}$$

Step 3 : Dissipation factor $(D) = \omega C_x R_x$

$$= 2 \times 3.142 \times 2000 \times 0.1\ \mu\text{F} \times 2\ \text{k}\Omega$$
$$= 2 \times 3.142 \times 4 \times 0.1$$
$$= 8 \times 3.142 \times 0.1$$
$$D = 2.5136$$

WIEN'S BRIDGE 11.14

The Wien bridge shown in Fig. 11.28 has a series RC combination in one arm and a parallel combination in the adjoining arm. Wien's bridge in its basic form, is designed to measure frequency. It can also be used for the measurement of an unknown capacitor with great accuracy.

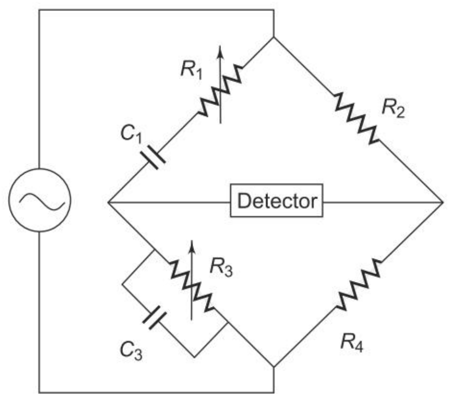


Fig. 11.28 Wein's bridge

The impedance of one arm is

$$Z_1 = R_1 - j/\omega C_1.$$

The admittance of the parallel arm is

$$Y_3 = 1/R_3 + j\,\omega\,C_3.$$

Using the bridge balance equation, we have $\quad Z_1\,Z_4 = Z_2\,Z_3.$

Therefore, $Z_1\,Z_4 = Z_2/Y_3$, i.e. $Z_2 = Z_1\,Z_4\,Y_3$.

$$\therefore \qquad R_2 = R_4\left(R_1 - \frac{j}{\omega C_1}\right)\left(\frac{1}{R_3} + j\,\omega C_3\right)$$

$$R_2 = \frac{R_1\,R_4}{R_3} - \frac{j\,R_4}{\omega\,C_1\,R_3} + j\,\omega\,C_3\,R_1\,R_4 + \frac{C_3\,R_4}{C_1}$$

$$R_2 = \left(\frac{R_1\,R_4}{R_3} + \frac{C_3\,R_4}{C_1}\right) - j\left(\frac{R_4}{\omega\,C_1\,R_3} - \omega\,C_3\,R_1\,R_4\right)$$

Equating the real and imaginary terms we have

$$R_2 = \frac{R_1\,R_4}{R_3} + \frac{C_3\,R_4}{C_1} \quad \text{and} \quad \frac{R_4}{\omega\,C_1\,R_3} - \omega\,C_3\,R_1\,R_4 = 0$$

$$\text{Therefore} \qquad \frac{R_2}{R_4} = \frac{R_1}{R_3} + \frac{C_3}{C_1} \tag{11.21}$$

$$\text{and} \qquad \frac{1}{\omega\,C_1\,R_3} = \omega\,C_3\,R_1 \tag{11.22}$$

$$\therefore \qquad \omega^2 = \frac{1}{C_1\,R_1\,R_3\,C_3}$$

$$\omega = \frac{1}{\sqrt{C_1\,R_1\,C_3\,R_3}}$$

$$\text{as} \qquad \omega = 2\,\pi f$$

$$\therefore \qquad f = \frac{1}{2\pi\sqrt{C_1\,R_1\,C_3\,R_3}} \tag{11.23}$$

The two conditions for bridge balance, (11.21) and (11.23), result in an expression determining the required resistance ratio R_2/R_4 and another expression determining the frequency of the applied voltage. If we satisfy Eq. (11.21) and also excite the bridge with the frequency of Eq. (11.23), the bridge will be balanced.

In most Wien bridge circuits, the components are chosen such that $R_1 = R_3 = R$ and $C_1 = C_3 = C$. Equation (11.21) therefore reduces to $R_2/R_4 = 2$ and Eq. (11.23) to $f = 1/2\pi RC$, which is the general equation for the frequency of the bridge circuit.

The bridge is used for measuring frequency in the audio range. Resistances R_1 and R_3 can be ganged together to have identical values. Capacitors C_1 and C_3 are normally of fixed values.

The audio range is normally divided into 20 – 200 – 2 k – 20 kHz ranges. In this case, the resistances can be used for range changing and capacitors C_1 and C_3 for fine frequency control within the range. The bridge can also be used for measuring capacitances. In that case, the frequency of operation must be known.

The bridge is also used in a harmonic distortion analyzer, as a Notch filter, and in audio frequency and radio frequency oscillators as a frequency determining element.

An accuracy of 0.5% – 1% can be readily obtained using this bridge. Because it is frequency sensitive, it is difficult to balance unless the waveform of the applied voltage is purely sinusoidal.

Example 11.11 *A Wien bridge circuit consists of the following:*

$R_1 = 4.7\ k\Omega$, $C_1 = 5\ nf$
$R_2 = 20\ k\ \Omega$, $C_3 = 10\ nf$
$R_3 = 10\ k\ \Omega$
$R_4 = 100\ k\ \Omega$.

Determine the frequency of the circuit.

Solution The frequency is given by the equation

$$f = \frac{1}{2\pi\sqrt{C_1\ R_1\ R_3\ C_3}}$$

$$f = \frac{1}{2\pi\sqrt{5\times10^{-9}\times4.7\times10^{3}\times10\times10^{-9}\times10\times10^{3}}}$$

$$f = \frac{1}{2\pi\sqrt{5\times10^{-10}\times4.7}}$$

$$f = \frac{10^{5}}{2\pi\sqrt{5\times4.7}} = 3.283\ \text{kHz}$$

Example 11.12 *Find the equivalent parallel resistance and capacitance that causes a Wien bridge to null with the following component values.*

$R_1 = 3.1\ k\Omega$
$C_1 = 5.2\ \mu F$
$R_2 = 25\ k\Omega$
$f = 2.5\ kHz$
$R_4 = 100\ k\Omega$

Solution Given $\omega = 2\,\pi f = 2 \times 3.14 \times 2500 = 15.71$ k rad/s.

Substituting the value of C_3 from Eq. (11.22) in Eq. (11.21) we get,

$$R_3 = \frac{R_4}{R_2}\left(R_1 + \frac{1}{\omega^2 R_1 C_1^2}\right)$$

$$= \frac{100\text{ k}}{25\text{ k}}\left(3.1\text{ k} + \frac{1}{(15.71\text{ k})^2 \times 3.1\text{ k} \times (5.2\times10^{-6})^2}\right)$$

$$= 12.4\text{ k}\Omega$$

$$C_3 = \frac{R_2}{R_4}\left(\frac{C_1}{1+\omega^2 R_1^2 C_1^2}\right)$$

$$= \frac{25\text{ k}}{100\text{ k}}\left(\frac{5.2\times10^{-6}}{1+(15.71\text{ k})^2\times(3.1\text{ k})^2\times(5.2\times10^{-6})^2}\right)$$

$$= 1.3\times10^{-6}\left(\frac{1}{1+64133.07}\right)$$

$$= 20.3\text{ pf}$$

The value of C_3 can also be found out by using equation $C_3 = \dfrac{1}{\omega^2 C_1 R_1 R_3}$.

Example 11.13 *An ac bridge with terminals ABCD has in*
Arm AB a resistance of 800 Ω in parallel with a capacitor of 0.5 μF,
Arm BC – a resistance of 400 Ω in series with a capacitor of 1 μF,
Arm CD – a resistance of 1000 Ω, Arm DA – a pure resistance R.
(a) Determine the value of frequency for which the bridge is balanced
(b) Calculate the value of R required to produce balance.

Solution The bridge configuration is of Wien Bridge.

Given : $C_1 = 0.5\ \mu F$, $R_1 = 800\ \Omega$
$C_2 = 1.0\ \mu F$, $R_2 = 400\ \Omega$
$R_4 = 1000\ \Omega$, $R_3 = R = ?$

Step 1 : Frequency calculated by

$$f = \frac{1}{2\pi\sqrt{R_1 C_1 R_2 C_2}}$$

$$= \frac{1}{2\pi\sqrt{800 \times 0.5\,\mu\text{F} \times 400 \times 1\,\mu\text{F}}}$$

$$= \frac{1}{2\pi\sqrt{800 \times 400 \times 0.5 \times 10^{-12}}}$$

$$= \frac{10^6}{2\pi\sqrt{800 \times 200}}$$

$$= \frac{10^6}{2\pi \times 400} = \frac{1000\text{ kHz}}{2 \times 3.14 \times 400} = \frac{1000}{314 \times 8} = 0.398\text{ kHz}$$

Step 2 : Also given,

$$\frac{R_2}{R_1} + \frac{C_1}{C_2} = \frac{R_4}{R_3}$$

$$\therefore \quad \frac{400}{800} + \frac{0.5\,\mu\text{F}}{1\,\mu\text{F}} = \frac{1000}{R}$$

$$\therefore \quad 0.5 + 0.5 = \frac{1000}{R}$$

$$\therefore \quad R = 1000\ \Omega$$

WAGNER'S EARTH (GROUND) CONNECTION 11.15

When performing measurements at high frequency, stray capacitances between the various bridge elements and ground, and between the bridge arms themselves, becomes significant. This introduces an error in the measurement, when small values of capacitance and large values of inductance are measured.

An effective method of controlling these capacitances, is to enclose the elements by a shield and to ground the shield. This does not eliminate the capacitance, but makes it constant in value.

Another effective and popular method of eliminating these stray capacitances and the capacitances between the bridge arms is to use a Wagner's ground connection. Figure 11.29 shows a circuit of a capacitance bridge. C_1 and C_2 are the stray capacitances. In Wagner's ground connection, another arm, consisting of R_w and C_w forming a potential divider, is used. The junction of R_w and C_w is grounded and is called Wagner's ground connection. The procedure for adjustment is as follows.

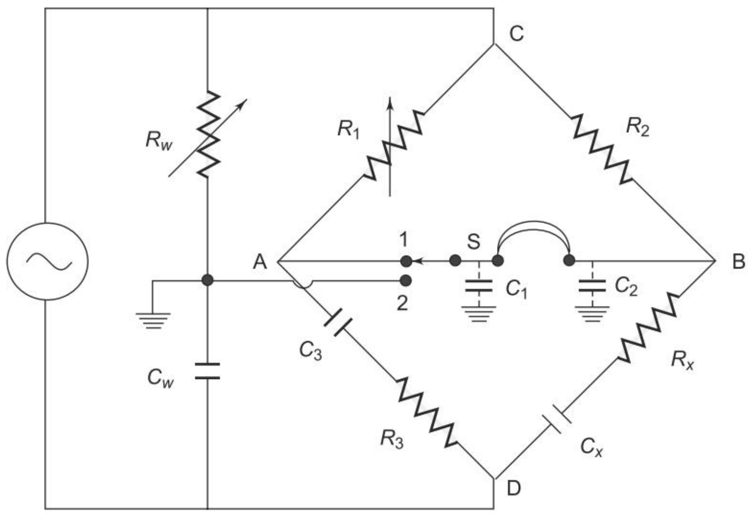


Fig. 11.29 Wagner's earth connection

The detector is connected to point 1 and R_1 is adjusted for null or minimum sound in the headphones. The switch S is then connected to point 2, which connects the detector to the Wagner ground point. Resistor R_w is now adjusted for minimum sound. When the switch 'S' is connected to point 1, again there will be some imbalance. Resistors R_1 and R_3 are then adjusted for minimum sound and this procedure is repeated until a null is obtained on both switch positions 1 and 2. This is the ground potential. Stray capacitances C_1 and C_2 are then effectively short-circuited and have no effect on the normal bridge balance.

The capacitances from point C to D to ground are also eliminated by the addition of Wagner's ground connection, since the current through these capacitors enters Wagner's ground connection.

The addition of the Wagner ground connection does not affect the balance conditions, since the procedure for measurement remains unaltered.

RESONANCE BRIDGE 11.16

One arm of this bridge, shown in Fig. 11.30, consists of a series resonance circuit. The series resonance circuit is formed by R_d, C_d and L_d in series. All the other arms consists of resistors only.

Using the equation for balance, we have $Z_1 Z_4 = Z_2 Z_3$,
where $Z_1 = R_b$, $Z_2 = R_c$, $Z_3 = R_a$, and $Z_4 = R_d + j\omega L_d - j/\omega C_d$.

Therefore $$R_b\left(R_d + j\omega L_d - \frac{j}{\omega C_d}\right) = R_a R_c$$

$$\therefore \quad R_b R_d + j\,\omega L_d R_b - \frac{j R_b}{\omega C_d} = R_a R_c$$

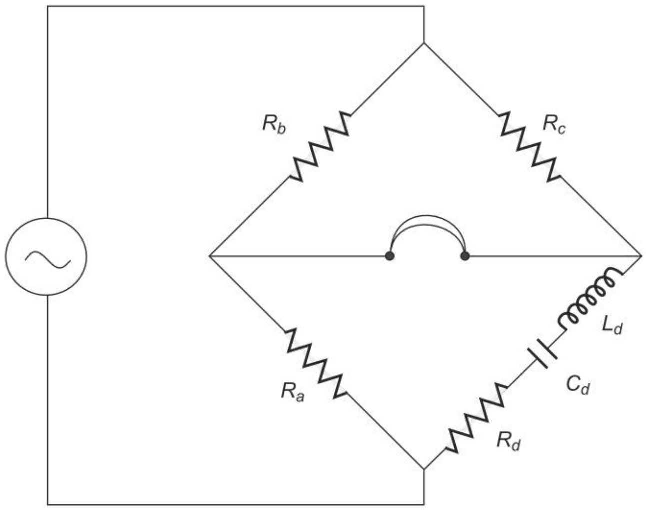


Fig. 11.30 Resonance bridge

Equating the real and imaginary terms

we get $\quad R_b\, R_d = R_a\, R_c$ and $j\,\omega L_d - \dfrac{j}{\omega C_d} = 0$

Therefore $\quad R_d = \dfrac{R_a\, R_c}{R_b}$ and $\omega L_d = \dfrac{1}{\omega C_d}$ i.e. $\omega^2 = \dfrac{1}{L_d\, C_d}$

Therefore
$$f = \frac{1}{2\pi\sqrt{L_d\, C_d}} \tag{11.24}$$

The bridge can be used to measure unknown inductances or capacitances. The losses R_d can be determined by keeping a fixed ratio R_a/R_b and using a standard variable resistance to obtain balance. If an inductance is being measured, a standard capacitor is varied until balance is obtained. If a capacitance is being measured, a standard inductor is varied until balance is obtained.The operating frequency of the generator must be known in order to calculate the unknown quantity. Balance is indicated by the minimisation of sound in the headphones.

MAXWELL–WIEN BRIDGE 11.17

As seen before, a positive phase angle of inductive impedance can be compensated by the negative phase angle of capacitive impedance, which is placed in the opposite arms *CD*. As shown in Fig. 11.31, the unknown inductance can be determined in terms of capacitance.

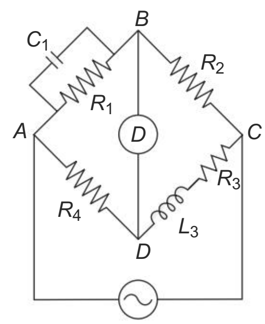


Fig 11.31 Maxwell's–Wein bridge

Balance condition is obtained when

$$Z_1Z_3 = Z_2Z_4$$

But

$$Z_1 = R_1//jXc_1,\ \frac{1}{Z_1} = \frac{1}{R_1} + \frac{1}{jXc_1} = \frac{1}{R_1} + j\omega C_1 = \frac{1 + j\omega C_1 R_1}{R_1}$$

Therefore, $$Z_1 = \frac{R_1}{1 + j\omega\, C_1\, R_1}$$

$$Z_2 = R_2,\ Z_4 = R_4 \text{ and } Z_3 = R_3 + j\omega L_3$$

Using balance condition, $Z_1Z_3 = Z_2Z_4$

Therefore, $$\left(\frac{R_1}{1 + j\omega\, C_1\, R_1}\right)(R_3 + j\omega L_3) = R_2R_4$$

Therefore, $$R_1R_3 + j\omega L_3R_1 = R_2R_4\,(1 + j\omega C_1R_1)$$

$$R_1R_3 + j\omega L_3R_1 = R_2R_4 + j\omega C_1R_1\, R_2R_4$$

Equating the real and imaginary terms, we have

$$R_1R_3 = R_2R_4 \text{ therefore } R_3 = \frac{R_2R_4}{R_1}$$

and $$j\omega L_3R_1 = j\omega C_1R_1\, R_2R_4$$

Therefore, $L_3 = C_1R_2R_4$

Hence the unknown resistance R_3 and unknown inductance L_3 can be determined

Example 11.11

From Fig. 11.31,

$$R_1R_3 = R_2R_4, \text{ therefore } R_3 = \frac{R_2R_4}{R_1} = \frac{600 \times 400}{1000} = 240\ \Omega$$

$L_3 = C_1R_2R_4 = 0.5 \times 10^{-6} \times 600 \times 400 = 12 \times 10^{-2}\, H = 0.12$ mH.

Hence $R_3 = 240\ \Omega$ and $L_3 = 0.12$ mH

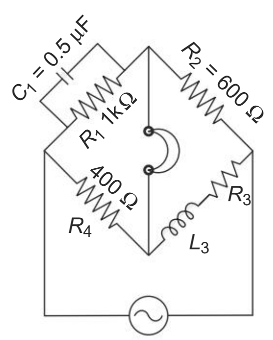


Fig. 11.31(a)

ANDERSON BRIDGE 11.18

The Anderson Bridge is a very important and useful modification of the Maxwell–Wien Bridge as shown in Fig 11.32.

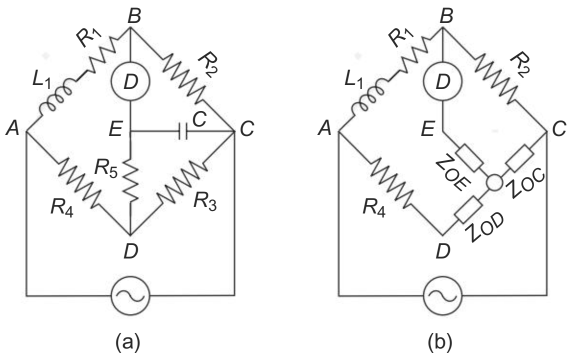


Fig 11.32 Anderson's bridge

The balance condition for this bridge can be easily obtained by converting the mesh impedances C, R_3, R_5 to an equivalent star with the star point 0 as shown in Fig 11.32(b) by using star/delta transformation.

As per delta to star transformation

$$Z_{OD} = \frac{R_3\, R_5}{(R_3 + R_5 + 1/j\omega C)} \qquad Z_{OC} = \frac{R_3 / j\omega C}{(R_3 + R_5 + 1/j\omega C)} = Z_3$$

Hence with reference to Fig. 11.32 (b) it can be seen that

$$Z_1 = (R_1 + j\omega L_1),\ Z_2 = R_2,\ Z_3 = Z_{OC} = \frac{R_3 / j\omega C}{(R_3 + R_5 + 1/ j\omega C)} \text{ and } Z_4 = R_4 + Z_{OD}$$

For balance condition,

$$Z_1 Z_3 = Z_2\, Z_4$$

Therefore, $(R_1 + j\omega L_1) \times Z_{OC} = Z_2 \times (Z_4 + Z_{OD})$

$$(R_1 + j\omega L_1) \times \left(\frac{R_3 /j\omega C}{(R_3 + R_5 + 1/j\omega C} \right) = R_2 \left(R_4 + \frac{R_3 R_5}{(R_3 + R_5 + 1 j\omega C} \right)$$

Simplifying,

$$(R_1 + j\omega L_1) \times \frac{R_3 /j\omega C}{(R_3 + \cancel{R_5 + 1}/j\omega C} = R_2 \left(R_4\ (R_3 + R_5 + 1/j\omega C) + \frac{R_3 R_5}{(R_3 + \cancel{R_5 + 1}/j\omega C} \right)$$

$$(R_1 + j\omega L_1) \times \frac{R_3}{j\omega C} = R_2\, R_4\, (R_3 + R_5 + 1/j\omega C) + R_2 R_3 R_5$$

$$\frac{R_1\, R_3}{j\omega C} + \frac{j\omega L_1 R_3}{j\omega C} = R_2 R_3 R_4 + R_2 R_4 R_5 + \frac{R_2 R_4}{j\omega C} + R_2 R_3 R_5$$

$$\frac{-j R_1 R_3}{\omega C} + \frac{L_1 R_3}{C} = R_2 R_3 R_4 + R_2 R_4 R_5 - \frac{j R_2 R_4}{\omega C} + R_2 R_3 R_5`$$

Equating the real terms and imaginary terms

$$\frac{L_1 R_3}{C} = R_2 R_3 R_4 + R_2 R_4 R_5 + R_2 R_3 R_5$$

$$L_1 = \frac{C}{R_3}\ (R_2 R_3 R_4 + R_2 R_4 R_5 + R_2 R_3 R_5)$$

$$L_1 = CR_2 \left[R_4 + \frac{R_4 R_5}{R_3} + R_5 \right]; \quad L_1 = CR_2 \left[R_4 + R_5 + \frac{R_4 R_5}{R_3} \right]$$

$$\frac{-j R_1 R_3}{\omega C} = \frac{-j R_2 R_4}{\omega C}; \quad R_1 R_3 = R_2 R_4, \text{ therefore, } R_1 = \frac{R_2 R_4}{R_3}$$

This method is capable of precise measurement of inductances and a wide range of values from a few µH to several Henries.

Example 11.14 *An inductive coil was tested by an Anderson bridge. The following were the values on balance.*
Arm AB unknown inductance having resistance R_1 and inductance L_1
Arm BC, CD, DA are resistors having 1000 Ω, 1000 Ω and 2000 Ω respectively
A capacitor of 10 µF and resistance 400 Ω are connected between CE and ED respectively, Source between A and C, r = 496. Determine L_1 and R_1.

Solution Given : $R_2 = 200\ \Omega$, $R_3 = 1000\ \Omega$, $R_4 = 1000\ \Omega$, $C = 10\ \mu\text{F}$, $r = 496$

Step 1: To calculate

$$R_1 = \frac{R_2 R_3}{R_4} = \frac{200 \times 1000}{1000} = 200\ \Omega$$

Step 2: Similarly to calculate

$$L_1 = \frac{CR_3}{R_4}\ (rR_4 + \mathrm{R}_2 R_4 + rR_2)$$

$$= \frac{10 \times 10^{-6} \times 1000}{1000} \times (496 \times 10^3 + 200 \times 10^3 + 496 \times 200)$$

$$= 10^{-5} \times 10^3 \times (496 + 200 + 0.496 \times 200)$$

$$= 10^{-2}\ (496 + 200 + 99.2)$$

$$= 795.2 \times 10^{-2}$$

$$= 7.952\ \text{H}$$

THE OWEN BRIDGE 11.19

In the arrangement shown in Fig 11.33, the unknown inductance is measured in terms of resistance and capacitance. This method has the advantage of being useful over a very wide range of inductances with capacitors of reasonable dimensions.

As per the balance condition,

$$Z_1 Z_3 = Z_2 Z_4$$

$$Z_1 = \frac{-j}{\omega C_1},\; Z_2 = R_2,\; Z_4 = R_4 - \frac{j}{\omega C_4},\; Z_3 = R_3 + j\omega L_3$$

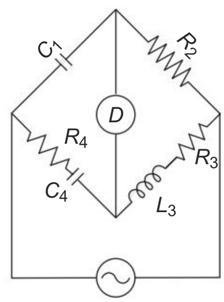


Fig 11.33 Owen's bridge

Hence substituting in the balance equation, we get

$$\left(\frac{-j}{\omega C_1}\right) \times (R_3 + j\omega L_3) = R_2 \times \left(R_4 - \frac{j}{\omega C_4}\right)$$

$$\frac{-j R_3}{\omega C_1} + \frac{-j \times j\omega L_3}{\omega C_1} = R_2 R_4 - \frac{j R_2}{\omega C_4}$$

$$\frac{-j R_3}{\omega C_1} + \frac{L_3}{C_1} = R_2 R_4 - \frac{j R_2}{\omega C_4}$$

Equating the real and the imaginary terms,

$$\frac{L_3}{C_1} = R_2 R_4 \text{ therefore, } L_3 = C_1 R_2 R_4$$

$$\frac{-j R_3}{\omega C_1} = \frac{-j R_2}{\omega C_4} \quad \text{Therefore, } \frac{R_3}{C_1} = \frac{R_2}{C_4} \quad \text{hence } R_3 = \frac{C_1 R_2}{C_4}$$

As seen from the values of R_3 and L_3 the term 'ω' does not appear in both the final equations. Hence the bridge is unaffected by frequency variations.

DE SAUTY BRIDGE 11.20

As seen from Fig. 11.34,

Let

C_2 be the capacitor whose capacitance is to be measured,

C_3 be a standard capacitor

R_1, R_2 be non-inductive resistors

Balance is obtained by varying either R_1 or R_2

For balance conditions, point B and D are at same potential.

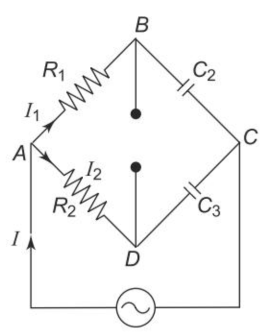


Fig 11.34 De Sauty bridge

Therefore, $V_B = V_D$

$$I_1 R_1 = I_2 R_2 \qquad (11.25)$$

$$\frac{-j}{\omega C_2} \times I_1 = \frac{-j}{\omega C_3} \times I_2 \qquad (11.26)$$

Dividing equation (11.25) and (11.26) we have

$$\frac{I_1 R_1}{\dfrac{-j}{\omega C_2} \times I_1} = \frac{I_2 R_2}{\dfrac{-j}{\omega C_3} \times I_2}$$

$$\frac{R_1 \times \omega C_2}{-j} = \frac{R_2 \times \omega C_3}{-j} \quad \text{therefore, } R_1 \times C_2 = R_2 \times C_3$$

Hence $R_1 C_2 = R_2 C_3$ therefore, $C_2 = \dfrac{R_2}{R_1} \cdot C_3$

The bridge has maximum sensitivity when $C_2 = C_3$.

If both the capacitors are not free from dielectric loss, to obtain perfect balance is difficult.

A perfect balance can only be obtained if air capacitors are used.

CAREY FOSTER / HEYDWEILLER BRIDGE 11.21

This bridge was basically designed and used by Carey Foster and was later on modified by Heydweiller for use in ac. Hence this bridge has both their names associated.

The two bridges are used for opposite purposes.

(i) If it is used for measurement of capacitance in terms of a standard mutual inductance, the bridge is known as Carey Foster's bridge.

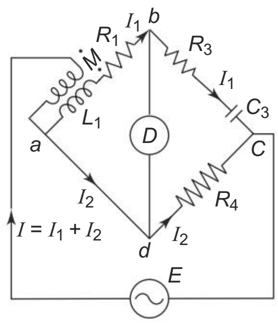


Fig 11.35 Carey Foster/Heydweiller bridge

(ii) It can also be used for measurement of mutual inductance in terms of a standard capacitance. Then the bridge is known as Heydweiller's bridge.

The bridge circuit is as shown in Fig 11.35. The bridge has a special feature, that one of its arms '*ad*' is short circuited and, therefore, the potential drop across this arm is zero.

In order to achieve balance,the potential drop across arm '*ab*' should also be zero. Hence for this reason negative coupling is required for the mutual inductance.

At balance

$$I_1 (R_1 + j\omega L_1) - (I_1 + I_2)\, j\omega M = 0 \tag{11.27}$$

and

$$I_1 \left(R_3 + \frac{1}{j\omega C_3} \right) = I_2 R_4 \tag{11.28}$$

From Eq. (11.27)

$$I_1 (R_1 + j\omega L_1) - I_1 j\omega M - I_2 j\omega M = 0$$

$$I_1 (R_1 + j\omega L_1 - j\omega M) = I_2 j\omega M$$

$$I_1 \left(R_3 + \frac{1}{j\omega C_3} \right) = I_2 R_4$$

Taking ratio of the above two equations, we have

$$\frac{I_1\,(R_1 + j\omega L_1 - j\omega M)}{I_1\left(R_3 + \dfrac{1}{j\omega C_3}\right)} = \frac{I_2\, j\omega M}{I_2 R_4}$$

$$\frac{(R_1 + j\omega L_1 - j\omega M)}{\left(R_3 + \dfrac{1}{j\omega C_3}\right)} = \frac{j\omega M}{R_4}$$

Cross-multiplying, we have

$$R_4\,(R_1 + j\omega L_1 - j\omega M) = j\omega M\left(R_3 + \frac{1}{j\omega C_3}\right)$$

$$R_4 R_1 + j\omega(L_1 - M)R_4 = j\omega M R_3 + \frac{j\omega M}{j\omega C_3}$$

$$R_4 R_1 + j\omega(L_1 - M)\,R_4 = j\omega M R_3 + \frac{M}{C_3}$$

Equating the real and imaginary terms, we have

$$R_4 R_1 = \frac{M}{C_3} \text{ therefore, } M = R_1 R_4 C_3 \tag{11.29}$$

$$j\omega\,(L_1 - M)\,R_4 = j\omega M R_3 \text{ therefore, } (L_1 - M)\,R_4 = M R_3$$

$$L_1\,R_4 = M R_3 + M R_4 \text{ therefore, } L_1\,R_4 = M\,(R_3 + R_4)$$

Hence, $$L_1 = \frac{M(R_3 + R_4)}{R_4} \tag{11.30}$$

But $M = R_1 R_4 C_3$. Hence, $L_1 = \dfrac{M(R_3 + R_4)}{R_4}$ $L_1 = \dfrac{R_1 R_4 C_3 (R_3 + R_4)}{R_4}$

Therefore, $L_1 = R_1 C_3\,(R_3 + R_4)$ (11.31)

If the bridge is used for measurement of capacitance equations (11.29) and (11.30) can be written as

$$C_3 = \frac{M}{R_1 R_4}$$

And $$R_3 = \frac{R_4\,(L_1 - M)}{M}$$

It can be seen that in the measurement of mutual inductance with this bridge, R_3 is a separate resistance while in the measurement of capacitance R_3 is not a separate unit but represents the equivalent series resistance of the capacitor and hence can be determined in terms of the bridge elements.

TYPES OF DETECTORS 11.22

1. For low frequency, the most convenient detector is the vibration galvanometer.
2. For ordinary laboratory work at frequencies up to a few 100 Hz, the moving coil type of instrument is usually employed. It has a high sensitivity.
3. In high voltage testing, the moving magnet type of vibration galvanometer with remote controlled tuning is used, (for 300 Hz – 1 kHz).
4. For higher AF frequencies (>800 Hz), the telephone (headphone) is the best detector.
   (Vibration galvanometers and headphones have no phase selectivity, i.e. they do not indicate whether it is resistance or reactance adjustments that are required.)
5. The ac galvanometer and separately excited dynamometer are phase selective, and are best suited at low frequencies. They have a high sensitivity.
6. In many cases, especially in bridges for routine use, a pointer instrument is used. It is advantageous if it can be made phase selective.
   These pointer instruments are generally moving coil milliammeters operated with some arrangement of copper oxide rectifiers (frequency range 40 Hz – 1 kHz).
7. Modern bridge techniques employ the amplifier as a regular feature.
8. At frequencies above about 3 kHz, and particularly at high AF or RF, a heterodyne or beat-tone detector is used.
9. With all detectors, the impedance should be selected to suit that of the bridge. A higher sensitivity can be obtained by using an interbridge transformer. Also, when headphones are used as detectors, precautions should be taken to eliminate capacitance effects between the observer and the headphones.
10. A moving magnet vibration galvanometer has a range of up to 1500 Hz.
11. An electrodynamometer can also be used as an ac detector.
12. Electrometers are used as detector because small capacitances possess a very large impedance when used with ac circuits at low frequency, and when measured in the bridge they form a high impedance branch. Hence, this detector is used to increase sensitivity.

PRECAUTIONS TO BE TAKEN WHEN USING A BRIDGE 11.23

Assuming that a suitable method of measurement has been selected and that the source and detector are given, there are some precautions which must be observed to obtain accurate readings.

The leads should be carefully laid out in such a way that no loops or long lengths enclosing magnetic flux are produced, with consequent stray inductance errors.

With a large L, the self-capacitance of the leads is more important than their inductance, so they should be spaced relatively far apart.

In measuring a capacitor, it is important to keep the lead capacitance as low as possible. For this reason the leads should not be too close together and should be made of fine wire.

In very precise inductive and capacitances measurements, leads are encased in metal tubes to shield them from mutual electromagnetic action, and are used or designed to completely shield the bridge.

Review Questions

1. What is a bridge.? What is the importance of a bridge?
2. Explain with diagram the operation of a Wheatstone bridge.
3. What is the criteria for balance of a Wheatstone bridge?
4. State the limitations of a Wheatstone bridge. How is it overcome?
5. Explain with a diagram the working of a unbalanced Wheatstone bridge.
6. Explain with a diagram the working principle of Kelvin's bridge.
7. Describe with diagram the operation of Kelvin's bridge.
8. How does the basic circuit of Kelvin's bridge differ from that of a Wheatstone bridge.
9. Derive the balance condition for a basic Kelvin's bridge.
10. Draw and explain a practical Kelvin's double bridge.
11. List and discuss the principle applications of a Kelvin's bridge.
12. Explain with a diagram how a bridge be used as an error detector.
13. Explain with a block diagram the working of a digital readout bridge.
14. How does the use of microprocessors are useful in bridge circuits?
15. Explain the operation of a microprocessor controlled bridge.
16. Describe how a Wheatstone bridge may be used to control various physical parameters.
17. Define the term 'null' as applied to bridge measurement.
18. What are some methods by which microprocessors are reducing the cost and complexity of analog measurements?
19. Explain with diagram a basic ac bridge.
20. Compare dc and ac bridges.
21. List various detectors used for ac measurements.
22. State the two conditions that must be satisfied to obtain bridge balance.
23. Describe how a similar angle bridge (comparison bridge) differs from a Wheatstone bridge.
24. Draw and derive the balance condition for a capacitance comparison bridge.
25. Draw and derive the balance condition for an inductance comparison bridge.
26. Explain with a diagram how Maxwell's bridge can be used to measure unknown inductance.
27. Draw the circuit diagram and obtain balance conditions for Maxwell's bridge. State the limitation of a Maxwell's bridge.
28. Draw the circuit diagram and obtain balance conditions for Hay's bridge.
29. Compare Maxwell's bridge and Hay's bridge.
30. Explain with a diagram how Schering's bridge can be used to measure unknown capacitance.
31. Draw the circuit diagram and obtain balance conditions for Schering's bridge.

32. Explain how dissipation factor of a capacitor can be measured.
33. Explain Wien's bridge with a diagram.
34. State and derive the two balance conditions for a Wien bridge.
35. How can a Wien bridge be used to measure frequency?
36. Explain with a diagram the working of a Maxwell Wien bridge.
37. Draw the circuit and derive the condition of balance for a Maxwell–Wien bridge.
38. Explain with a diagram the working of an Anderson bridge.
39. Draw the circuit and obtain the balance condition of an Anderson bridge.
40. Compare Maxwell–Wien bridge with that of an Anderson bridge.
41. Explain with a diagram the working of Owen's bridge.
42. Draw the circuit and obtain the balance condition of an Owen's bridge.
43. Compare Anderson bridge and Owen's bridge.
44. Explain with a diagram the operation of a De Sauty bridge.
45. Draw the circuit and obtain the balance condition of a De Sauty bridge.
46. Compare Schering bridge with De Sauty bridge.
47. What do you mean by Wagner's ground connection? What is its significance?
48. Explain with a diagram the working of a Wagner's ground connection.
49. Explain how stray capacitances can be eliminated using Wagner's ground connection.
50. Explain Resonance Bridge with a diagram.
51. Derive the balance conditions for a resonance bridge.

Multiple Choice Questions

1. A basic bridge consists of
   (a) two arms (b) three arms
   (c) four arms (d) single arm
2. Wheatsone bridge is used to measure
   (a) voltage (b) current
   (c) power (d) resistance
3. Kelvin's bridge is used to measure
   (a) voltage (b) current
   (c) power (d) resistance
4. An ac bridge uses a detector in the form of
   (a) ammeter (b) voltmeter
   (c) headphones (d) wattmeter
5. Maxwell's bridge is used to measure unknown
   (a) inductance (b) capacitance
   (c) resistance (d) Q
6. Maxwell's bridge is used to measure Q factor in the range
   (a) 1–10 (b) 30–50
   (c) 50–75 (d) 75–100
7. Hay's bridge is used to measure an inductance of
   (a) low Q (b) medium Q
   (c) high Q (d) very high Q
8. Schering bridge is used to measure unknown
   (a) inductance (b) capacitance
   (c) resistance (d) frequency
9. Schering bridge is also used to measure
   (a) Q factor
   (b) dissipation factor
   (c) resistance
   (d) frequency
10. Wien bridge in its basic form is used to measure unknown
    (a) inductance (b) capacitance
    (c) resistance (d) frequency
11. Anderson bridge is used to measure unknown
    (a) inductance (b) capacitance
    (c) resistance (d) frequency

12. To measure precise inductance from a few μH to several henries, the following bridge is used.
(a) Maxwell's (b) Hay
(c) Maxwell–Wien (d) Anderson

13. Wagner's ground is used to
(a) eliminate stray capacitances
(b) measure capacitance
(c) measure resistance
(d) measure inductance

Practice Problems

1. Calculate the value of R_x in a Wheatstone bridge if
(i) $R_1 = 400\ \Omega$, $R_2 = 5$ k, $R_3 = 2$ k
(ii) $R_1 = 10$ k, $R_2 = 40$ k, $R_3 = 15.5$ k
(iii) $R_1 = 5$ k, $R_2 = 40$ k, $R_3 = 10\ \Omega$
2. What resistance range must resistor R_3 have in order to measure unknown resistor in the range 1 – 100 kΩ using a Wheatstone bridge? Given $R_1 = 1$ k and $R_2 = 10$ k.
3. Calculate the value of R_x in Fig. Ex. 11.12, $R_a = 1600\ R_b$, $R_1 = 800\ R_b$ and $R_1 = 1.25\ R_2$.
4. Calculate the current through the galvanometer in the circuit diagram of Fig. 11.36.

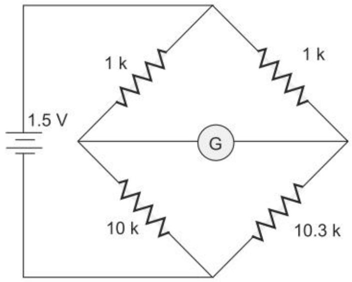


Fig. 11.36

5. If the sensitivity of the galvanometer in the circuit of Fig. 11.37 is 10 mm/μA, determine its deflection.

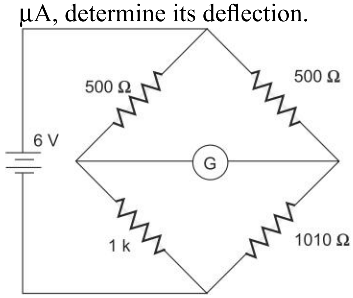


Fig. 11.37

6. A balanced ac bridge has the following constants.
Arm AB – R = 1 k in parallel with C = 0.047 μF
Arm BC – R = 2 k in series with C = 0.047 μF
Arm DC – unknown
Arm DA – C = 0.25 μF.
The frequency of the oscillator is 1 kHz. Determine the constants of arm CD.
7. A bridge is balanced at a frequency of 1 kHz and has the following constants.
Arm AB – 0.2 μF pure capacitor
Arm BC – 500 Ω pure resistance
Arm CD – unknown
Arm DA – R = 600 Ω in parallel with C = 0.1 μF.
Derive the balance condition and find the constants of arm CD, considered as a series circuit.
8. A 1000 Hz bridge has the following constants
Arm AB – R = 1 k in parallel with C = 0.25 μF
Arm BC – R = 1 k in series with C = 0.25 μF
Arm CD – L = 50 mH in series with R = 200 Ω
Arm DA – unknown
Find the constants of arm DA to balance the bridge.
Express the result as a pure R in series with a pure C or L, and as a pure R in parallel with a pure C or L.
9. An ac bridge has the following constants.
Arm AB – a pure capacitor C = 0.2 μF

Arm BC – a pure resistance R = 500 Ω

Arm CD a series combination of R = 50 Ω and L = 0.1 H

Arm DA a capacitor C = 0.5 μF in series with a resistance R_s

If ω = 2000 rad/s

(i) Find the value of R_s to obtain bridge balance.

(ii) Can complete balance be obtained by the adjustment of R_s?

If not, specify the position and value of an adjustable resistance to complete the balance.

10. A Maxwell–Wien bridge consists of the following:

Arm AB having resistance value of 1.2 kΩ in parallel with a capacitor of 1 μF

Arm BC having resistance value of 500 Ω

Arm AD having resistance value of 300 Ω

Arm BD having resistance and inductance in series.

Determine the value of the unknown resistance and unknown inductance.

11. An Anderson's bridge consists of the following:

Arm AD having resistance value of 500 Ω

Arm CD having a resistance of 1000 Ω

Arm ED having a resistance of 600 Ω

Arm EC having a capacitor of 0.5 μF

Arm BC having resistance value of 300 Ω

Arm AB having resistance and inductance in series.

Determine the value of the unknown resistance and unknown inductance.

12. A Owen's bridge consists of the following:

Arm AB having a capacitor of 0.5 μF

Arm BC having resistance value of 600 Ω

Arm AD having resistance value of 300 Ω in series with a capacitor 0.75 μF

Arm BD having resistance and inductance in series

Determine the value of the unknown resistance and unknown inductance:

13. A De Sauty bridge consists of the following:

Arm AB having a resistance of 1 kΩ,

Arm BC having a capacitor value of 0.75 μF

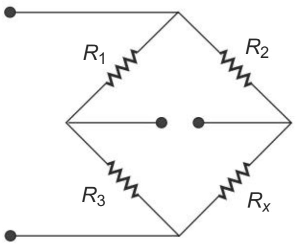
Arm AD having resistance value of 300 Ω

Arm BD having unknown capacitor

Determine the value of the unknown capacitor.

Bridge Arrangements (Summary)

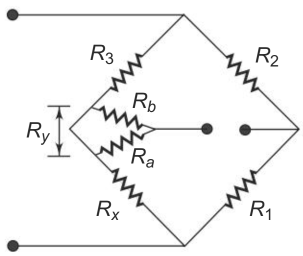
A. Resistance Measurement



Wheatstone's Bridge

$$R_x = \frac{R_2 R_3}{R_1}$$

Universally used circuit for resistance measurement. It can be used to make measurements from 1 Ω to 1 MΩ with an accuracy of ± 0.25%.



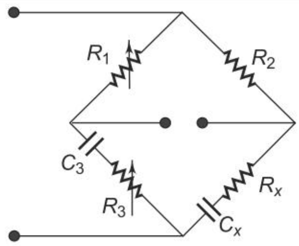
Kelvin's Double Bridge

$$R_x = \frac{R_1 R_3}{R_2}$$

Used for measuring small resistances as low as 0.001 Ω with an accuracy of ± 2%.

Fig. 11.31 (a)

B. Capacitance Measurement

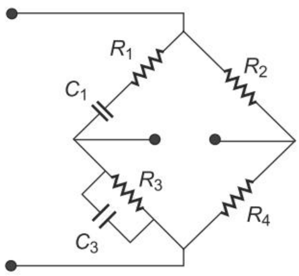


Capacitance Comparison

$$R_x = \frac{R_2 R_3}{R_1}$$

$$C_x = \frac{C_3 R_1}{R_2}$$

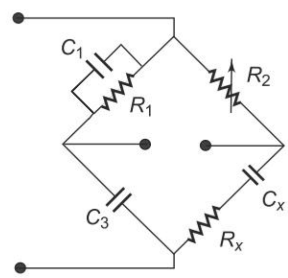
Used to measure unknown capacitances with reference to a standard capacitor. A fixed resistance ratio and variable standards are used. Balance is obtained by alternately varying C_3 and R_1.



Wien's Bridge

$$C_3 = \left(\frac{R_2}{R_4} - \frac{R_1}{R_3}\right)C_1$$

Used to measure capacitance to a high degree of accuracy when frequency and resistance standards are employed.



Schering's Bridge

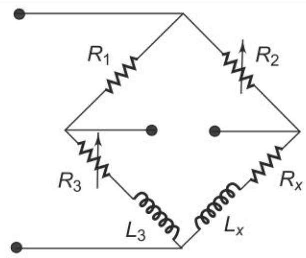
$$R_x = \frac{R_2 C_1}{C_3}$$

$$C_x = \frac{R_1 C_3}{R_2}$$

Most widely used bridge for capacitance measurements. Used for measurements of capacitance in the range of 100 pf – 1 µF, with an accuracy of ± 0.2%.

Fig. 11.31 (b)

C. Inductance Measurements

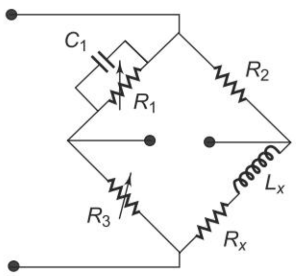


Inductance Comparison Bridge

$$R_x = \frac{R_2 R_3}{R_1}$$

$$L_x = \frac{L_3 R_2}{R_1}$$

Used to measure unknown inductances with reference to a standard inductor. A fixed resistance ratio and variable standards are used. Balance is obtained by alternately varying L_3 and R_3.

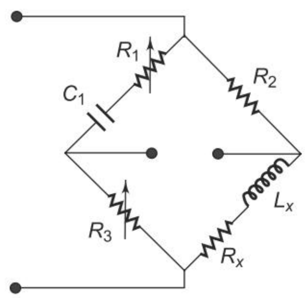


Maxwell's Bridge

$$R_x = \frac{R_2 R_3}{R_1}$$

$$L_x = C_1 R_2 R_3$$
$$Q = \omega C_1 R_1$$

Most widely used for Inductance measurements of low Q. Best suited for inductance measurement, since comparison with a capacitor is more ideal than with an inductor. It can measure from 1 µH to 1000 H with a ± 2% error (for $Q < 10$).



Hay's Bridge

$$R_x = \frac{R_2 R_3}{R_1 [1 + Q^2]}$$

$$L_x = \frac{R_2 R_3 C_1}{[1 + 1/Q^2]}$$

$$Q = \frac{1}{\omega R_1 C_1}$$

Mostly used for measurement of high Q (>10) coils. Not used for low Q, since the measurement depends on Q. It can measure from 1 µH to 100 H with a ± 2% error.

Fig. 11.31 (c)

Further Reading

1. B. Haque, *A.C. Bridges Methods*, Sir Issac Pitman & Sons, 1942.
2. Philco Technological Centre, *Electronic Precision Measurement Techniques and Experiment*.
3. *Handbook of Electronic Measurements*, Vols. I & II, Polytechnic Institute of Brooklyn; 1956. (Microwave Research Institute)
4. Larry D. Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley & Sons, 1987.

Chapter

Recorders 12

INTRODUCTION 12.1

A recorder is a measuring instrument that displays a time-varying signal in a form easy to examine, even after the original signal has ceased to exist.

Recorders generally provide a graphic recod of variations in the quantity being measured, as well as an easily visible scale on which the indication is displayed.

The variety of recording instruments in the central monitoring and control stations of many industrial and utility plants is proof of their importance in industrial work. They provide a continuous, written record of the changes taking place in the quantity being measured. This chart record may be scaled off in electrical values (mV/mA) or in terms of some non-electrical quantity, such as temperature or pressure.

Many recording instruments include an additional provision for some sort of controlling action. If the control function is the primary one, the measuring instrument is called a controller.

The recorder usually provides an instantaneous indication for monitoring at the same time as it makes a graphic record.

Electronic recording instruments may be divided into three groups.

The easiest type is simply a meter having an indicating needle and a writing pen attached to the needle. If a strip of paper is pulled at a constant velocity under the writing pen (at a 90° angle to the direction of pen motion), the moving pen plots the time function of the signal applied to the meter. A highly special designed D'Arsonval movement is used to drive the writing pen. This type is called a *galvanometer recorder*.

Another recorder is the null or *potentiometric recorder*, operating on a self-balancing comparison basis by servomotor action. This recorder is basically a voltage responsive positional servo system using a motor to move a writing device back and forth across a piece of paper. The servo system can be made extremely accurate, rugged and powerful.

The *magnetic recorder* is the third type. In it, a thin magnetic tape or wire, is magnetised in accordance with a varying signal as the tape passes rapidly across a magnetic recording head. The frequency response of magnetic recorders can extend from 0 Hz to a few kHz to nearly 10 MHz. Because of the wide bandwidth of tape recorders, several modes of recording (direct, FM and digital) can be used.

There are two types of recording devices, (i) Analog, and (ii) Digital. Analog recorders may be (i) graphic or (ii) magnetic. Graphic recorders are devices which display and store a pen and ink record of some physical quantity. They are of three types (i) strip chart recorder, (ii) circular chart recorder, and an (iii) X-Y recorder. These and other types of recorders are discussed in detail in the following sections.

STRIP CHART RECORDER 12.2

Strip chart recorders are those in which data is recorded on a continuous roll of chart paper moving at a constant speed. The recorder records the variation of one or more variables with respect to time. The basic element of a strip chart recorder consists of a pen (stylus) used for making marks on a movable paper, a pen (stylus) driving system, a vertically moving long roll of chart paper and chart paper drive mechanism and a chart speed selector switch, (as shown in Fig. 12.1(a)).

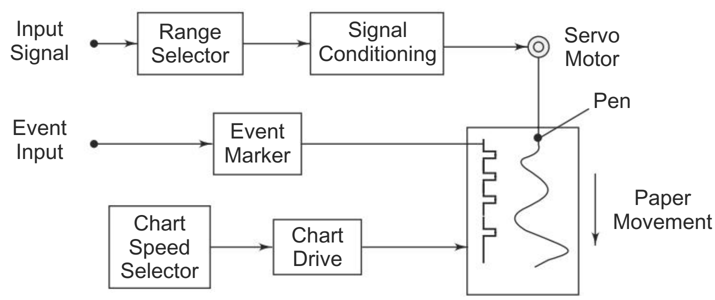


Fig. 12.1 (a) Basic strip chart recorder

Most recorders use a pointer attached to the stylus, so that the instantaneous value of the quantity being recorded can be measured directly on a calibrated scale. The assembly of a strip chart recorder is shown in Fig. 12.1(b). This recorder uses a single pen and is servo driven.

Most strip chart recorders use a servo feedback system, to ensure that the displacement of the pen (stylus) across the paper tracks the input voltage in the required frequency range.

A potentiometer system is generally used to measure the position of the writing head (stylus).

The chart paper drive system generally consists of a stepping motor which controls the movement of the chart paper at a uniform rate.

The data on the strip chart paper can be recorded by various methods.

1. Pen and Ink Stylus The ink is supplied to the stylus from a refillable reservoir by capillary action. Modern technology has replaced these pens by disposable fibre tip pens. In addition, multichannel operation can be performed, i.e. at any instant, a maximum of six pens can be used to record data. When using multiple pens, staggering of the pens are necessary to avoid mechanical interference.

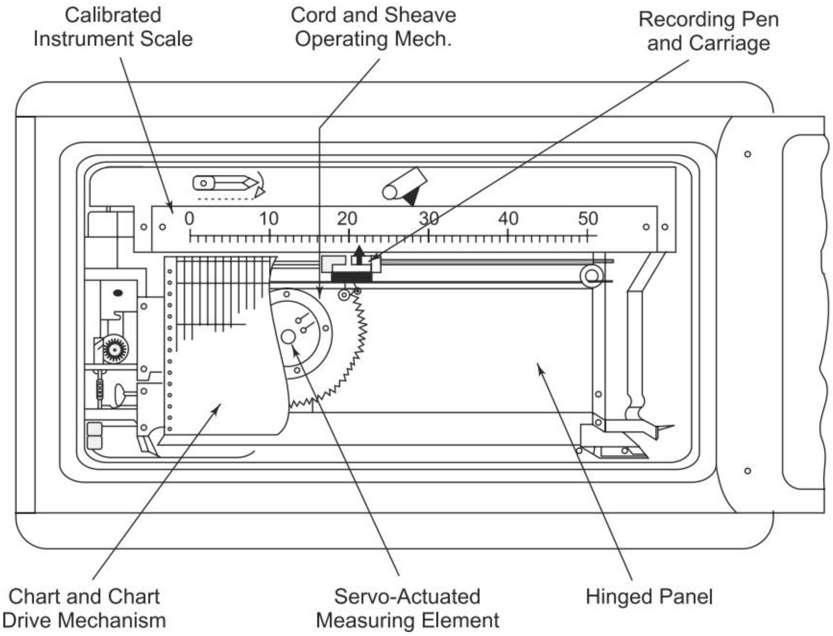


Fig. 12.1 (b) Assembly of a single pen servo operated strip chart recorder

2. Impact Printing The original impact system consisted of a carbon ribbon placed between the pointer mechanism and paper, which provided the ink for recording data. The mark was made on the paper by pressing the pointer mechanism on it. The advantage of impact printing over the pen and ink method is that, it can record data on up to 20 variables simultaneously. This is achieved with the help of a wheel with an associated ink pad which provides the ink for the symbol on the wheel. The wheel is moved across the paper in response to the variable being recorded.

In some mechanisms, pressure sensitive paper is used. The markings on the paper are done with chopper bar, which applies the pressure on the paper. The frequency of the chopper bar is once per second.

3. Thermal Writing In this system, a special movable pen which is thermally heated by passing an electric current through it is used. This system requires a thermally sensitive paper which changes its colour on application of heat.

4. Electric Writing This technique is based on the principle of electrostatics.

In this method, a special chart paper is used. This paper consists of a paper base coated with a layer of coloured dye (black, blue or red), which in turn is coated with a thin surface of aluminium.

The stylus (pen) consists of a tungsten wire moving over the aluminium surface. Markings on the paper are achieved by applying a potential of 35 V to the stylus. This causes an electric discharge which removes the aluminium, revealing the coloured dye.

5. Optical Writing In this technique of writing, a special photo sensitive chart paper, sensitive to ultra violet light is used. This technique is mostly used in galvanometer system.

Ultra violet light is used to reduce unwanted effects from ambient light. The paper can be developed in daylight or under artificial light without the need for special chemicals, which is not possible if ordinary light is used.

Most recorders use a pointer attached to the stylus. This pointer moves over a calibrated scale giving the instantaneous value of the quantity being recorded.

(a) *Paper drive system* The paper drive system should move the paper at a uniform speed. A spring wound mechanism may be used in most recorders. A synchronous motor is used for driving the paper.

(b) *Chart speed* Chart speed is a term used to express the rate at which the recording paper in a strip chart recorder moves. It is expressed in in/s or mm/s and is determined by mechanical gear trains. If the chart speed is known, the period of the recorded signal can be calculated as

$$\text{Period} = \frac{\text{time}}{\text{cycle}} = \frac{\text{time base}}{\text{chart speed}}$$

and frequency can be determined as $f = 1/\text{period}$.

Example 12.1 *The chart speed of a recording instrument is 40 mm/s. One cycle of the signal is recorded over 5 mm (this is referred to sometimes as the time base). Determine the frequency of the signal.*

Solution

$$\text{Period} = \frac{\text{time}}{\text{cycle}} = \frac{\text{time base}}{\text{chart speed}} = \frac{5\ \text{mm/cycle}}{40\ \text{mm/s}} = \frac{5\ \text{mm}}{\text{cycle}} \times \frac{\text{s}}{40\ \text{mm}}$$

Therefore, period = 5/40 s/cycle = 1/8 s/cycle = 0.125 s/cycle and frequency $f = 1/\text{period} = 1/0.125$ s/cycle = 8 cycles/s.

Example 12.2 *If the frequency of a signal to be recorded with a strip-chart recorder is 20 Hz, what must be the chart speed used to record one complete cycle on 5 mm of recording paper?*

Solution Given frequency = 20 Hz and time base = 5 mm

$$\text{Period} = 1/\text{frequency} = 1/20 = 0.05\ \text{s}$$

$$\text{Period} = \frac{\text{time base}}{\text{chart speed}}, \quad \text{therefore} \quad 0.05 = \frac{5\ \text{mm/cycle}}{\text{chart speed}}$$

$$\text{Chart speed} = \frac{5\ \text{mm}}{\text{cycle}} \times \frac{1}{0.05\ \text{s/cycle}}$$

$$= \frac{5 \times 100}{5}\ \text{mm/s} = 100\ \text{mm/s}$$

There are basically two types of strip chart recorders, the (i) galvanometer type, and the (ii) null type (potentiometric).

GALVANOMETER TYPE RECORDER 12.3

The D'Arsonval movement used in moving coil indicating instruments can also provide the movement in a galvanometer recorder.

The D'Arsonval movement consists of a moving coil placed in a strong magnetic field, as shown in Fig. 12.2(a).

In a galvanometer type recorder, the pointer of the D'Arsonval movement is fitted with a pen-ink (stylus) mechanism.

The pointer deflects when current flows through the moving coil. The deflection of the pointer is directly proportional to the magnitude of the current flowing through the coil.

As the signal current flows through the coil, the magnetic field of the coil varies in intensity in accordance with the signal. The reaction of this field with the field of the permanent magnet causes the coil to change its angular position. As the position of the coil follows the variation of the signal current being recorded, the pen is accordingly deflected across the paper chart.

The paper is pulled from a supply roll by a motor driven transport mechanism. Thus, as the paper moves past the pen and as the pen is deflected, the signal waveform is traced on the paper.

The recording pen is connected to an ink reservoir through a narrow bore tube. Gravity and capillary action establish a flow of ink from the reservoir through the tubing and into the hollow of the pen.

Galvanometer type recorders are well suited for low frequency ac inputs obtained from quantities varying slowly at frequencies of upto 100 c/s, or in special cases up to 1000 c/s.

Because of the compact nature of the galvanometer unit (or pen motor) this type of recorder is particularly suitable for multiple channel operation. Hence it finds extensive use in the simultaneous recording of a large number of varying transducers outputs.

This recorder uses a curvilinear system of tracing. The time lines on the chart must be arcs of radius R (where R is the length of the pointer), and the galvanometer shaft must be located exactly at the center of curvature of a time line arc. Improper positioning of the galvanometer or misalignment of the chart paper in the recorder can give a distorted response, i.e. having a negative rise time or a long rise time. One method of avoiding the distorted appearance of recordings in curvilinear coordinates is to produce the recording in rectangular coordinates. In this design, the chart paper is pulled over a sharp edge that defines the locus of the point of contact between the paper and the recording stylus. The stylus is rigidly attached to the galvanometer coil and wipes over the sharp edge as the coil rotates.

In one of the recorders, the paper used is usually heat sensitive, and the stylus is equipped with a heated tip long enough to guarantee a hot point of contact with the paper, regardless of the stylus position on the chart. Alternatively the paper can be electrically sensitive, in which case the stylus tip would serve to carry current into the paper at the point of contact.

The recorders can work on ranges ranging from a few mA/mV to several mA/mV. These moving galvanometer type recorders are comparatively inexpensive instruments, having a narrow bandwidth of 0 – 10 Hz. They have a sensitivity of about 0.4 V/mm, or from a chart of 100 mm width a full scale deflection of 40 mV is obtained.

In most instruments, the speed of the paper through the recorder is determined by the gear ratio of the driving mechanism. If it is desired to change the speed of the paper, one or more gears must be changed.

Paper speed is an important consideration for several reasons.

1. If the paper moves too slowly, the recorded signal variations are bunched up and difficult to read.
2. If the paper moves too fast, the recorded waveform will be so spread out that greater lengths of paper will be required to record the variations of the signal. It also makes the task of reading and interpreting the waveforms more difficult.
3. Also, the operator can determine the frequency components of the recorded waveform, if he knows how fast the paper has moved past the pen position. The paper is usually printed with coordinates, such as graph paper.

Some recorders contain a timing mechanism that prints a series of small dots along the edge of the paper chart, as the paper moves through the recorder. This time marker produces one mark per second.

These types of recorders are mostly used as optical recorders, and contain a light source provided by either an ultra violet or tungsten lamp.

A small mirror is connected to the galvanometer movement and the light beam is focussed on this mirror, as shown in Fig. 12.2(b).

The beam reflected from the mirror is focussed into a spot on a light sensitive paper.

As the current passes through the coil, the mirror deflects. The movement of the light beam is affected by the deflection of the small mirror, and the spot on the paper also varies for the same reason, thus tracing the waveform on the paper.

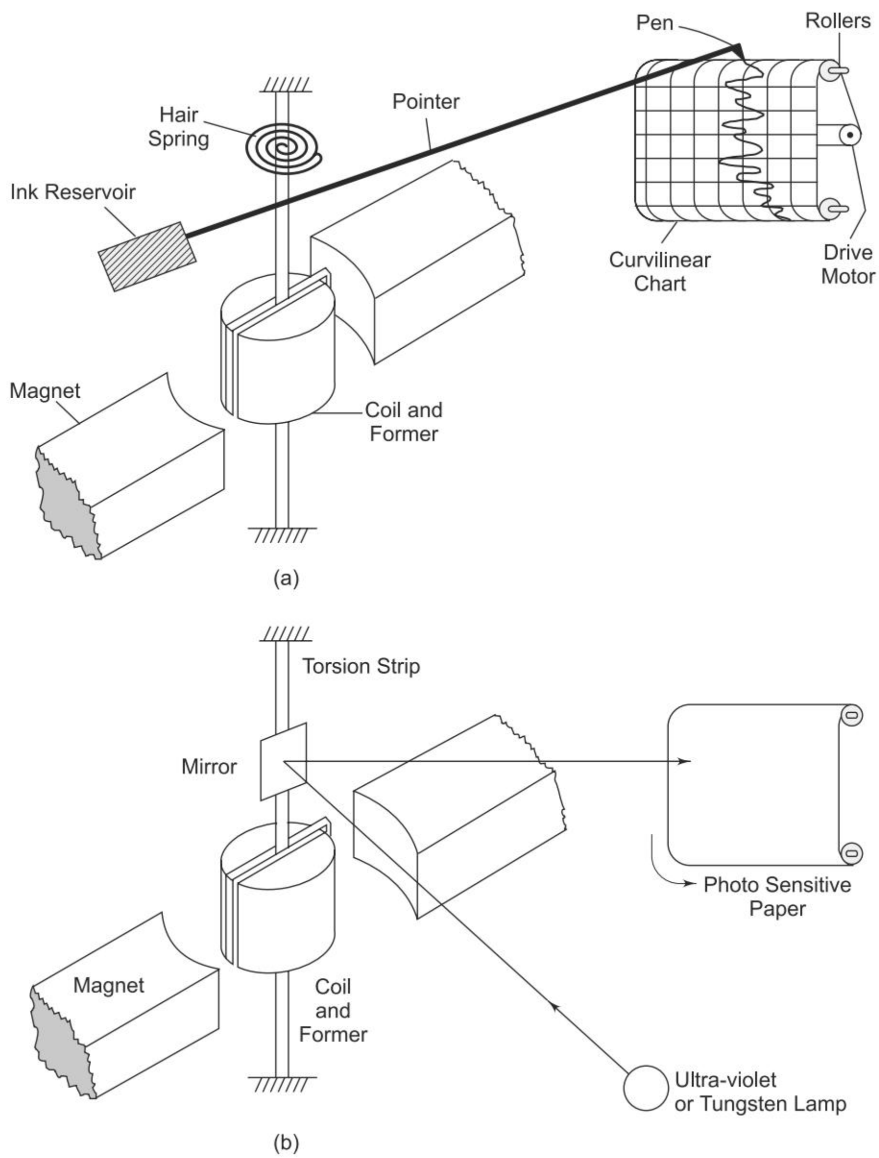


Fig. 12.2 (a) Galvanometer type recorder (b) Optical galvanometer recorder

NULL TYPE RECORDER (POTENTIOMETRIC RECORDERS) 12.4

These recorders work on the principle of self-balancing or null conditions.

When an input is given to the measuring circuit of the recorder from a sensor or transducer, it upsets the balance of the measuring circuit, producing an error voltage which operates some other device, which in turn restores the balance or brings the system to null conditions.

The magnitude of the error signal indicates the amount of movement of this balance restoring device and the direction of the movement indicates the direction of the quantity being measured. The different types of null recorders are as follows:

1. Potentiometric recorders
2. Bridge recorders
3. LVDT recorders (Linear Variable Differential Transformer)

12.4.1 Potentiometric Recorders

The basic disadvantage of a galvanometer type recorder is that is has a low input impedance and a limited sensitivity.

This disadvantage can be overcome by using an amplifier between the input terminals and the display or indicating instruments. This amplifier provides a high input impedance and improved sensitivity at the cost of low accuracy.

To improve the accuracy of the instrument, the input signal is compared with a reference voltage using a poteniometer circuit.

The self-balancing feature is obtained with a servo motor, a motor whose speed and direction of rotation follows the output of an amplifier. In a dc system, this is simply a reversible motor, such as the type that uses a permanent magnet for its field. In the ac system, it takes the form of a two-phase motor.

Figure 12.3 is a basic circuit of a potentiometric or self-balancing recorder.

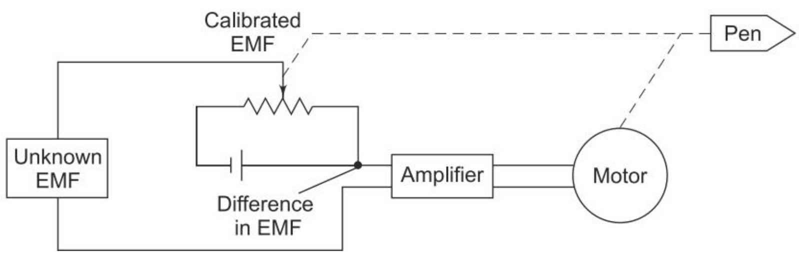


Fig. 12.3 Basic circuit of a self-balancing or potentiometric recorder

The difference between the input signal and the potentiometer voltage is the error signal. This error signal is amplified and is used to energize the field coil of a dc motor. In this circuit, instead of obtaining a balance between two opposing voltages by rotating the arm of the voltage divider, an error current is allowed to flow, either clockwise or counter clockwise, depending on which voltage is higher. This error serves as the input to the electronic detector, and the amplified error is then fed to the balancing motor. This motor is so connected that it turns in a direction that rotates the voltage divider arm (geared) to it in the direction that reduces the error. As the error becomes smaller, the motor slows down and finally stops at the point where the error is zero, thus producing the null balance.

This is achieved by mechanically connecting the wiper/variable arm to the armature of the dc motor. The pen is also mechanically connected to the wiper. Hence as the wiper moves in a particular direction, the pen also moves in synchronism in the same direction, thereby recording the input waveform. The wiper comes to rest when the unknown signal voltage is balanced against the voltage of the potentiometer. This technique results in graphical recorders having a very high input impedance.

A sensitivity of 4 V/mm is attained with an error of less than ± 0.25% with a bandwidth of 0.8 Hz.

A motor synchronised to power line frequency is used to drive the chart drive for most potentiometer recorders.

Hence the speeds of the chart drive can be changed by the use of a gear train which uses different gear ratios.

Potentiometer recorders are mostly used for the recording and control of process temperature.

Figure 12.4 is the basic block diagram of a dc self balancing system. Instrument that record changes of only one measured variable are called single point recorders.

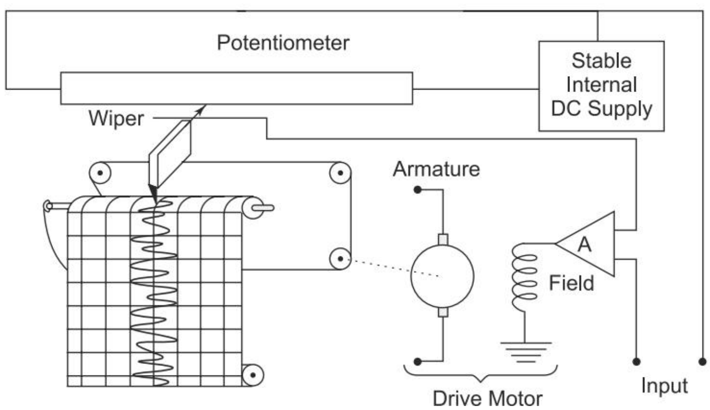


Fig. 12.4 Block diagram of self-balancing potentiometer recorder

Multipoint recorders are those in which one recorder may be used for recording several inputs. These may have as many as 24 inputs, with traces displayed in six colours. The data is recorded on an 8 inch chart, at frequencies from dc to 5 kHz, models up to 36 channels are also available.

12.4.2 Bridge Type Recorders

When a non-electrical quantity such as temperature is to be recorded, the transducer converts the temperature changes in to corresponding electrical variations.

If a thermistor or resistance thermometer were used as the transducer, the changes in temperature would produce variations in the resistance of the tranducer, rather than a change in voltage. In this case, the thermistor is made part of the bridge circuit, as shown in Fig. 12.5.

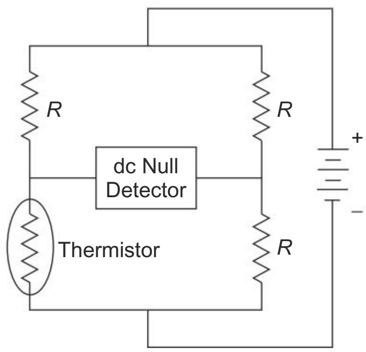


Fig. 12.5 Thermistor arranged in a dc bridge circuit

The resistance changes in the thermistor cause corresponding changes in the bridge output. These changes are applied to the detector. The bridge balance (null) can be restored by varying the resistance of another arm of the bridge, while recording in terms of current, voltage or temperature. Depending on the kind of voltage supplied to the bridge, the output can be chosen to be dc or ac.

12.4.3 Linear Servo Motor Recorder (LVDT)

Some temperature indicating and recording instruments incorporate a linear servo motor which dispenses with the conventional servo motor and error-prone gear train linkage and motor brushes. Improvement in reliability, speed of response, minimum resistance to movement and high accuracy are obtained with the use of linear servo motors.

The requirements of dc linear motor are as follows

1. It should produce motion in straight line in response to a direct current.
2. It should reverse its motion if the current polarity is reversed.
3. It should be easy to control and have a low inertia.
4. Its use and power requirements must be comparable with existing drive systems.

If two permanent magnets, parallel to one another, are fixed to a bench, a small free permanent magnet placed between them will move linearly in a direction which depend on the position of the north and south poles of the fixed magnets.

If a current is passed through a small coil, in place of the permanent magnet, the coil will move in a similar fashion but with the advantage that the direction of motion can be reversed by reversing the current flow.

Figure 12.6 indicates the simple nature of motor action.

Current flow through the coil assembly generates the motor action. The force *F* should be sufficient to overcome the friction in the moving parts supporting the coil, the inertia of the coil assembly, plus a suitable safety factor.

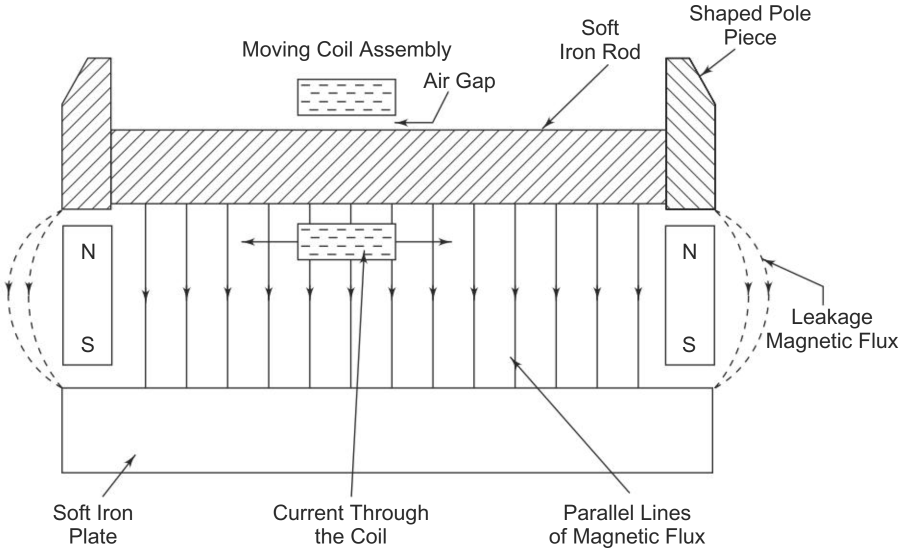


Fig. 12.6 LVDT (Basic principle)

The coil in the linear motor is supported by two small pulleys at the rear and the front with a spring affixed to the coil and resting on a slide wire. The pulleys are adjusted to maintain a clearance air gap around the rod. The scale pointer is part of the coil holder. This arrangement provides a direct drive for the scale pointer and slide wire contacts, enabling motion in straight line, which

(i) permits the use of a straight former for the slide wire, and
(ii) enables a flat scale to be used.

Ferrodynamic (LVDT) recorders are versatile instruments used for measuring and recording process variables such as pressure, vacuum, flow, level, and angle of rotation. These instruments work on the null balance principle and are used in conjunction with suitable primary transmitters (transducers), as different transformers or induction transducers. The accuracy of these instruments is of the order of ± 0.6% of full scale division. The threshold sensitivity is about ± 0.25% of span and the response time is about 6 seconds.

Principle of Operation

The instrument comprises the following main units.

1. Ferrodynamic compensating converter
2. Solid state amplifier
3. Balancing motor
4. Linkage mechanism
5. Chart drive mechanism (for recorders)

The ferrodynamic converter consists of a sturdy moving coil placed in an alternating magnet field. The magnitude and phase of the induced EMF in the moving coil depends upon the angle of rotation of the coil. A bias winding is introduced over the exciting winding, which can bodily shift the output characteristics of the ferrodynamic converter (LDVT), as shown in Fig. 12.7.

From Fig. 12.7, it can be seen that for a – 20° to + 20° rotation of the moving coil, the voltage induced varies by 2 V. By having a bias winding, it is possible to obtain signals of range – 1 V to + 1 V, 0 V to + 2 V, and + 1 V to + 3 V.

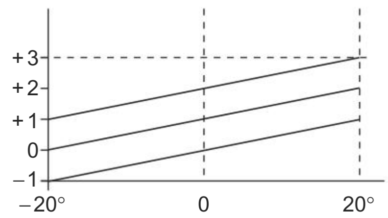


Fig. 12.7 Output characteristics of an LVDT (Ferro-dynamic converter)

The output from the primary transmitter (LVDT) is connected in phase opposition to the output from the ferrodynamic compensating converter. The resultant difference voltage, is amplified by the solid state electronic amplifier which drives the balancing motor. Through mechanical linkage, the balancing motor in turn positions the moving coil of the converter in such a way that the unbalance voltage becomes zero. The motor's output through some mechanical linkage and cam mechanism is also fed to the pointer and the pen, which moves on a scale calibrated in terms of the measured variable.

Chart Drive Mechanism

The usefulness of a recording instrument depends upon

(i) The selection of the proper chart speed, and
(ii) The selection of the suitable drive element.

The selection of chart speeds is an important factor for separating the records of specific deviation.

A drive element should be a positive drive, it should require minimum maintenance and be able to run under certain extreme ambient conditions.

CIRCULAR CHART RECORDER 12.5

As the name implies, the data is recorded on a flat circular chart. The basic assembly of a single pen circular chart recorder is shown in Fig. 12.8.

It consists of a measuring element, an operating mechanism, a chart drive, and a recording device, which may all be mounted on a single panel. The chart is usually mounted on a flat supporting plate and fastened in position by spring clips, which prevent it from curling. The measuring element could be a helical pressure tube or any other element. The operating mechanism consists of levers a and b and links c which convey motion from the measuring element to the recording device.

For optimum recording conditions, light uniform pressure, and a smooth flat chart surface must be ensured. The pen arm must be accurately fitted and locked. The chart is driven at a uniform rate by some timing device.

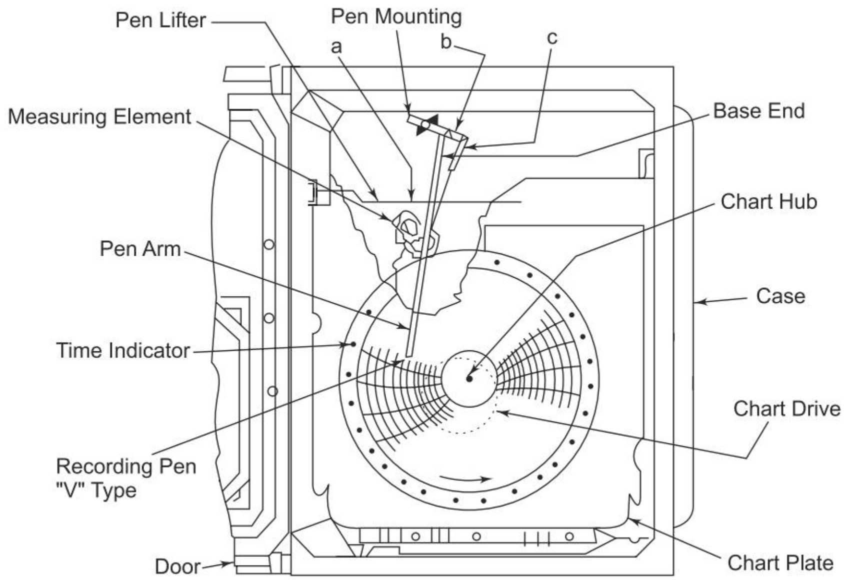


Fig. 12.8 Assembly of a single pen direct-acting circular chart recorder

Circular Chart Drive A conventional drive assembly consists essentially of a centre chart support spindle and an drive motor. The chart is generally centered and locked around the centre spindle. The chart speed may vary from one rotation in 15 minutes to one rotation in 30 days. A 30 cm circular chart has a maximum time-line length of approximately 100 cm in a revolution, whereas a 15 cm strip chart may move 25 m of chart in the same period. The various drives for circular charts are classified as follows.

1. Mechanical (spring clock drive).
2. Pneumatic (air lock drive).
3. Electric (synchronous regulated dc motor or motor wound spring).
4. Dual powered drive (duplex), i.e. a synchronous motor and spring clock mechanical drive.
5. Externally controlled drives.

X–Y RECORDER 12.6

In most research fields, it is often convenient to plot the instantaneous relationship between two variables [$Y = f(x)$], rather than to plot each variable separately as a function of time.

In such cases, the X–Y recorder is used, in which one variable is plotted against another variable.

In an analog X–Y recorder, the writing head is deflected in either the x-direction or the y-direction on a fixed graph chart paper. The graph paper used is generally squared shaped, and is held fixed by electrostatic attraction or by vacuum.

The writing head is controlled by a servo feedback system or by a self balancing potentiometer. The writing head consist of one or two pens, depending on the application.

In practice, one emf is plotted as a function of another emf in an X–Y recorder.

In some cases, the X–Y recorder is also used to plot one physical quantity (displacement, force, strain, pressure, etc.) as a function of another physical quantity, by using an appropriate transducer, which produces an output (EMF) proportional to the physical quantity.

The motion of the recording pen in both the axis is driven by servo-system, with reference to a stationary chart paper. The movement in x and y directions is obtained through a sliding pen and moving arm arrangement.

A typical block diagram of an X–Y recorder is illustrated in Fig. 12.9.

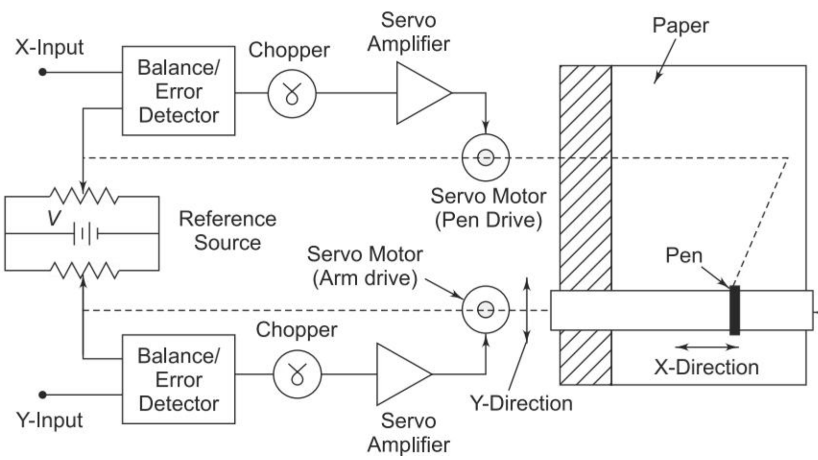


Fig. 12.9 Basic X–Y recorder

Referring to Fig. 12.9, each of the input signals is attenuated in the range of 0–5 mV, so that it can work in the dynamic range of the recorder. The balancing circuit then compares the attenuated signal to a fixed internal reference voltage. The output of the balancing circuit is a dc error signal produced by the difference between the attenuated signal and the reference voltage. This dc error signal is then converted into an ac signal with the help of a chopper circuit. This ac signal is not sufficient to drive the pen/arm drive motor, hence, it is amplified by an ac amplifier. This amplified signal (error signal) is then applied to actuate the servo motor so that the pen/arm mechanism moves in an appropriate direction in order to reduce the error, thereby bringing the system to balance. Hence as the input signal being recorded varies, the pen/arm tries to hold the system in balance, producing a record on the paper.

The action described above takes place in both the axes simultaneously. Hence a record of one physical quantity with respect to another is obtained.

Some X–Y recorders provides x and y input ranges which are continuously variable between 0.25 mV/cm and 10 V/cm, with an accuracy of ± 0.1% of the full scale. Zero offset adjustments are also provided.

The dynamic performance of X–Y recorders is specified by their slewing rate and acceleration. A very high speed X–Y recorder, capable of recording a signal up to 10 Hz at an amplitude of 2 cm peak to peak, would have a slewing rate of 97 cm/s and a peak acceleration of 7620 cm/s.

An X–Y recorder may have a sensitivity of 10 μV/mm, a slewing speed of 1.5 ms and a frequency response of about 6 Hz for both the axis. The chart size is about 250 × 180 mm. The accuracy of X–Y recorder is about ± 0.3%.

Applications of X–Y Recorders

These recorders are used to measure the following.

1. Speed-torque characteristics of motors.
2. Regulation curves of power supply.
3. Plotting characteristics of active devices such as vacuum tubes, transistors, zener diode, rectifier diodes, etc.
4. Plotting stress-strain curves, hysteresis curves, etc.
5. Electrical characteristics of materials, such as resistance versus temperature.

12.6.1 Digital X–Y Plotters

The rapid increase in the development in digital electronics has led to the replacement of analog X–Y recorders by digital X–Y plotters. The latter provide increased measurement and graphics capabilities. Digital X–Y plotters use an open loop stepping motor drive, in place of the servo motor drive used in analog X–Y recorders.

Digital measurement plotting systems provide the following features:

1. Simultaneous sampling and storage of a number of nput channels.
2. A variety of trigger modes, including the ability to display pre-triger data.
3. Multi-pen plotting of the data.
4. Annotation of the record with date, time and set up conditions.
5. An ability to draw grids and axis.

Communication with such devices can be done by means of the IEEE 488 or RS232 interface.

Graphic plotters are used to obtain hard copy from digital data input. By the use of appropriate software and hardware, these devices can draw grids, annotate charts and differentiate data by the use of different colours and line types. They are specified by their line quality, plotting speed and paper size.

MAGNETIC RECORDERS 12.7

The major advantage of using a magnetic tape recorder is that once the data is recorded, it can be replayed an almost indefinite number of times.

The recording period may vary from a few minutes to several days. Speed translation of the data captured can be provided, i.e. fast data can be slowed down and slow data speeded up by using different record and reproduce speeds.

The recorders described earlier have a poor high frequency response. Magnetic tape recorder, on the other hand, have a good response to high frequency, i.e. they can be used to record high frequency signals. Hence, magnetic tape recorders are widely used in instrumentation systems.

Basic Components of a Tape Recorder

A magnetic tape recorder consists of the following basic components.

1. Recording Head
2. Magnetic Head
3. Reproducing Head
4. Tape transport mechanism
5. Conditioning devices

Magnetic Recording The basic elements of a simple magnetic recording system are illustrated in Fig. 12.10(a).

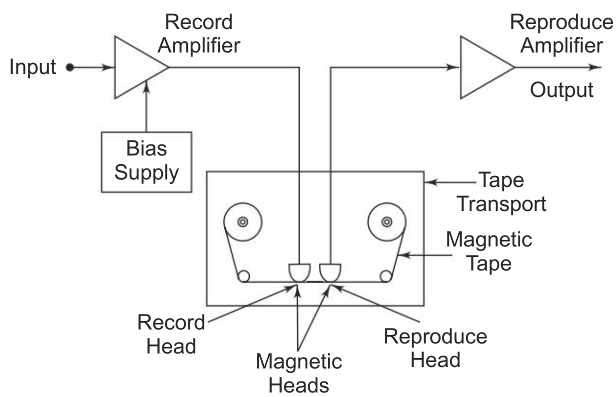


Fig. 12.10 (a) Elementary magnetic tape recorder

The magnetic tape is made of a thin sheet of tough, dimensionally stable plastic, one side of which is coated with a magnetic material.

Some form of finely powdered iron oxide is usually cemented on the plastic tape with a suitable binder. As the tape is transferred from one reel, it passes across a magnetising head that impresses a residual magnetic pattern upon it in response to an amplified input signal.

The methods employed in recording data on to the magnetic tape include direct recording, frequency modulation (FM) and pulse code modulation (PCM).

Modulation of the current in the recording head by the signal to be recorded linearly modulates the magnetic flux in the recording gap. As the tape moves under the recording head, the magnetic particles retain a state of permanent magnetisation proportional to the flux in the gap. The input signal is thus converted to a spatial variation of the magnetisation of the particles on the tape. The reproduce head detects these changes as changes in the reluctance of its magnetic circuit which induce a voltage in its winding. This voltage is proportional to the rate of change of flux. The reproduce head amplifier integrates the signal to provide a flat frequency characteristics.

Since the reproduce head generates a signal which is proportional to the rate of change of flux, the direct recording method cannot be used down to dc. The lower limit is around 100 Hz and the upper limit for direct recording, around 2 MHz. The upper frequency limit occurs when the induced variation in magnetisation varies over a distance smaller than the gap in the reproduce head.

The signal on an exposed tape can be retrieved and played out at any time by pulling the tape across the magnetic head, in which a voltage is induced.

It is possible to magnetise the tape longitudinally or along either of the other two main axis, but longitudinal magnetisation is the best choice.

Figure 12.10(b) shows simply how the tape is magnetised. If a magnetic field is applied to any one of the iron oxide particles in a tape and removed, a residual flux remains. The relationship between the residual flux and the recording field is determined by the previous state of magnetisation and by the magnetisation curves of the particular magnetic recording medium.

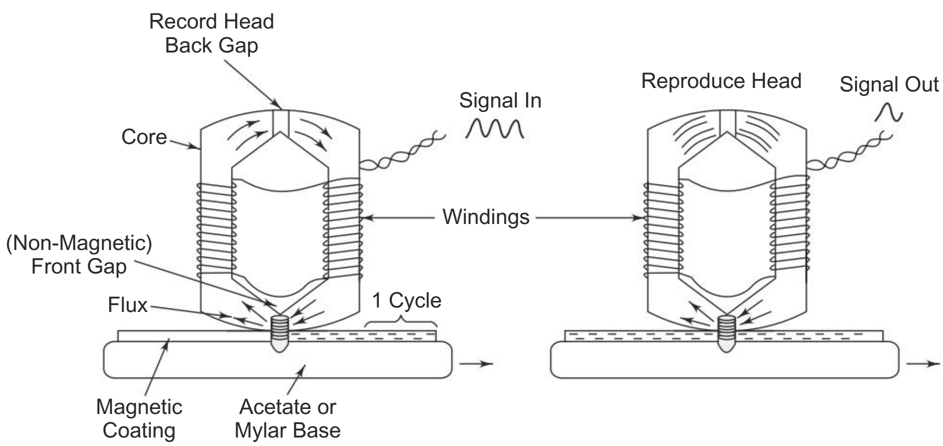


Fig. 12.10 (b) Magnetisation of tape

A simple magnetic particle on the tape might have the $B - H$ curve shown in Fig. 12.10(c) where H is the magnetising force and B the flux density in the particle.

Consider the material with no flux at all, i.e. the condition at point 0.

Now if the current in the coil of the recording head [Fig. 12.10(b)] is increased from 0 in a direction that gives positive values of H, the flux density increases along the path 0 – 1– 2, until the material is eventually saturated. If the operating point is brought from 0 only as far as 1, and H is brought back to 0, B follows a minor hysteresis loop back to point 6. A greater value of coil current would leave a higher residual flux, and a lower current a lower residual; a very simple recording process results.

However, the linearity between residual flux and recording current is very poor. Hence to obtain linearity in direct recording, FM is used. In all systems, the signal is reproduced by passing the magnetised tape over a magnetic head similar to the recording head. The magnetisation of the particles on the tape induces a varying flux in the reproducing head and a voltage is induced in the coil, proportional to the rate of change of flux.

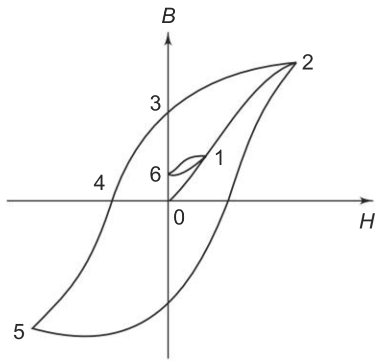


Fig. 12.10 (c) Typical magnetisation curve

Methods of Recording

There are three methods of magnetic tape recording which are used for instrumentation purposes.

1. Direct recording
2. FM recording
3. Pulse Duration Modulation recording (PDM)

FM recorders are generally used for instrumentation purposes. PDM recording is used in instrumentation for special applications where a large number of slowly changing variables have to be recorded simultaneously.

Direct Recording

This type of recording is described in the introduction of Sec. 12.7.

Advantage of Direct Recording

1. This recording process has a wide frequency response, ranging from 50 Hz – 2 MHz for a tape speed of 3.05 m/s. It provides the greatest bandwidth obtainable for a given recorder.
2. It requires simple electronic circuits.
3. It has a good dynamic response and takes overloads without increase in distortion. In general, instrumentation recorders have a signal to noise ratio of 22 – 30 db at 1% total harmonic distortion.
4. It is used to record signals where information is contained in the relation between frequency and amplitude, such as the spectrum analysis of noise.
5. It can be used for recording voice signals.

Disadvantage

1. The direct recording process is characterised by some amplitude instability caused by random surface inhomogeneities in the tape

coating. Some portion may not be perfectly recorded owing to dirt or poor manufacture.

2. This recorder is only used when maximum bandwidth is required.

FREQUENCY MODULATION (FM) RECORDING 12.8

When a more accurate response to dc voltages is required, an FM system is generally used.

In this FM system, the input signal is used to frequency modulate a carrier, i.e. the carrier signal is frequency modulated by the input signal (FM modulation), which is then recorded on the tape in the usual way.

The central frequency is selected with respect to the tape speed and frequency deviation selected for the tape recorders is ± 40% about the carrier frequency.

The reproduce head reads the tape in the usual way and sends a signal to the FM demodulator and low pass filter, and the original signal is reconstructed. The signal to noise ratio (S/N) of an FM recorder is of the order of 40 – 50 db, with an accuracy of less then ± 1%.

This ± 1 db flat frequency response of FM recorders can go as high as 80 kHz at 120 in/s tape speed, when using very high carrier frequencies (above 400 kHz).

When high frequency (HF) is not needed, and with a view to conserving tape. A tape speed range selector is generally provided. When the tape speed is changed, the carrier frequency also changes in the same proportion. Therefore, no matter what tape speed is being used, the recorded wavelength of a given dc input remains the same, since ± 40% full scale frequency deviation is utilised in all cases. A common set of specifications are given in Table 12.1.

Table 12.1

Tape speed in/s	*Carrier frequency kHz*	*Flat frequency response ± 0.5 db, Hz*	*RMS (S/N) ratio*
120.0	108.0	0 – 20,000	50
60.0	54.0	0 – 10,000	50
30.0	27.0	0 – 5,000	49
15.0	13.5	0 – 2,500	48
7.5	6.75	0 – 1,250	47
3.75	3.38	0 – 625	46
1.88	1.68	0 – 312	45

Input to the tape recorders is generally at the 1 V level, and so most transducers require amplification before recording.

An FM recording system is illustrated in Fig. 12.11. In this system a carrier oscillator frequency f_c, called the centre frequency, is modulated by the level of the input signal.

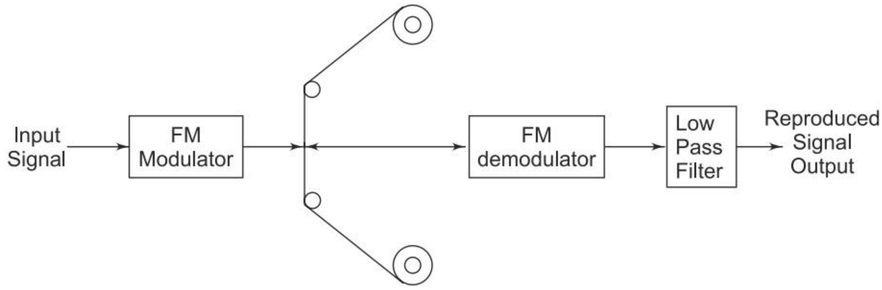


Fig. 12.11 FM recording system

When there is no input signal, i.e. zero input, the modulation is at centre frequency f_c. If a positive input signal is applied, the frequency deviates from the centre frequency by some amount in a certain direction; the application of a negative input voltage deviates the carrier frequency in the opposite direction.

The output of the modulation, which is fed to the tape, is a signal of constant frequency for dc inputs, and varying frequency for ac inputs. The variation of frequency is directly proportional to the amplitude of the input signal.

On playback, the output of the reproduce head is demodulated and fed through a low pass filter which removes the carrier and other unwanted frequencies reproduced due to the modulation process.

The operation of FM modulation can be easily checked by applying a known input voltage and measuring the output frequency with an electronic counter. This signal is applied to the tape with no further conditioning, as the signal is independent of the amplitude.

The FM demodulator converts the difference between the centre frequency and the frequency on the tape, to a voltage proportional to the difference in the frequencies. This system can thus record frequencies from dc to several thousand Hertz. Residual carrier signals and out of band noise are removed by a low pass filter.

Advantages of FM Recording

1. FM recording is useful primarily when the dc component of the input signal is to be preserved.
2. This system has a wide frequency range and can record from dc voltages to several kHz.
3. There is no drop-out effect due to inhomogeneties of the tape material.
4. Independent of amplitude variations, and accurately reproduces the waveform of the input signal.
5. Used extensively for recording voltages derived from non-electrical quantities, such as force, acceleration and pressure.
6. It is extremely useful for multiplexing in an instrumentation system.

Disadvantages

1. FM recording is extremely sensitive to tape speed fluctuations.
2. FM recording circuitry is more complicated than that of direct recording systems.
3. FM system has a limited frequency response.
4. It requires a high tape speed.
5. It requires a high quality of tape transport and speed control.

Pulse Duration Modulation

Pulse width modulation is also called pulse duration modulation (PDM). In this system, the amplitude and the starting time of each pulse is kept fixed, but the width of the pulse is made proportional to the amplitude of the signal at that instant. This type of system is mostly used for Digital recording.

DIGITAL DATA RECORDING 12.9

Digital magnetic tapes are often used as storage devices in digital data processing applications. Digital tape units are of two types, incremental and synchronous.

Incremental digital recorders are commanded to step ahead (increment) for each digital character to be recorded. Input data may be at a relatively slow, or even discontinuous rate. In this way, each character is equally and precisely spaced along the tape.

In a synchronous digital recorder, the tape moves at a constant speed (about 75 cm/s) while a large number of data characters are recorded. The data inputs are at precise rates, up to tens of thousands of characters per second. The tape is rapidly brought up to speed, recording takes place, and the tape is brought to a fast stop. In this way a block of characters (a record) is written with each character spaced equally along the tape. Blocks of data are usually separated from each other by an erased area on the tape called the record gap. The synchronous tape unit starts and stops the tape for each block of data to be recorded.

Characters are represented on magnetic tape by a coded combination of 1-bit in appropriate tracks across the tape width. The recording technique used in most instrumentation tape recorders is the industry accepted IBM format of Non-Return Zero (NRZ) recording.

In this system the tape is magnetically saturated at all times in either the positive or the negative direction.

The NRZ method uses the change in flux direction on the tape to indicate 1 bit, and no change in flux direction to indicate 0 bit. This method is illustrated in Fig. 12.12, where the binary numbr 11101011 is represented by a flux pattern in the NRZ system.

The simplest method of coding the recording head field is to reverse its direction.

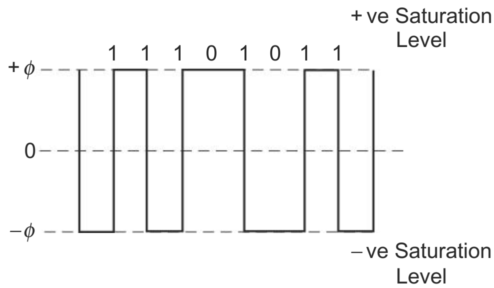


Fig. 12.12 NRZ method of digital recording

In digital data recording, a recording field of amplitude sufficient to produce magnetic saturation through the complete tape layer thickness is reversed to record a 1 signal and kept constant to record a 0 signal.

Reproduction of this recording is achieved by using a timing signal obtained from a separate clock track, corresponding to the time when a 1 or 0 is recorded.

Self-clocking systems, where the recording field is reversed at regular intervals and 1 or 0 signals are recorded between these clock signals, are also used.

It is evident that the highest resolution is obtained in NRZ recording by adjusting the field amplitude, so that maximum longitudinal decrement occurs in the surface layer of the tape. In practice, larger fields are usually employed to ensure more reliable recordings on a coated thicker tape. To minimise the effects of dropouts, large recording fields are used and resolution is sacrificed for increased reliability.

Present high density data recording on oxide powder tape is in the range of 1500 – 2000 flux reversals per inch. (By using thin metallic coatings with high coercive force, extensions up to 10000 reversals per inch are possible in future.)

Since magnetisation is independent of frequency and amplitude but relies only on the polarity of the recording current, the usual problems of non-linearity and distortion found in direct and FM recordings do not exist. The write coils of the tape head require only sufficient current, of the correct polarity, to saturate the tape. Two of the problems encountered in digital recording are signal drop out and spurious pulses (losing or adding data). Signal dropout or loss of pulses becomes serious when the packing density increases (a large number of bits per unit tape length).

As a check on dropout errors, most tape systems include a parity check. This check involves keeping track of the number of 1 bits of information initially recorded on the tape by writing a parity check pulse on an extra tape track. If the number of 1's recorded is even it is called an even parity check, and if the number of 1's recorded is odd, then it is an odd parity check. When a dropout occurs, the parity check does not agree with the actual recorded data and a parity error is detected.

Some systems use the parity error system to insert missing bits in the appropriate places in addition to indicating that a parity error has occurred.

Another scheme, called bipolar or alternate mark inversion, is illustrated in Fig. 12.13.

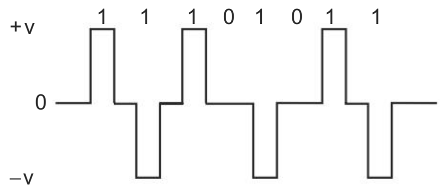


Fig. 12.13 Bipolar (Alternate mark inversion

This format has no residual dc component and has zero power in the spectrum at zero frequency, as shown in Fig. 12.13. These are pulses of 50% duty cycle (they are only half as wide as the pulse interval allows) and by inverting the polarity of alternate 1 bits. The bipolar format is really a three state signal (+ V, 0, – V).

Advantages of Digital Data Recording

1. High accuracy.
2. Insensitivity to tape speed.
3. Use of simple conditioning equipment.
4. The information is fed directly to a digital computer for processing and control.

Disadvantages of Digital Data Recording

1. Poor tape economy.
2. The information from transducers is in analog form, hence an A/D converter is required.
3. A high quality tape and tape transport mechanism are required.

OBJECTIVES AND REQUIREMENTS OF RECORDING DATA 12.10

1. Recording is often carried out in order to preserve the details of measurement at a particular time.
2. The accuracy of the recording must be the same as the accuracy of the measurement, for best results.
3. A record should be legible and capable of being maintained properly.
4. Most of the critical parameters which influence the performance of the process or equipment has to be recorded for taking necessary action from time to time.
5. The recorded chart at a glance provides an overall picture of the performance of the unit. (All parameters automatically regulated are invariably recorded to depict the performance of automatic regulating loop.)
6. The recorded chart also reflects immediately what actions the operator had taken during his shift.
7. The necessary data for determining the efficiency, etc. is easily and readily provided.

8. Charts are also used as permanent records to provide answers to queries which may come up at a later time with respect to product quality.
9. It is also very valuable from the point of preventive maintenance.
10. Manufacturers of equipment often ask for the use of a recorder for recording certain parameters which are very critical for the performance of the equipment. The recorded chart indicates whether the equipment has been used as per their instructions or not.

RECORDER SELECTIONS FOR PARTICULAR APPLICATIONS 12.11

When selecting a recorder for a particular application, the following points should be considered.

The basic consideration is the frequency of the waveform to be recorded. For signals in the frequency range of about 125 Hz to a few thousand Hz, an optical recorder is well suited. This is a strip chart recorder that directs a light beam through an optical system and onto a photographic plate. The paper used in such instruments may require darkroom developing or the paper may be light sensitive and self develop when exposed to light. (See Fig. 12.2(b).) Optical recorders can also be used for low frequencies. The disadvantage is that the photographic plate is much costlier than the paper in instruments using a pen and ink system.

For signals in the frequency range of 50 – 125 Hz, a servo-type strip chart recorder with a preamplifier is most suitable. The preamplifier is recommended to provide additional energy to the pen or stylus at the rate required to record waveforms. At frequencies of 10 Hz or less, a servo-type recorder offers the user sensitivity, linearity, stability, mechanical sluggishness and sufficient energy to drive the pen or stylus, as well as a control device if it is part of the control system.

RECORDER SPECIFICATIONS 12.12

The following general features should be considered whenever one is examining a recorder.

1. The type of writing instrument desired, such as pen and ink, heated stylus or electric stylus.
2. The type of recorded chart desired, such as paper, wax-coated paper, or photographic film with or without darkroom developing.
3. Maximum amplification, if the signal is to be recorded.
4. Frequency response (Hz).
5. Recording speed adjustors.
6. Input signals voltage or current.
7. Input impedance, should be several hundred thousand ohms for a general purpose recorder.
8. Charts speed expressed in inches or centimeters per second, per min. or per hour.

POTENTIOMETRIC RECORDER (MULTIPOINT) 12.13

12.13.1 Principle of Operation

The thermocouple or millivolt signal is amplified by a non-inverting MOSFET, chopper stabilised, feedback amplifier. This configuration has a very high input impedance and the current passing through the signal source is a maximum of 0.5 nA (without broken sensor protection). With the use of span control, the output signal is adjusted to 5 V (nominal), for an input signal change, e.g. to full scale deflection of the pen.

The pre-amplifier output signal is then compared with a reference voltage picked off the measuring slide wire, which is energised by a stabilised power supply, and the difference amplified by a servo amplifier, whose output drives a linear motor. The motor carriage carries the indicating pointer pen and the sliding contact on the slide wire.

The motor itself consists of a coil assembly, travelling in a magnetic field (Fig. 12.14). The servo amplifier drives the carriage in the appropriate direction, to reduce the difference signal to zero.

An input filter circuit reduces any spurious signals picked up by the input leads. The B–E junctions of the transistor are used for overload protection.

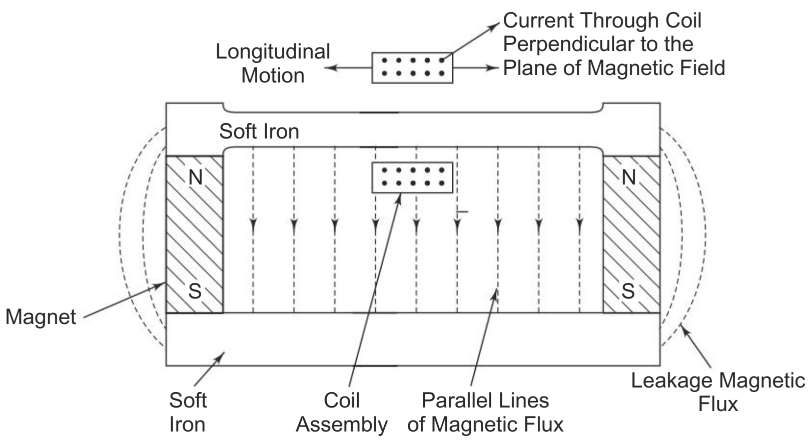


Fig. 12.14 Linear motor operating principle

In multipoint versions of the recorder, a signal selector switch is driven by a synchronous motor. The pen changeover and dotting action are actuated by a second synchronous motor (Fig. 12.15). The pen operation is electrically synchronised to the rotation of the signal selector switch—should they get out of step, the pen motor stops at a predetermined point and restarts only when the signal selector switch has rotated to its correct alignment. The linear motor is muted during the signal changeover, but dotting always takes place with the system live. The standard dotting interval is 6s (Fig. 12.15). The linear motor consists of a coil assembly travelling in a magnetic field. The direction of the field is as shown in Fig. 12.14. The current passing through the coil produces a magnetic field which is perpendicular to the existing field and causes the carriage to move in a direction given by Fleming's left hand rule.

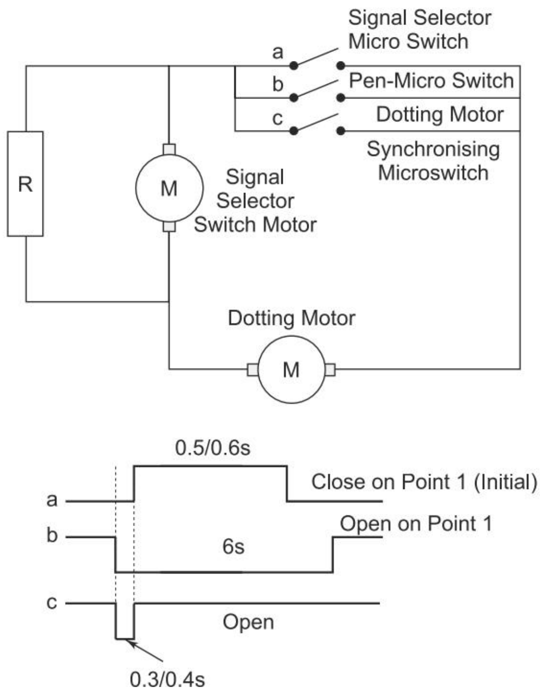


Fig. 12.15 Multipoint recording synchronising circuit

12.13.2 Ranging Circuit

With the indicating pointer positioned at the minimum of the scale, the output from the preamplifier should be zero. A small voltage appears across R_1 and R_4 due to current feeds from R_2 and R_3 respectively. R_1 is kept less than R_4 (in general, but $R_1 > R_4$ in thermocouples) and additional current is passed through R_1 from RV_1, so that the output can be adjusted to read zero corresponding to a zero input signal. When a signal input equivalent to full scale is given, RV_3 is adjusted to make the indicating pointer read the maximum on the scale. For thermocouple inputs, the circuit remains as above. For automatic cold junction compensation, the variation in the base-emitter voltage of the transistor with respect to temperature is made use of. The resistance R_7 injects current through R_4. The value of R_4 depends on the type of thermocouple used. Figure 12.16(a) shows a ranging circuit.

For input signals greater than 100 mV, an additional range card is used. Figure 12.16(b) shows the diagram of a resistance thermometer. For resistance thermometer input, a separate power supply gives the bridge supply as 6.2 V. One

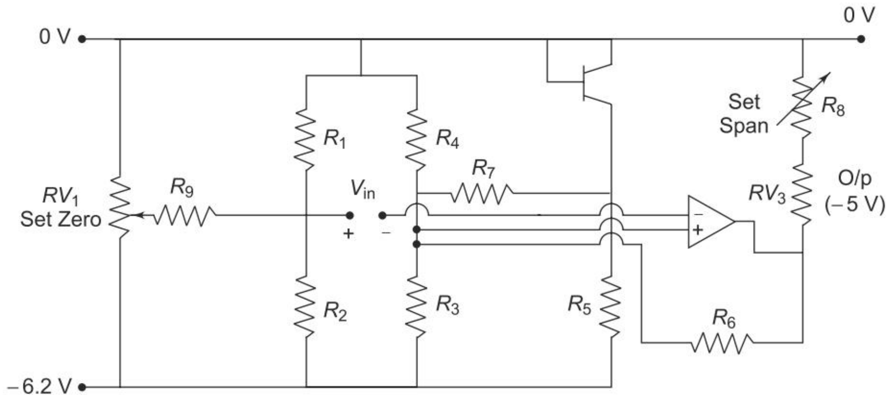


Fig. 12.16 (a) Ranging circuit (Millivolt and thermocouple inputs)

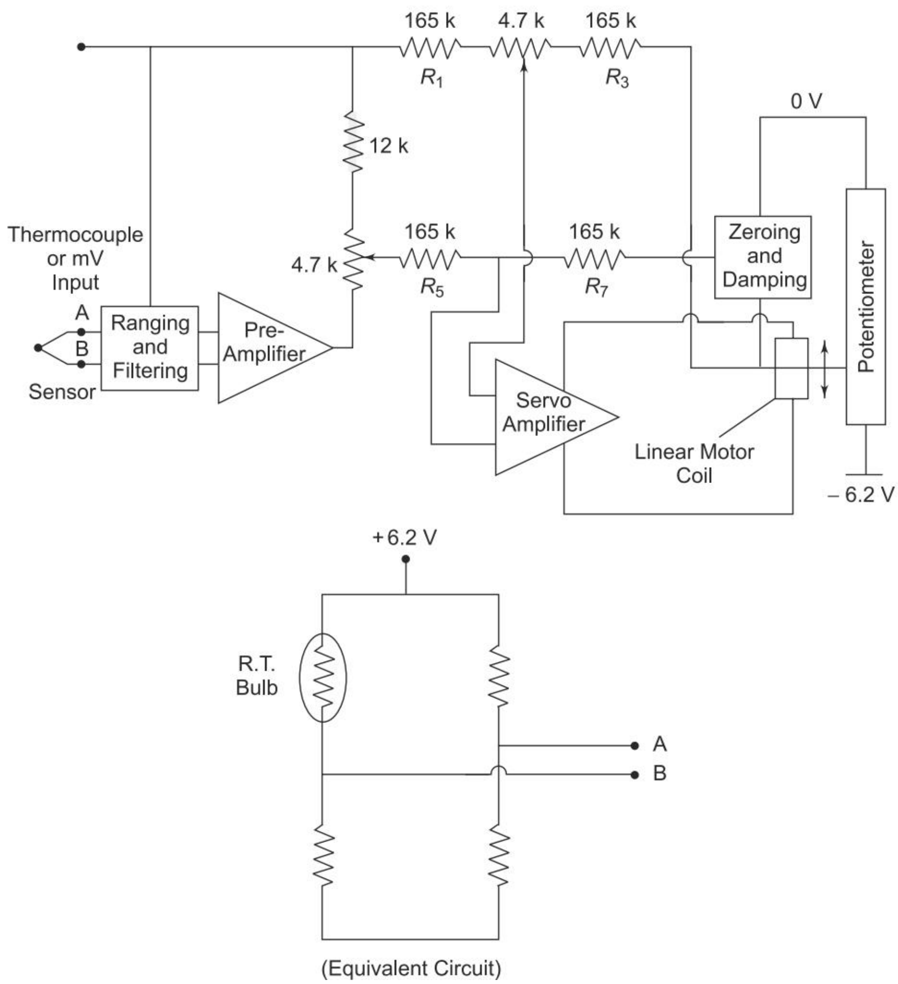


Fig. 12.16 (b) Basic diagram of a resistance thermometer

of the arms of the bridge is a resistance bulb (the standard three wires connection of the resistance thermometer). The other three arms of the bridge are mounted on an additional range card. The bridge gives 10 mV nominal output for the range of temperatures under consideration.

12.13.3 Servo Amplifier

The servo amplifier shown in Fig. 12.16(c) compares the output from the preamplifier with the voltage picked off the measuring slide wire. The output of the preamplifier and the voltage picked up by the slide wire obtained via an emitter follower stage are connected in a high resistance bridge circuit ($R_1 = R_3 = R_5 = R_7 = 165$ kΩ), the balance of which is adjusted for maximum common mode rejection by adjusting the potentiometer, R_{V_1} (4.7 kΩ). The output is fed to the differential amplifier (consisting of a matched pair Opamps).

The signal voltage picked from the slide wire is also used to control the damping of the servo system. For this, signal voltage from the slide wire is fed to a differentiator circuit using an Opamp and its associated circuit and damping control.

The indicating point may be adjusted to set scale minimum, when the preamplifier output is zero. This is done by adjusting a set zero potentiometer (1 k value) coming in the damping circuit.

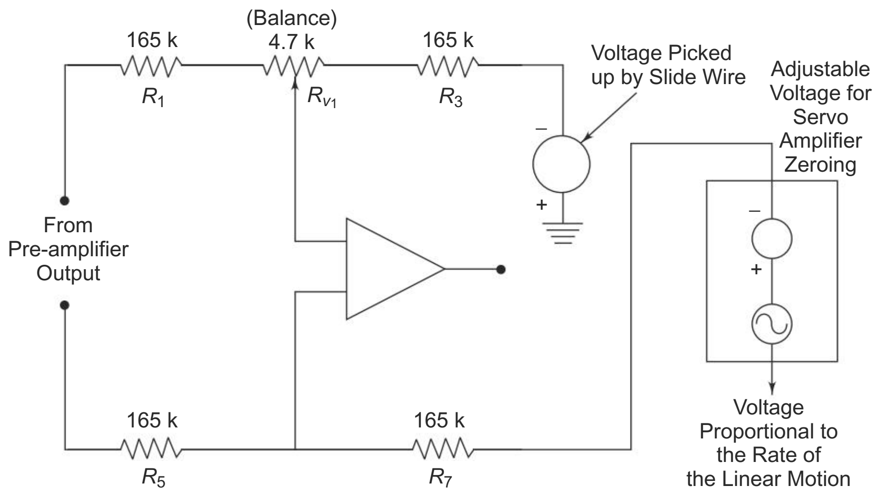


Fig. 12.16 (c) Schematic section for servo amplifier

The output of the differential amplifier is fed to a dc amplifier (consisting of typical NPN/PNP transistor stages), which give sufficient output to drive the linear motor in the appropriate direction. This reduces the differential input signal to the servo amplifier to zero.

12.13.4 Multipoint Variation

In multipoint recorders, each signal to be measured is connected in turn to the ranging circuit input terminals by means of a rotating selector switch, which is driven from a synchronous motor having a speed of 10 rpm via an intermittent motor. Each signal is sampled for a period of 3 or 6 seconds, depending on the switch motor fitted.

A second synchronous motor (10 rpm – 24 V ac) is mounted and fitted with cams and microswitches (Fig. 12.15).

The cam at the extreme end actuates the pen lifting (and dotting) bar and initiates the pen dotting. As the stylus lifts off the paper, a pawl and ratchet system in the pen head rotates the stylus and ink well to the next point.

A synchronising microswitch, mounted under the pen head, is operated by a pip on the stylus shaft. The contacts are open while the pen stylus is at initial point 1.

Fitted to the shaft of the dotting motor is a second cam, which operates a second microswitch in the synchronising circuit. The contact switches are open for a period of 0.3 – 0.4s, once in each revolution of the dotting motor, which rotates once during the sampling period of each print, i.e. 6s.

The third synchronising microswitch is operated by the projection on the side of the plastic intermittent advance gear on the same shaft of the signal switch. The contact of this microswitch closes for a period of 0.5 – 0.6s once in each revolution of the signal switch, i.e. once every 72s for a 6s sampling.

When the recorder goes out of synchronisation, if the cams and pen stylus are set correctly, the dotting motor will continue operating until the stylus comes around to the initial point. At this point the pen microswitch opens, as does the dotting microswitch. The dotting motor then stops.

The signal selector switch motor operates continuously, and when the microswitch operated by this motor closes, the dotting starts again.

(Note: The closing time of the signal selector microswitch is slightly more than the open time of the dotting motor microswitch, ensuring that the dotting motor microswitch be closed again before the signal selector switch motor microswitch opens.)

12.13.5 Multiple Recorders (6 Point Recorders)

The controller consists of a high gain opamp in comparator configuration, with one input from the pre-amplifier and another from the set point potentiometer. The output of the comparator is connected to the Schmitt trigger via both high and low control microswitches, operated by control cams, as shown in Fig. 12.17.

The Schmitt trigger circuit maintains the relay state after the input is disconnected, i.e. while the signal selector switches scan the other signal input. The control microswitch ensures that the relay does not change state during signal changeover, while the pre-amplifier input is open-circuit. Once the relay gets energised, it remains in that state via its own (normally open) contacts and through the Reset button.

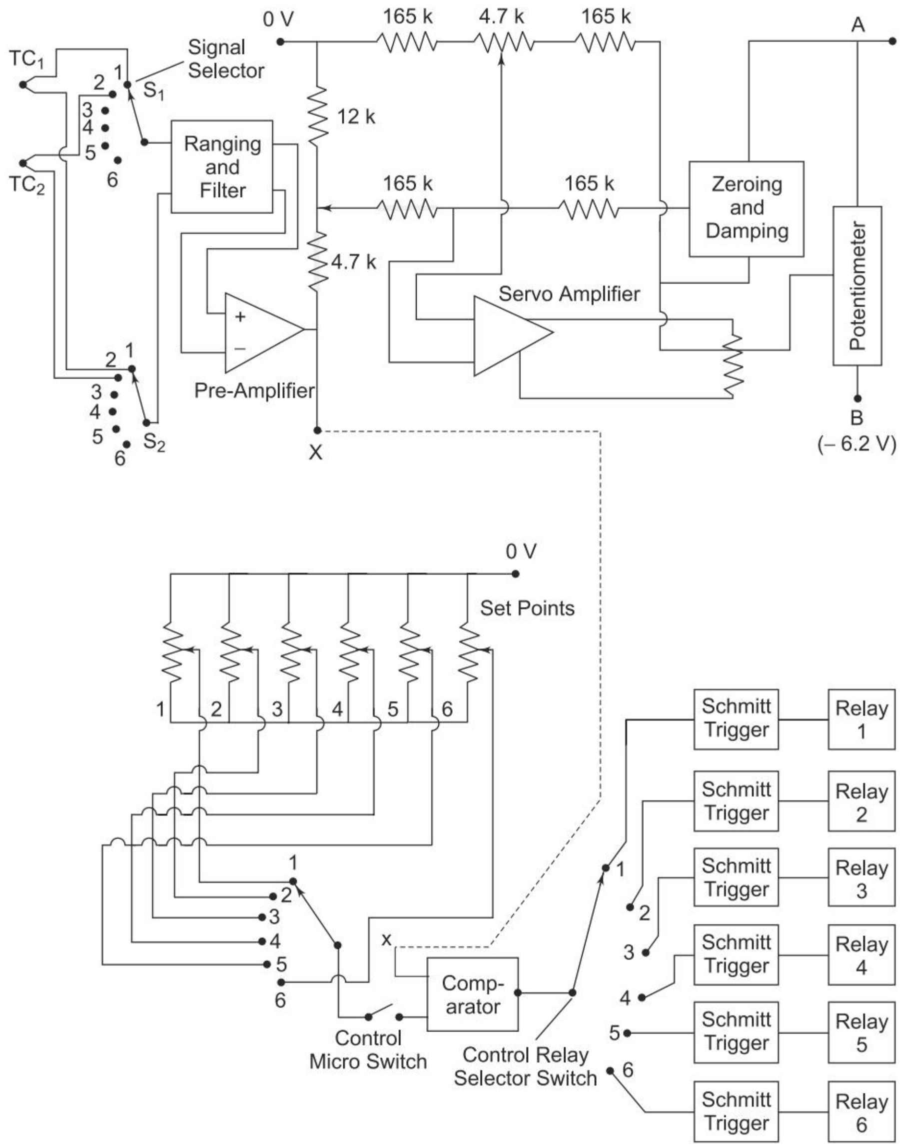


Fig. 12.17 Multipoint decoder having 6 on/off control points for a 6 point recorder

When the Reset button is pressed, the relay gets de-energised, which again gets latched if the same condition ($V_{in} > V_{set}$) prevails for the same or any other channel.

DIGITAL MEMORY WAVEFORM RECORDER (DWR) 12.14

The digital memory waveform recorder shown in Fig. 12.18, provides capability which is difficult to achieve by other methods. The block diagram (Fig. 12.18) is of a unit capable of recording four independent signals simultaneously.

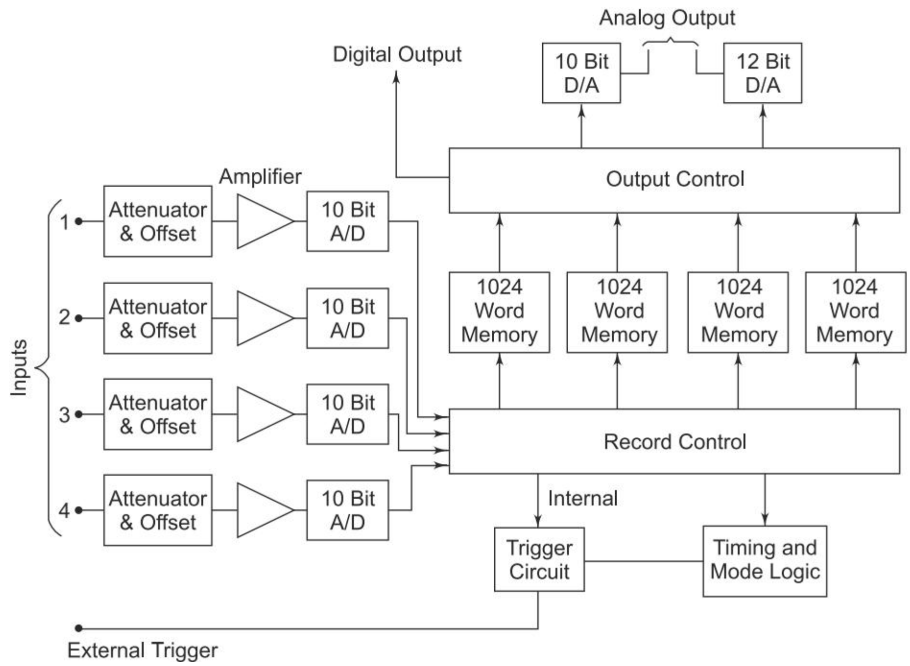


Fig. 12.18 Digital memory waveform recorder

The practical implementation of this concept rests on the ready availability, small size and reasonable cost of digital memories and associated digital hardware.

Consider a signal channel, the analog voltage input signal is digitised by a 10 bit A/D converter with a resolution of 0.1% (1 part in 1024) and frequency response of 25 kHz.

The total digital memory of 4096 words can be used for a single channel 2048 words can be used for each of the two channels an 1024 for each of the four channels.

The analog input voltage is sampled at adjustable rates (up to 100000 samples per second) and the data points are read into the memory. A maximum of 4096 points are storable in this particular instrument.

(Sampling rate and memory size must be selected to suit duration and waveform of the physical event being recorded. Some models have sampling rates up to 200 MHz.)

Once the sampled record of the event is captured in the memory, many useful manipulations are possible, since the memory can be read without erasing it (non-destructive readout).

By reading the memory out slowly through a digital to analog converter, the original event (which could have been extremly rapid) is reproduced as a slowly changing voltage, that is easy to record permanently in large size on a slow X–Y plotter.

If the memory is read out rapidly and repetitively, an input event which has a one shot transient becomes a repetitive waveform that now may be observed easily on an ordinary oscilloscope (not a storage CRO). The digital memory also may be readout directly (without going through a digital to analog converter) to say, a computer where a stored program can manipulate the data in almost any desired way.

Pre-trigger recording allows the device to record the input signal preceding the trigger point, a unique and often useful capability. In ordinary triggering, the recording process is started by the rise of the input signal (or some external triggering signal) above some preset threshold value. If this threshold is set too low, random noise will trigger the system; too high a threshold prevents recording of the initial rise of the desired signal. The digital recorder can be set to record continuously (new data coming into the memory pushes out the old data, once the data memory is full), until the trigger signal is received. Then the recording process is stopped, thereby freezing in the memory data received prior to the trigger signal.

An adjustable trigger delay allows operator control of the stop point, so that the trigger may occur near the beginning, middle or end of the stored information.

While digital memory waveform recorders are marketed without attached recording devices for analog outputs, such packages are available as a digital memory oscillographs or oscilloscopes.

The usual 40 Hz frequency response of pen/ink recorders can be extended to 20 kHz with a plug-in module using waveform digitising principles. This principle of waveform digitising makes direct recording easier. Direct writing instruments of high recording quality are available for recording a very wide range of signals.

APPLICATIONS OF A STRIP CHART RECORDER 12.15

The following are some of the thousands of applications for recorders in industry.

1. Temperature Recording A strip chart recorder may be used to provide a graphical record of temperature as a function of time. There are two primary methods used for recording temperature, the thermocouple method and the resistance method.

The thermocouple method utilises commercially available thermocouples that cover a wide range of temperature. These serve very well as temperature-sensing elements and are readily compatible with strip-chart recorders.

The circuit shown in Fig. 12.19(a), is a basic circuit used to record the thermocouple voltage.

A thermocouple (discussed in Sec. 3.6) is made by joining two dissimilar metals. A potential difference, which is proportional to the temperature, exists across the junctions. This potential difference has an almost linear relationship with the temperature and is very repeatable.

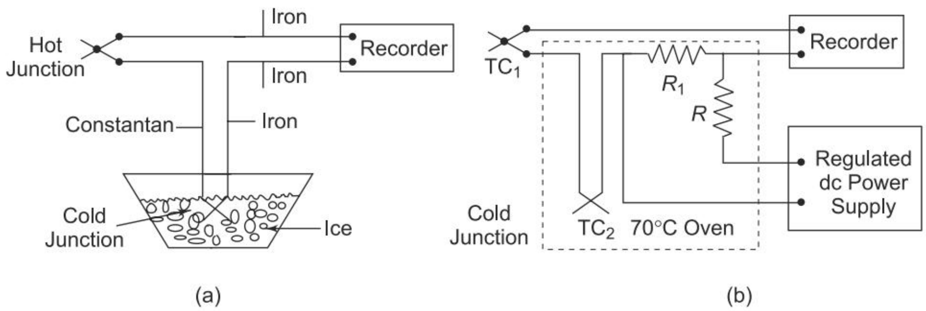


Fig. 12.19 (a) Basic circuit used to record thermocouple voltage, (b) Basic circuit with artificial reference junction voltage used to measure thermocouple voltage

Elaborate tables of temperature versus potential difference have been developed for certain pairs of metals that are used for commercial thermocouples. These tables allow one to determine the hot junction temperature when the cold junction or reference junction temperature is 0°C. If the reference temperature is not 0°C, correction factors may be used.

However, it is usually convenient to maintain the reference junction at 0°C or to develop an artificial reference junction emf by using a circuit as shown in Fig. 12.19(b).

Thermocouple TC_2 is maintained at 70°C by the oven. The emf developed by TC_2 is balanced out by the voltage drop across R, due to current flowing from the regulated dc power supply, until the total emf is equl to the emf of TC_2 at 0°C.

(The two metals connected to the recorder terminal are of the same metal.)

2. Sound Level Recording It is frequently desirable to obtain a record of the sound level over a period of time, near highways, airports, hospitals, schools or residence.

This can be done with an ordinary microphone and a strip-chart recorder, provided the output signal from the microphone is of sufficient amplitude to drive the recorder.

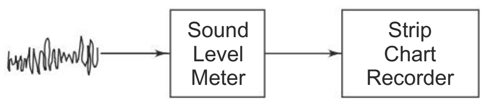


Fig. 12.20 Setup for measuring sound level as a function of time

A better setup for obtaining sound level data is to use a sound level meter and a strip chart recorder, as shown in Fig. 12.20.

3. Recording Amplifier Drift Transistor amplifiers are sensitive to temperature changes. Temperature changes causes the bias voltage of the transistor to change, thereby changing the operating or quiescent point, generally by a small amount. This change of the point is called Drift.

Since this is generally a slow process, a strip chart recorder may be used to monitor and record the drift by connecting the recorder to the output of the amplifier, as shown in Fig. 12.21.

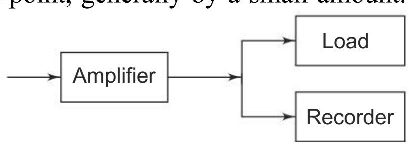


Fig. 12.21 Basic circuit setup to record amplifier drift

Review Questions

1. Differentiate between indicator and recorder.
2. List various types of recorders.
3. State the working principle of a strip chart recorder.
4. List three major types of a strip chart recorder.
5. List various pen mechanisms used for recording.
6. Define chart speed.
7. Describe the operation of a strip chart recorder.
8. Explain the working principle of a galvanometer type recorder.
9. Describe the operation of a galvanometer type recorder with the help of a diagram. State the disadvantages of a galvanometer type recorder.
10. State the limitations of a galvanometer recorder and how is it eliminated.
11. Why does the galvanometer recorder uses a curvilinear chart.
12. State the importance of paper speed.
13. Explain the basic principle of null type recorders.
14. State different types of null type detectors.
15. Describe the operation of null type recorder with the help of a diagram.
16. Compare between galvanometer and null type recorders.
17. Describe with diagram the working of a bridge type recorders.
18. Explain with diagram the operation of a LVDT type recorder.
19. State the requirement of a dc linear motor used in LVDT type recorder.
20. List three functions that a recorder may simultaneously serve in industrial applications.
21. List minimum five specifications that should be considered while selecting a recording instrument.
22. Describe three applications for recording instruments.
23. What is the approximate maximum frequency response of a galvanometer type recorder?
24. What are the primary functions of a galvanometer recorder?
25. Differentiate between galvanometer type recorder and potentiometric recorder.
26. Define single point recorder and multipoint recorder.
27. Define a circular recorder.
28. Explain the working principle of a circular recorder.
29. Describe the operation of a circular chart recorder. List various applications of a circular chart recorder.
30. Define a plotter.
31. Differentiate between plotter and recorder.
32. Describe the operation of a X-Y plotter. List the various applications of a XY plotter.
33. Classify different types of recorders used in instrumentation.
34. State the advantage of using a potentiometer circuit.
35. What is a servo system?
36. List the various objective and requirements of recording data.
37. List the various specification of a recorder.
38. List various applications of strip chart recorder.
39. State the purpose of error detector in a recorder.
40. How is automatic balance achieved in a recording instrument?
41. Give the constructional features of strip chart recorders. State its various applications.
42. An ADC signal cannot be reproduced if recorded on a direct recording system. Comment.
43. What are the basic components of a magnetic recorder.

44. Explain with diagram the operation of a magnetic recorder.
45. State the advantages of direct recording.
46. Explain with diagram the working of an FM recording system. State the advantages and disadvantages of FM recording.
47. Describe with the help of a block diagram the operation of an *X-Y* recorder. List the applications of an *X-Y* recorder.
48. Differentiate between strip chart and *X-Y* recorders.
49. Differentiate between circular chart and *X-Y* recorder.
50. Explain the working principle of a digital memory waveform recorder.
51. List the various mode of operation of a digital recorder.
52. Explain NRZ method of recording.
53. State the advantages and disadvantages of digital recording.
54. List the feature provided by a digital X-Y plotter.
55. Describe the operation of a digital memory waveform recorder.
56. Explain how transients can be recorded using digital waveform recorders.
57. Explain the concept of Pre-triggering as used in digital waveform recorders.
58. Explain A to D converter and D to A converter.
59. Explain the working principle of a digital *X-Y* plotter.
60. Explain the working principle of a digital recorder
61. Describe the operation of a digital recorder.

Multiple Choice Questions

1. A recorder is an instrument used for
   (a) recording (b) indicating
   (c) display (d) measurement
2. A strip chart recorder uses
   (a) a long roll of paper
   (b) a circular paper
   (c) a stationary paper
3. A galvanometer type recorder uses
   (a) taut band
   (b) PMMC movement
   (c) moving iron type
4. A null type recorder uses
   (a) amplifier (b) inductor
   (c) capacitor (d) potentiometer
5. A galvanometer recorder has
   (a) very high input impedance
   (b) high input impedance
   (c) low input impedance
   (d) very low input impedance
6. A circular recorder uses
   (a) rectilinear chart
   (b) curvilinear chart
   (c) square chart
   (d) circular chart
7. A circular chart uses principle of
   (a) electrostatic
   (b) mechanical and link levers
   (c) galvanometer
   (d) self-balancing potentiometer
8. A *X-Y* recorder uses the principle of
   (a) galvanometer
   (b) mechanical levers
   (c) electrostatic
   (d) self-balancing potentiometer
9. An *X-Y* plotter, plots
   (a) one variable with respect to time
   (b) an emf against a physical quantity
   (c) one variable against another variable
10. An *X-Y* plotter uses
    (a) a single pen
    (b) a single pen/arm mechanism
    (c) a double pen
    (d) a double pen/arm mechanism

11. A digital memory waveform recorder records
    (a) slow variation
    (b) transient
    (c) high frequency
    (d) very low variation
12. Pre-triggering is used in digital memory waveform recorder to record
    (a) input signal preceding the trigger point
    (b) to trigger the memory
    (c) to start the operation

Practice Problems

1. The frequency of a signal to be recorded with a strip chart recorder is 15 Hz. What chart speed must be used to record one complete cycle on a 5 mm of recording paper?
2. If the frequency of a signal to be recorded with a strip chart recorder is 30 Hz, what chart speed must be used to record one complete cycle on 5 mm of recording paper?
3. The chart speed of a recording instrument is 10 mm/s. If the time base of the recorded signal is 20 mm, what is the frequency of the recorded signal?
4. The chart speed of a recording instrument is 25 mm/s. If the time base of the recorded signal is 10 mm, what is the frequency of the recorded signal?

Further Reading

1. Austin E. Fribance, *Industrial Instrumentation Fundamentals*, 6th Edition, Tata McGraw-Hill, 1984.
2. Douglas M. Considine, *Process Instruments and Control Handbook*.
3. Jones, Instrument Technology, Vol. 4, *Instrumentation Systems*, B.E. Notlingl, Butterworth, 1987.
4. R.K. Jain, *Mechanical and Industrial Measurements*, Khanna Publishers, 1983.
5. A.K. Sawhney, *Electrical and Electrical Measurements*, Khanna Publishers.
6. Larry D. Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley & Sons, 1987.
7. Oliver Cage, *Electronic Measurements and Instrumentation*, McGraw-Hill, 1975.
8. E.O. Doebelin, *Measurement Systems: Applications and Design*, 4th edition, McGraw-Hill, 1990.

Chapter 13

Transducers

INTRODUCTION 13.1

A transducer is defined as a device that receives energy from one system and transmits it to another, often in a different form.

Broadly defined, the transducer is a device capable of being actuated by an energising input from one or more transmission media and in turn generating a related signal to one or more transmission systems. It provides a usable output in response to a specified input measurand, which may be a physical or mechanical quantity, property, or conditions. The energy transmitted by these systems may be electrical, mechanical or acoustical.

The nature of electrical output from the transducer depends on the basic principle involved in the design. The output may be analog, digital or frequency modulated.

Basically, there are two types of transducers, electrical, and mechanical.

ELECTRICAL TRANSDUCER 13.2

An electrical transducer is a sensing device by which the physical, mechanical or optical quantity to be measured is transformed directly by a suitable mechanism into an electrical voltage/current proportional to the input measurand.

An electrical transducer must have the following parameters:

1. ***Linearity*** The relationship between a physical parameter and the resulting electrical signal must be linear.
2. ***Sensitivity*** This is defined as the electrical output per unit change in the physical parameter (for example V/°C for a temperature sensor). High sensitivity is generally desirable for a transducer.
3. ***Dynamic Range*** The operating range of the transducer should be wide, to permit its use under a wide range of measurement conditions.
4. ***Repeatability*** The input/output relationship for a transducer should be predictable over a long period of time. This ensures reliability of operation.
5. ***Physical Size*** The transducer must have minimal weight and volume, so that its presence in the measurement system does not disturb the existing conditions.

Advantages of Electrical Transducers The main advantages of electrical transducers (conversion of physical quantity into electrical quantities) are as follows:

1. Electrical amplification and attenuation can be easily done.
2. Mass-inertia effects are minimised.
3. Effects of friction are minimised.
4. The output can be indicated and recorded remotely at a distance from the sensing medium.
5. The output can be modified to meet the requirements of the indicating or controlling units. The signal magnitude can be related in terms of the voltage current. (The analog signal information can be converted in to pulse or frequency information. Since output can be modified, modulated or amplified at will, the output signal can be easily used for recording on any suitable multichannel recording device.)
6. The signal can be conditioned or mixed to obtain any combination with outputs of similar transducers or control signals.
7. The electrical or electronic system can be controlled with a very small power level.
8. The electrical output can be easily used, transmitted and processed for the purpose of measurement.

Electrical transducers can be broadly classified into two major categories, (i) Active, (ii) Passive.

An **active transducer** generates an electrical signal directly in response to the physical parameter and does not require an external power source for its operation. Active transducers are self generating devices, which operate under energy conversion principle and generate an equivalent output signal (for example from pressure to charge or temperature to electrical potential).

Typical example of active transducers are piezo electric sensors (for generation of charge corresponding to pressure) and photo voltaic cells (for generation of voltage in response to illumination).

Passive transducers operate under energy controlling principles, which makes it necessary to use an external electrical source with them. They depend upon the change in an electrical parameter (R, L and C).

Typical example are strain gauges (for resistance change in response to pressure), and thermistors (for resistance change corresponding to temperature variations).

Electrical transducers are used mostly to measure non-electrical quantities. For this purpose a detector or sensing element is used, which converts the physical quantity into a displacement. This displacement actuates an electric transducer, which acts as a secondary transducer and give an output that is electrical in nature. This electrical quantity is measured by the standard method used for electrical measurement. The electrical signals may be current, voltage, or frequency; their production is based on R, L and C effects.

A transducer which converts a non-electrical quantity into an analog electrical signal may be considered as consisting of two parts, the sensing element, and the transduction element.

The sensing or detector element is that part of a transducer which responds to a physical phenomenon or to a change in a physical phenomenon. The response of the sensing element must be closely related to the physical phenomenon.

The transduction element transforms the output of a sensing element to an electrical output. This, in a way, acts as a secondary transducer.

Transducers may be further classified into different categories depending upon the principle employed by their transduction elements to convert physical phenomena into output electrical signals.

The different electrical phenomena employed in the transduction elements of transducers are as follows.

1. Resistive
2. Inductive
3. Capacitive
4. Electro magnetic
5. Piezo-electric
6. Photo-emissive
7. Photo-resistive
8. Potentiometric
9. Thermo-electric
10. Frequency generating

SELECTING A TRANSDUCER 13.3

The transducer or sensor has to be physically compatible with its intended application. The following should be considered while selecting a transducer.

1. ***Operating range*** Chosen to maintain range requirements and good resolution.
2. ***Sensitivity*** Chosen to allow sufficient output.
3. ***Frequency response and resonant frequency*** Flat over the entire desired range.
4. ***Environmental compatibility*** Temperature range, corrosive fluids, pressure, shocks, interaction, size and mounting restrictions.
5. ***Minimum sensitivity*** To expected stimulus, other than the measurand.
6. ***Accuracy*** Repeatability and calibration errors as well as errors expected due to sensitivity to other stimuli.
7. ***Usage and ruggedness*** Ruggedness, both of mechanical and electrical intensities versus size and weight.
8. ***Electrical parameters*** Length and type of cable required, signal to noise ratio when combined with amplifiers, and frequency response limitations.

RESISTIVE TRANSDUCER 13.4

Resistive transducers are those in which the resistance changes due to a change in some physical phenomenon. The change in the value of the resistance with a change in the length of the conductor can be used to measure displacement.

Strain gauges work on the principle that the resistance of a conductor or semiconductor changes when strained. This can be used for the measurement of displacement, force and pressure.

The resistivity of materials changes with changes in temperature. This property can be used for the measurement of temperature.

13.4.1 Potentiometer

A resistive potentiometer (pot) consists of a resistance element provided with a sliding contact, called a wiper. The motion of the sliding contact may be translatory or rotational. Some have a combination of both, with resistive elements in the form of a helix, as shown in Fig. 13.1(c). They are known as helipots.

Translatory resistive elements, as shown in Fig. 13.1(a), are linear (straight) devices. Rotational resistive devices are circular and are used for the measurement of angular displacement, as shown in Fig. 13.1(b).

Helical resistive elements are multi turn rotational devices which can be used for the measurement of either translatory or rotational motion. A potentiometer is a passive transducer since it requires an external power source for its operation.

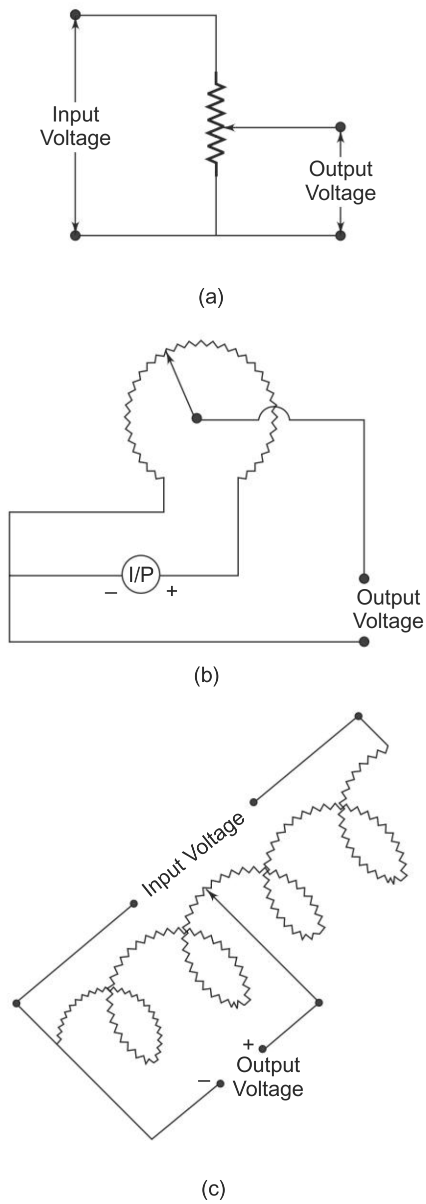


Fig. 13.1 (a) Translatory Type (b) Rotational Type (c) Helipot (Rotational)

Advantage of Potentiometers

1. They are inexpensive.
2. Simple to operate and are very useful for applications where the requirements are not particularly severe.
3. They are useful for the measurement of large amplitudes of displacement.
4. Electrical efficiency is very high, and they provide sufficient output to allow control operations.

Disadvantages

1. When using a linear potentiometer, a large force is required to move the sliding contacts.

2. The sliding contacts can wear out, become misaligned and generate noise.

13.4.2 Resistance Pressure Transducer

Measurement in the resistive type of transducer is based on the fact that a change in pressure results in a resistance change in the sensing elements. Resistance pressure transducers are of two main types. First, the electromechanical resistance transducer, in which a change of pressure, stress, position, displacement or other mechanical variation is applied to a variable resistor. The other resistance transducer is the strain gauge, where the stress acts directly on the resistance. It is very commonly used for stress and displacement measurement in instrumentation.

In the general case of pressure measurement, the sensitive resistance element may take other forms, depending on the mechanical arrangement on which the pressure is caused to act.

Figure 13.1(d) and (e) show two ways by which the pressure acts to influence the sensitive resistance element, i.e. by which pressure varies the resistance element. They are the bellow type, and the diaphragm type. (Yet another is the Bourdon tube of pressure gauge).

In each of these cases, the element moved by the pressure change is made to cause a change in resistance. This resistance change can be made part of a bridge circuit and then taken as either ac or dc output signal to determine the pressure indication.

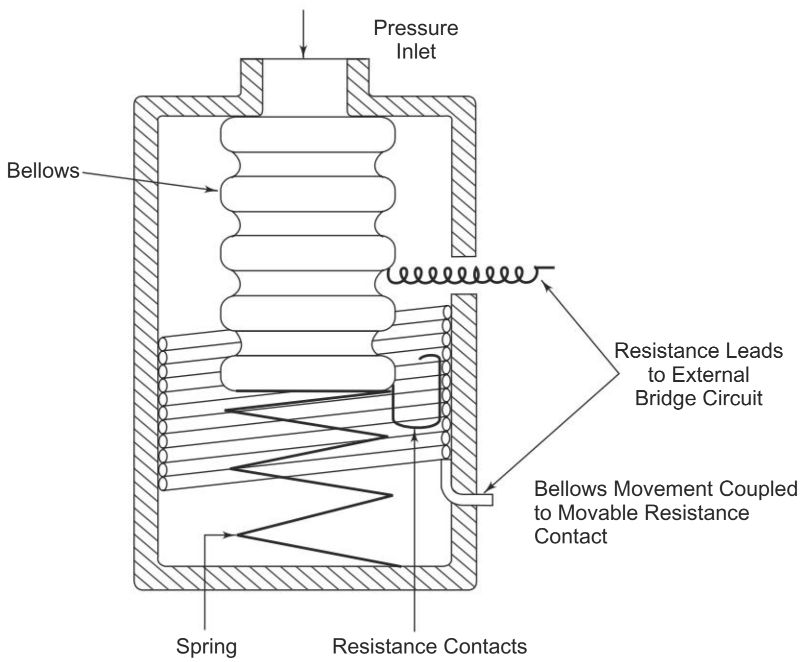


Fig. 13.1(d) Resistance pressure transducer

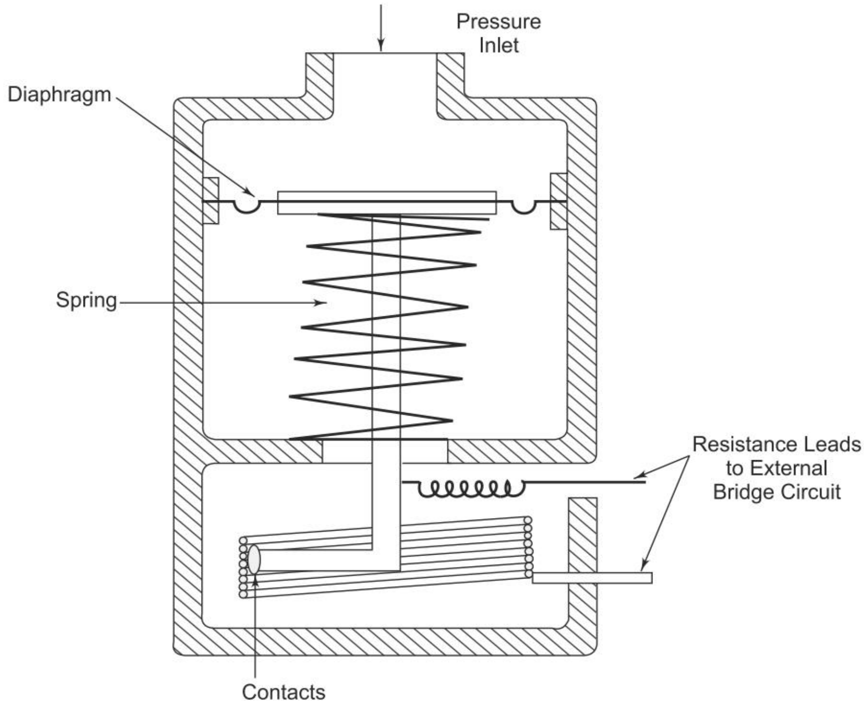


Fig. 13.1(e) Sensitive diaphragm moves the resistance contact

RESISTIVE POSITION TRANSDUCER 13.5

The principle of the resistive transducer is that the physical variable under measurement causes a resistance change in the sensing element. (A common requirement in industrial measurement and control work is to be able to sense the position of an object, or the distance it has moved).

One type of displacement transducer uses a resistive element with a sliding contact or wiper linked to the object being monitored or measured. Thus the resistance between the slider and one end of the resistance element depends on the position of the object. Figure 13.2(a) gives the construction of this type of transducer.

Figure 13.2(b) shows a typical method of use. The output voltage depends on the wiper position and is therefore a function of the shaft position. This voltage may be applied to a voltmeter calibrated in cms for visual display.

(Typical commercial units provide a choice of maximum shaft strokes, from an inch or less to 5 ft or more.) Deviation from linearity of the resistance versus distance specifications can be as low as 0.1 – 1.0%.

Considering Fig. 13.2(b), if the circuit is unloaded, the output voltage V_o is a certain fraction of V_t, depending upon the position of the wiper.

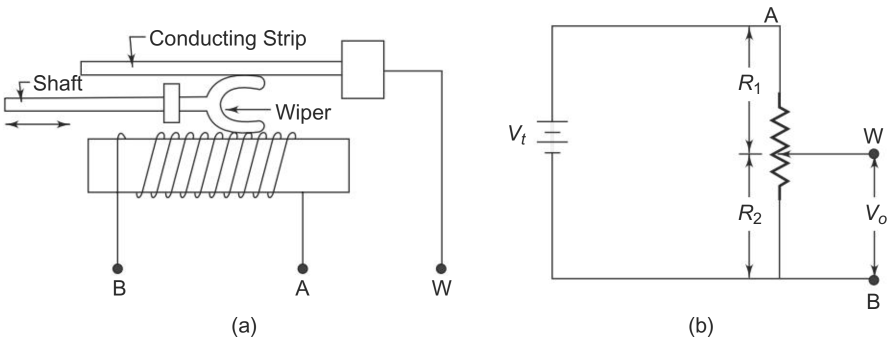


Fig. 13.2 (a) Construction of resistance position transducer (b) Typical method

Therefore,
$$\frac{V_o}{V_t} = \frac{R_2}{R_1 + R_2}$$

When applied to resistive position sensors, this equation shows that output voltage is proportional to R_2, i.e. the position of the wiper of the potentiometer. If the resistance of the transducer is distributed uniformly along the length of travel of the wiper, the resistance is perfectly linear.

Example 13.1 *A displacement transducer with a shaft stroke of 3.0 in. is applied to the circuit of Fig. 13.2(b). The total resistance of the potentiometer is 5 kΩ. The applied voltage V_t is 5 V. When the wiper is 0.9 in. from B, what is the value of the output voltage?*

Solution
$$R_2 = \frac{0.9 \text{ in.}}{3.0 \text{ in.}} \times 5 \text{ k} = \frac{9}{30} \times 5 \text{ k} = 1500\ \Omega$$

Therefore
$$\frac{V_o}{V_t} = \frac{R_2}{R_1 + R_2};\ V_o = \frac{R_2}{R_1 + R_2} \times V_t$$

$$V_o = \frac{1500}{5 \text{ k}} \times 5\ V = \frac{1500}{1 \text{ k}} = 1.5\ V$$

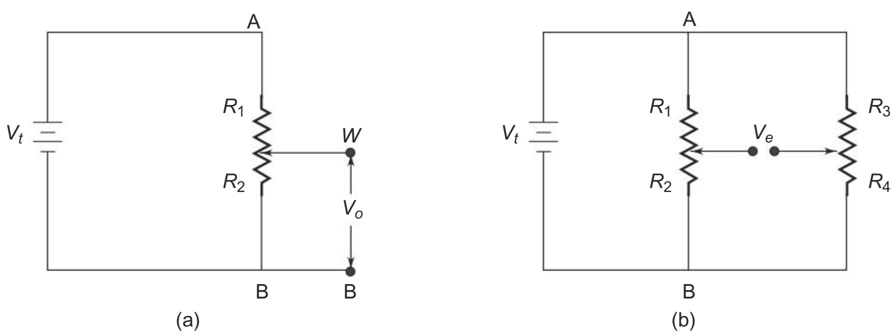


Fig. Ex. 13.1

Example 13.2 *A resistive transducer with a resistance of 5 kΩ and a shaft stroke of 3.0 in. is used in the arrangement in Fig. Ex. 13.1. Potentiometer R_3–R_4 is also 5 k and V_t is 5.0 V. The initial position to be used as a reference point is such that $R_1 = R_2$ (i.e. the shaft is at the centre). At the start of the test, potentiometer R_3–R_4 is adjusted so that the bridge is balanced ($V_e = 0$). Assuming that the object being monitored moves a maximum resistance of 0.5 in. towards A, what will be the new value of V_c? (Shaft distance is 5 in.)*

Solution If the wiper moves 0.5 in. towards A from the centre, it will have moved 3 in. from *B*.

$$R_2 = \frac{3.0}{5.0} \times 5\text{ k} = 3\text{ k}\Omega$$

$$V_e = VR_2 - VR_4 = \left(\frac{R_2}{R_1 + R_2}\right) \times V_t - \left(\frac{R_4}{R_3 + R_4}\right) \times V_t$$

$$= \left(\frac{3\text{ k}}{5\text{ k}}\right) \times 5\ V - \left(\frac{2.5\text{ k}}{5\text{ k}}\right) \times 5\ V$$

$$= 3\ V - 2.5\ V = 0.5\ V$$

STRAIN GAUGES 13.6

The strain gauge is an example of a passive transducer that uses the variation in electrical resistance in wires to sense the strain produced by a force on the wires.

It is well known that stress (force/unit area) and strain (elongation or compression/unit length) in a member or portion of any object under pressure is directly related to the modulus of elasticity.

Since strain can be measured more easily by using variable resistance transducers, it is a common practice to measure strain instead of stress, to serve as an index of pressure. Such transducers are popularly known as strain gauges.

If a metal conductor is stretched or compressed, its resistance changes on account of the fact that both the length and diameter of the conductor changes. Also, there is a change in the value of the resistivity of the conductor when subjected to strain, a property called the *piezo-resistive effect*. Therefore, resistance strain gauges are also known as *piezo resistive gauges*.

Many detectors and transducers, e.g. load cells, torque meters, pressure gauges, temperature sensors, etc. employ strain gauges as secondary transducers.

When a gauge is subjected to a positive stress, its length increases while its area of cross-section decreases. Since the resistance of a conductor is directly proportional to its length and inversely proportional to its area of cross-section, the resistance of the gauge increases with positive strain. The change in resistance value of a conductor under strain is more than for an increase in resistance due to its dimensional changes. This property is called the piezo-resistive effect.

The following types of strain gauges are the most important.

1. Wire strain gauges
2. Foil strain gauges
3. Semiconductor strain gauges

13.6.1 Resistance Wire Gauge

Resistance wire gauges are used in two basic forms, the unbonded type, and the bonded type.

1. Unbonded Resistance Wire Strain Gauge An unbonded strain gauge consists of a wire streched between two points in an insulating medium, such as air. The diameter of the wire used is about 25 μm. The wires are kept under tension so that there is no sag and no free vibration. Unbonded strain gauges are usually connected in a bridge circuit. The bridge is balanced with no load applied as shown in Fig. 13.3.

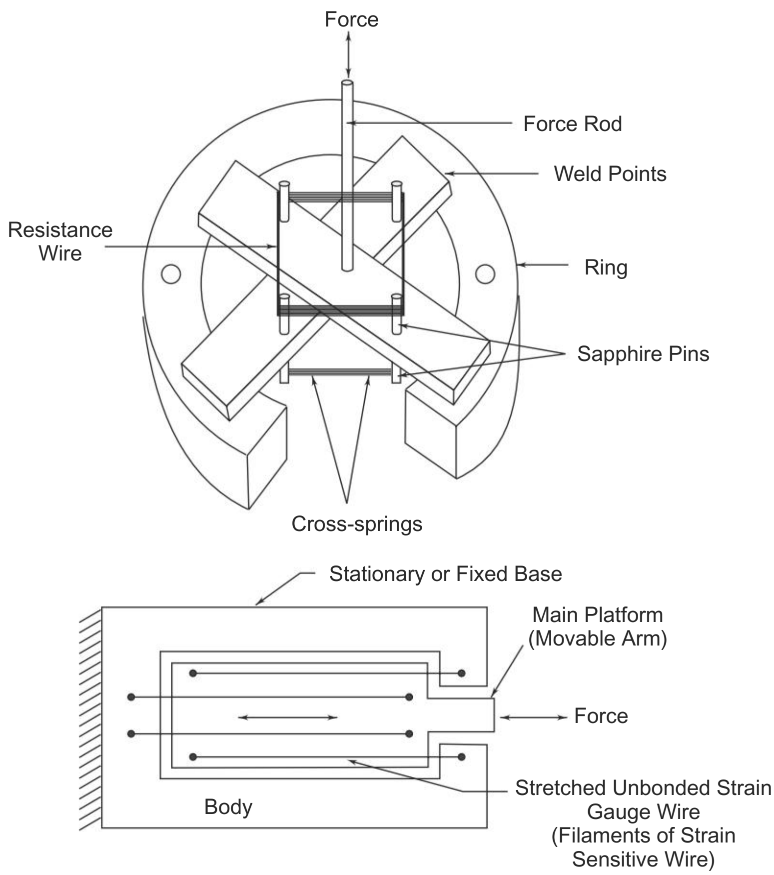


Fig. 13.3 Unbonded strain gauge

When an external load is applied, the resistance of the strain gauge changes, causing an unbalance of the bridge circuit resulting in an output voltage. This voltage is proportional to the strain. A displacement of the order of 50 μm can be detected with these strain gauges.

2. Bonded Resistance Wire Strain Gauges A metallic bonded strain gauge is shown in Fig. 13.4.

A fine wire element about 25 μm (0.025 in.) or less in diameter is looped back and forth on a carrier (base) or mounting plate, which is usually cemented to the member undergoing stress. The grid of fine wire is cemented on a carrier which may be a thin sheet of paper, bakelite, or teflon. The wire is covered on the top with a thin material, so that it is not damaged mechanically. The spreading of the wire permits uniform distribution of stress. The carrier is then bonded or cemented to the member being studied. This permits a good transfer of strain from carrier to wire.

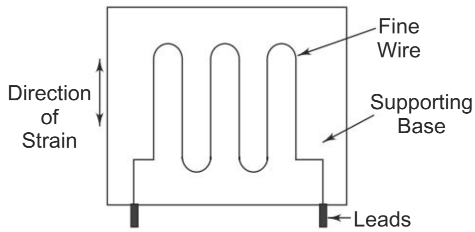


Fig. 13.4 Bonded resistance wire strain gauge

A tensile stress tends to elongate the wire and thereby increase its length and decrease its cross-sectional area. The combined effect is an increase in resistance, as seen from the following equation

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

where ρ = the specific resistance of the material in Ωm
l = the length of the conductor in m
A = the area of the conductor in m^2

As a result of strain, two physical parameters are of particular interest.

1. The change in gauge resistance.
2. The change in length.

The measurement of the sensitivity of a material to strain is called the gauge factor (GF). It is the ratio of the change in resistance $\Delta R/R$ to the change in the length $\Delta l/l$

i.e.
$$\text{GF } (K) = \frac{\Delta R/R}{\Delta l/l} \tag{13.1}$$

where K = gauge factor
ΔR = the change in the initial resistance in Ω's
R = the initial resistance in Ω (without strain)
Δl = the change in the length in m
l = the initial length in m (without strain)

Since strain is defined as the change in length divided by the original length, i.e. $$\sigma = \frac{\Delta l}{l}$$

Eq. (13.1) can be written as

$$K = \frac{\Delta R/R}{\sigma} \tag{13.2}$$

where σ is the strain in the lateral direction.

The resistance of a conductor of uniform cross-section is

$$R = \rho \frac{\text{length}}{\text{area}}$$

$$R = \rho \frac{l}{\pi r^2}$$

Since $$r = \frac{d}{2} \quad \therefore \quad r^2 = \frac{d^2}{4}$$

$$\therefore \quad R = \rho \frac{l}{\pi d^2/4} = \rho \frac{l}{\pi/4\, d^2} \tag{13.3}$$

where ρ = specific resistance of the conductor
l = length of conductor
d = diameter of conductor

When the conductor is stressed, due to the strain, the length of the conductor increases by Δl and the simultaneously decreases by Δd in its diameter. Hence the resistance of the conductor can now be written as

$$R_s = \rho \frac{(l + \Delta l)}{\pi/4(d - \Delta d)^2} = \frac{\rho(l + \Delta l)}{\pi/4\,(d^2 - 2d\,\Delta d + \Delta d^2)}$$

Since Δd is small, Δd^2 can be neglected

$$\therefore \quad R_s = \frac{\rho\,(l + \Delta l)}{\pi/4\,(d^2 - 2d\,\Delta d)}$$

$$= \frac{\rho\,(l + \Delta l)}{\pi/4 d^2 \left(1 - \frac{2\Delta d}{d}\right)} = \frac{\rho l\,(1 + \Delta l/l)}{\pi/4 d^2 \left(1 - \frac{2\Delta d}{d}\right)} \tag{13.4}$$

Now, Poisson's ratio μ is defined as the ratio of strain in the lateral direction to strain in the axial direction, that is,

$$\mu = \frac{\Delta d/d}{\Delta l/l} \tag{13.5}$$

$$\therefore \quad \frac{\Delta d}{d} = \mu \frac{\Delta l}{l} \tag{13.6}$$

Substituting for $\Delta d/d$ from Eq. (13.6) in Eq. (13.4), we have

$$R_s = \frac{\rho l\,(1+\Delta l/l)}{(\pi/4)\,d^2\,(1-2\mu\,\Delta l/l)}$$

Rationalising, we get

$$R_s = \frac{\rho l\,(1+\Delta l/l)}{(\pi/4)\,d^2\,(1-2\mu\,\Delta l/l)}\,\frac{(1+2\mu\,\Delta l/l)}{(1+2\mu\,\Delta l/l)}$$

$$\therefore \qquad R_s = \frac{\rho l}{(\pi/4)\,d^2}\left[\frac{(1+\Delta l/l)}{(1-2\mu\,\Delta l/l)}\,\frac{(1+2\mu\,\Delta l/l)}{(1+2\mu\,\Delta l/l)}\right]$$

$$\therefore \qquad R_s = \frac{\rho l}{(\pi/4)\,d^2}\left[\frac{1+2\mu\,\Delta l/l+2\Delta l/l+2\mu\,\Delta l/e\,\Delta l/l}{1-4\mu^2\,(\Delta l/l)^2}\right]$$

$$\therefore \qquad R_s = \frac{\rho l}{(\pi/4)\,d^2}\left[\frac{1+2\mu\,\Delta l/l+\Delta l/l+2\mu\,\Delta l^2/l^2}{1-4\mu^2\,\Delta l^2/l^2}\right]$$

Since Δl is small, we can neglect higher powers of Δl.

$$\therefore \qquad R_s = \frac{\rho l}{(\pi/4)\,d^2}\,[1+2\,\mu\,\Delta l/l+\Delta l/l]$$

$$R_s = \frac{\rho l}{(\pi/4)\,d^2}\,[1+(2\,\mu+1)\,\Delta l/l]$$

$$R_s = \frac{\rho l}{(\pi/4)\,d^2}\,[1+(1+2\,\mu)\,\Delta l/l]$$

$$\therefore \qquad R_s = \frac{\rho l}{(\pi/4)\,d^2}+\frac{\rho l}{(\pi/4)\,d^2}\,(\Delta l/l)\,(1+2\,\mu)$$

Since from Eq. (13.3), $R = \dfrac{\rho l}{(\pi/4)\,d^2}$

$$\therefore \qquad R_s = R+\Delta R \tag{13.7}$$

where $\Delta R = \dfrac{\rho l}{(\pi/4)\,d^2}\,(\Delta l/l)\,(1+2\,\mu)$

$\therefore$ The gauge factor will now be

$$K = \frac{\Delta R/R}{\Delta l/l} = \frac{(\Delta l/l)\,(1+2\mu)}{\Delta l/l}$$

$$= 1+2\,\mu$$

$$\therefore \qquad K = 1+2\,\mu \tag{13.8}$$

Example 13.3 *A resistance strain gauge with a gauge factor of 2 is cemented to a steel member, which is subjected to a strain of 1×10^{-6}. If the original resistance value of the gauge is 130 Ω, calculate the change in resistance.*

Solution Given:

$$K = \frac{\Delta R/R}{\Delta l/l}$$

Therefore, $\Delta R = K\,R\,\Delta l/l$

$$\Delta R = 2 \times 130 \times 1 \times 10^{-6} = 260\ \mu\Omega$$

The initial resistance value R of a strain gauge is typically around 120 Ω and the gauge factor may be from (for Nickel) – 12 to + 6. A gauge factor of 2 is reasonable for most strain gauges. Semiconductor gauge have higher sensitivities.

The strain gauge is normally used in a bridge arrangement in which the gauge forms one arm of the bridge. The bridge may be ac or dc actuated. A simple dc arrangement is shown in Fig. 13.5. Only one of the gauges is an active element, producing an output proportional to the strain. The other (dummy) gauge is not strained, but simply balances the bridge (compensation). Since the resistance of the fine wire element is sensitive to temperature as well as stress variation, any change in temperature will cause a change in the bridge balance conditions. This effect can cause error in the strain measurement (thereby affecting the accuracy). Hence, when temperature variations are significant, or when unusual accuracy is required, some compensation must be used. The dummy gauge accomplishes this, because it is placed in the same temperature environment as the active gauges, but not subjected to strain. Consequently, the temperature causes the same change of resistance in the two strain gauges and the bridge balance is not affected by the temperature.

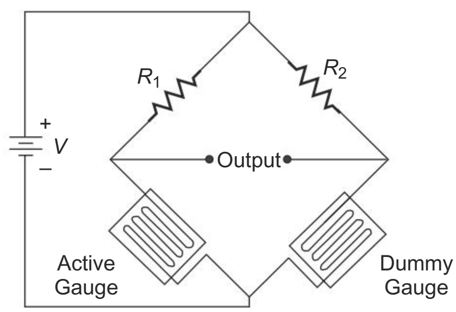


Fig. 13.5 Strain gauge used in bridge arrangement

If the two resistors R_1 and R_2 have negligible temperature coefficients, the bridge retains its balance under conditions of no-strain, at any temperature within its operating range.

(However, one of the two gauges in mounted so that its sensitivity direction is at right angles to the direction of strain.)

The resistance of this dummy gauge is not affected by the deformation of the material and it therefore acts like a passive resistance, with regard to strain measurement.

Since only one gauge responds to the strain, the strain causes bridge unbalance just as in the case of a single gauge.

13.6.2 Types of Strain Gauges (Wire)

1. Grid type
2. Rossette type
3. Torque type
4. Helical type

The *grid arrangement* of the wire element in a bonded strain gauge creates a problem not encountered in the use of unbonded strain gauges. To be useful as a strain gauge, the wire element must measure strain along one axis. Therefore complete and accurate analysis of strain in a rigid member is impossible, unless the direction and magnitude of stress are known. The measuring axis of a strain gauge is its longitudinal axis, which is parallel to the wire element, as shown in Fig. 13.6.

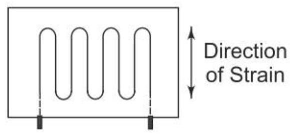


Fig. 13.6 Grid type strain gauge

When a strain occurs in the member being measured, along the transverse axis of the gauge, it also affects the strain being measured parallel to the longitudinal axis. This introduces an error in the response of the gauge.

In most applications, some degree of strain is present along the transverse axis and the transverse sensitivity must be considered in the final gauge output. Transverse sensitivity cannot be completely eliminated, and in highly accurate measurements the resultant gauge error must be compensated for.

If the axis of the strain in a component is unknown, strain gauges may be used to determine the exact direction. The standard procedure is to place several gauges at a point on the member's surface, with known angles between them. The magnitude of strain in each individual gauge is measured, and used in the geometrical determination of the strain in the member. Figure 13.7 shows a three-element strain gauge, called a *Rossette gauge*, in which the angle between any two longitudinal gauge axes is 45°.

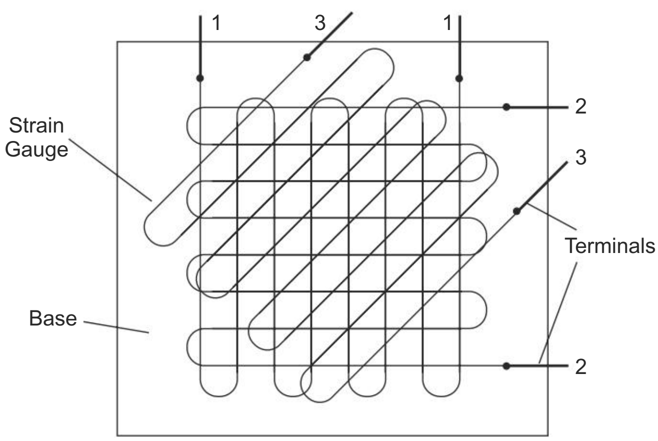


Fig. 13.7 Rossette gauge

The 45° Rossette gauge is, in general, the most popular one. There are different shapes and sizes of strain gauges for various purposes.

Serving a similar, but not specialised, purpose are gauges with specially modified grid configurations, such as those shown in Figs. 13.8(a), (b) and (c).

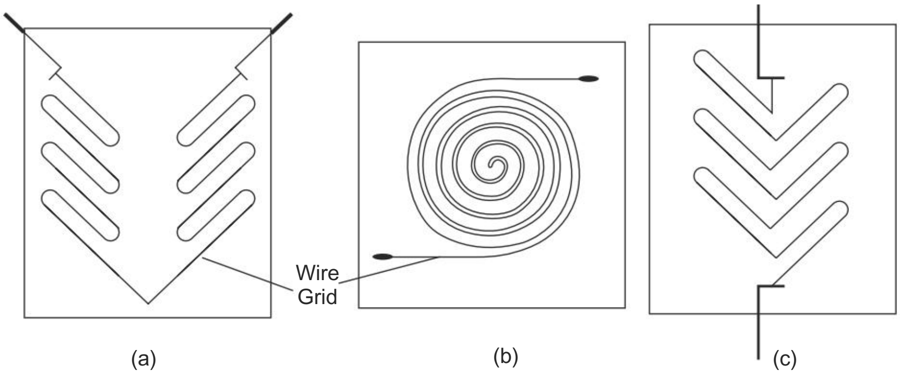


Fig. 13.8 (a) and (c) Torque type gauge (b) Helical gauge

A measurement of this type would be useful at the cross-point of an X-shaped frame.

The latest in strain gauges is the etched foil strain gauge. This device uses the technique of PCB design. Its physical and electrical characteristics are superior to bonded wire strain gauges in almost every respect.

The size of strain gauges varies with application. They can be as small as 3 sq.mm. Usually they are larger, but not more than 2.5 mm long and 12.5 mm wide.

To obtain good results, it is desirable that a resistance wire strain gauge have the following characteristics.

1. The strain gauge should have a high value of gauge factor (a high value of gauge factor indicates a large change in resistance for particular strain, implying high sensitivity).
2. The resistance of the strain gauge should be as high as possible, since this minimises the effects of undesirable variations of resistance in the measurement circuit. Typical resistances of strain gauges are 120 Ω, 350 Ω and 1000 Ω.
   A high resistance value results in lower sensitivity. Hence, in order to get high sensitivity, higher bridge voltages have to be used. The bridge voltage is limited by the maximum current carrying capacity of the wires, which is typically 30 mA.
3. The strain gauge should have a low resistance temperature coefficient. This is necessary to minimise errors on account of temperature variation, which affects the accuracy of measurements. (Temperature compensation is also used.).
4. The strain gauge should not have hysteresis effects in its response.

5. In order to maintain constancy of calibration over the entire range of the strain gauge, it should have linear characteristics, i.e. the variation in resistance should be a linear function of the strain.
6. Strain gauges are frequently used for dynamic measurements and hence their frequency response should be good. Linearity should be maintained within specified accuracy limits over the entire frequency range.
7. Leads used must be of materials which have low and stable resistivity and low resistance temperature coefficient.

13.6.3 Foil Strain Gauge

This class of strain gauges is an extension of the resistance wire strain gauge. The strain is sensed with the help of a metal foil. The metals and alloys used for the foil and wire are nichrome, constantan (Ni + Cu), isoelastic (Ni + Cr + Mo), nickel and platinum.

Foil gauges have a much greater dissipation capacity than wire wound gauges, on account of their larger surface area for the same volume. For this reason, they can be used for a higher operating temperature range. Also, the large surface area of foil gauges leads to better bonding.

Foil type strain gauges have similar characteristics to wire strain gauges. Their gauge factors are typically the same.

The advantage of foil type strain gauges is that they can be fabricated on a large scale, and in any shape. The foil can also be etched on a carrier.

Etched foil gauge construction consists of first bonding a layer of strain sensitive material to a thin sheet of paper or bakelite. The portion of the metal to be used as the wire element is covered with appropriate masking material, and an etching solution is applied to the unit. The solution removes that portion of the metal which is not masked, leaving the desired grid structure intact.

This method of construction enables etched foil strain gauges to be made thinner than comparable wire units, as shown in Fig. 13.9. This characteristics, together with a greater degree of flexibility, allows the etched foil to be mounted in more remote and restricted places and on a wide range of curved surfaces.

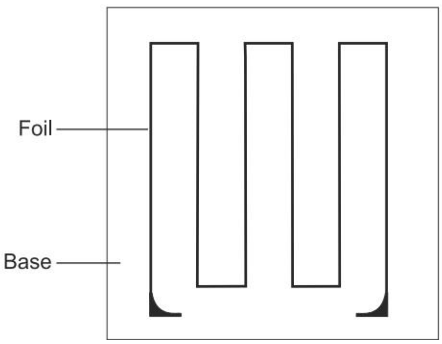


Fig. 13.9 Foil type strain gauge

The longitudinal sensitivity of the foil gauge is approximately 5% greater than that of similar wire elements. The transverse strain sensitivity of this gauge is smaller 1/3 to 1/2 of similar wire gauges. The hysteresis of the foil gauge is also 1/3 to 1/2 of a wire strain gauge.

(The term hysteresis, as used in strain gauges, is defined as follows. If the resistance of a strain gauge is measured with no strain applied, and the gauge is then stressed to its maximum usable resistance value, the measured resistance, after the stress is removed, differs from the original value. The inability of the gauge element to resume the exact physical form it had before being elongated, produces the difference in resistance. This effect is called *hysteresis*.)

The resistance film formed is typically 0.2 mm thick. The resistance value of commercially available foil gauges is between 50 and 1000 Ω. The resistance films are vacuum coated with ceramic film and deposited on a plastic backing for insulation.

13.6.4 Semiconductor Strain Gauge

To have a high sensitivity, a high value of gauge factor is desirable. A high gauge factor means relatively higher change in resistance, which can be easily measured with a good degree of accuracy.

Semiconductor strain gauges are used when a very high gauge factor is required. They have a gauge factor 50 times as high as wire strain gauges. The resistance of the semiconductor changes with change in applied strain.

Semiconductor strain gauges depend for their action upon the piezo resistive effect, i.e. change in value of the resistance due to change in resistivity, unlike metallic gauges where change in resistance is mainly due to the change in dimension when strained. Semiconductor materials such as germanium and silicon are used as resistive materials.

A typical strain gauge consists of a strain material and leads that are placed in a protective box, as shown in Fig. 13.10. Semiconductor wafer or filaments which have a thickness of 0.05 mm are used. They are bonded on suitable insulating substrates, such as teflon.

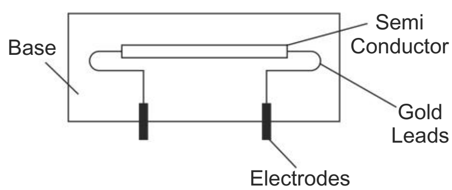


Fig. 13.10 Semiconductor strain gauge

Gold leads are generally used for making contacts. These strain gauges can be fabricated along with an IC Op Amp which can act as a pressure sensitive transducer. The large gauge factor is accompanied by a thermal rate of change of resistance approximately 50 times higher than that for resistive gauges. Hence, a semiconductor strain gauge is as stable as the metallic type, but has a much higher output.

Simple temperature compensation methods can be applied to semiconductor strain gauges, so that small values of strain, that is micro strains, can also be measured.

The gauge factor of this type of semiconductor strain gauge is 130 ± 10% for a unit of 350 Ω, 1” long, 1/2” wide and 0.005” thick. The gauge factor is determined at room temperature at a tensile strain level of 1000 micro strain (1000 micro in/in. of length). The maximum operating tensile strain is ± 3000 micro strain, with a power dissipation of 0.1 W. The semiconductor strain gauge also has low hysteresis and is susceptible to regular methods of temperature compensation. The semiconductor strain gauge has proved itself to be a stable and practical device for operation with conventional indicating and recording systems, to measure small strains from 0.1–500 micro strain.

Advantages of Semiconductor Strain Gauge

1. Semiconductor strain gauges have a high gauge factor of about + 130. This allows measurement of very small strains, of the order of 0.01 micro strain.
2. Hysteresis characteristics of semiconductor strain gauges are excellent, i.e. less than 0.05%.
3. Life in excess of 10×10^6 operations and a frequency response of 10^{12} Hz.
4. Semiconductor strain gauges can be very small in size, ranging in length from 0.7 to 7.0 mm.

Disadvantages

1. They are very sensitive to changes in temperature.
2. Linearity of semiconductor strain gauges is poor.
3. They are more expensive.

RESISTANCE THERMOMETER* 13.7

The resistance of a conductor changes when its temperature is changed. This property is utilised for the measurement of temperature. The resistance thermometer is an instrument used to measure electrical resistance in terms of temperature, i.e. it uses the change in the electrical resistance of the conductor to determine the temperature.

The main part of a resistance thermometer is its sensing element. The characteristics of the sensing element determines the sensitivity and operating temperature range of the instrument.

(There are three common types of temperature sensitive resistive elements in use, the wire wound resistance, the thermistor and the PTC semiconductor resistance.)

The sensing element may be any material that exhibits a relatively large resistance change with change in temperature. Also, the material used should be stable in its characteristics, i.e. neither its resistance nor its temperature coefficient of resistance should undergo permanent change with use or age.

To maintain the calibration of a resistance thermometer, it is necessary to consider its stability. The need for stability frequently limits the temperature range over which the sensing element may be used.

Another desirable characteristic for a sensing element is a linear change in resistance with change in temperature.

The speed with which a resistive element responds to changes in temperature is important when the measured temperature is subjected to rapid variations. The

*Also refer Sec. 13.20.2 (RTD's

smaller a given sensing element, the less heat required to raise its temperature, and the faster its response.

Platinum, nickel and copper are the metals most commonly used to measure temperature. The resistivity of platinum tends to increase less rapidly at higher temperatures than for other metals, hence it is a commonly used material for resistance thermometers. The temperature range over which platinum has stability is – 260–1100°C.

Figure 13.11(a) shows an industrial platinum resistance thermometer.

The changes in resistance caused by changes in temperature are detected by a Wheatstone bridge, as shown in Fig. 13.11(b). Hence, the temperature sensing element, which may be nickel, copper or platinum contained in a bulb or well, along with the balancing bridge, form the essential components of a temperature measuring system based upon this principle.

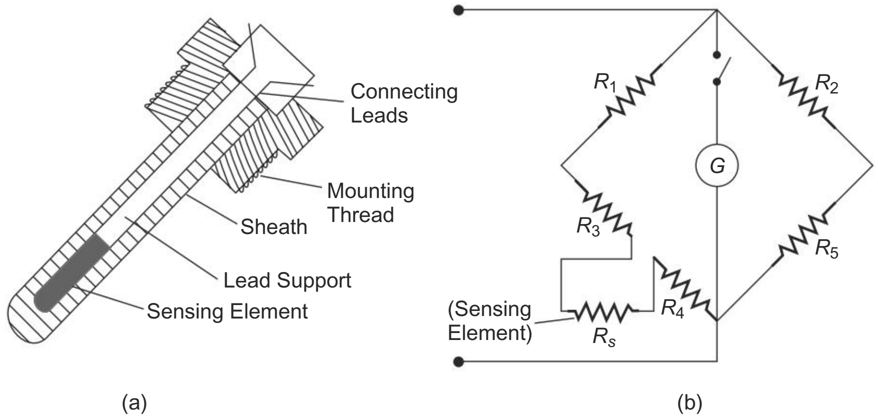


Fig. 13.11 (a) Industrial platinum resistance thermometer (b) Bridge circuit

The sensing element R_s is made of a material having a high temperature coefficient, and R_1, R_2, and R_5 are made of resistances that are practically constant under normal temperature changes.

When no current flows through the galvanometer, the normal principle of Wheatstone's bridge states the ratio of resistance is

$$\frac{R_1}{R_2} = \frac{R_s}{R_5}$$

In normal practice, the sensing element is away from the indicator, and its leads have a resistance, say R_3, R_4.

Therefore,

$$\frac{R_1}{R_2} = \frac{R_3 + R_s + R_4}{R_5}$$

Now it resistance R_s changes, balance cannot be maintained and the galvanometer shows a deflection, which can be calibrated to give a suitable temperature scale.

Advantages of Resistance Thermometers The measurement of temperature by the electrical resistance method has the following advantages and characteristics.

1. The measurement is very accurate.
2. It has a lot of flexibility with regard to choice of measuring equipment.
3. Indicators, recorders or controllers can also be operated.
4. More than one resistance element can be clubbed to the same indicating/recording instrument.
5. The temperature sensitive resistance element can be easily installed and replaced.
6. The accuracy of the measuring circuit can be easily checked by substituting a standard resistor for the resistive element.
7. Resistive elements can be used to measure differential temperature.
8. Resistance thermometers have a wide working range without loss of accuracy, and can be used for temperature ranges (–200°C to + 650°C).
9. They are best suited for remote indication.
10. The resistive element response time is of the order of 2 to 10s
11. The limits of error of a resistive element are ± 0.25% of the scale reading.
12. The size of the resistive element may be about 6 – 12 mm in diameter and 12 – 75 mm in length.
13. Extremely accurate temperature sensing.
14. No necessity of temperature compensation.
15. Stability of performance over long periods of time.

Limitations of Resistance Thermometer

1. High cost
2. Need for bridge circuit and power source
3. Possibility of self-heating
4. Large bulb size, compared to a thermocouple

THERMISTOR 13.8

The electrical resistance of most materials changes with temperature. By selecting materials that are very temperature sensitive, devices that are useful in temperature control circuits and for temperature measurements can be made.

Thermistor (THERMally sensitive resISTOR) are non-metallic resistors (semiconductor material), made by sintering mixtures of metallic oxides such as manganese, nickel, cobalt, copper and uranium.

Thermistors have a Negative Temperature Coefficient (NTC), i.e. resistance decreases as temperature rises. Figure 13.12 shows a graph of resistance vs temperature for a thermistor. The resistance at room temperature (25°C) for typical commercial units ranges from 100 Ω to 10 MΩ. They are suitable for use only up to about 800°C. In some cases, the resistance of thermistors at room temperature may decrease by 5% for each 1°C rise in temperature. This high sensitivity to temperature changes makes the thermistor extremely useful for precision temperature measurements, control and compensation.

The smallest thermistors are made in the form of beads. Some are as small as 0.15 mm (0.006 in.) in diameter. These may come in a glass coating or sealed in the tip of solid glass probes. Glass probes have a diameter of about 2.5 mm and a length which varies from 6 – 50 mm. The probes are used for measuring the temperature of liquids. The resistance ranges from 300 Ω to 100 MΩ.

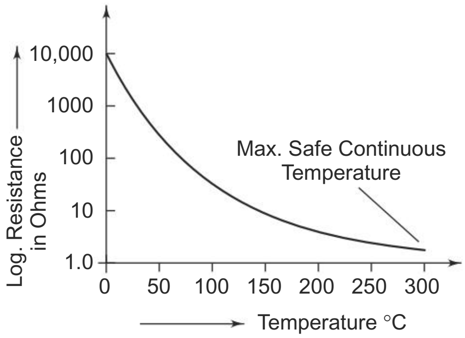


Fig. 13.12 Resistance vs temperature graph of a thermistor

Where greater power dissipations is required, thermistors may be obtained in disc, washer or rod forms.

Disc thermistors about 10 mm in diameter, either self supporting or mounted on a small plate, are mainly used for temperature control. These thermistors are made by pressing themistors material under several tons of pressure in a round die to produce flat pieces 1.25 – 25 mm in diameter and 0.25 – 0.75 mm thick, having resistance values of 1 Ω to 1 MΩ. These are sintered and coated with silver on two flat surfaces.

Washer thermistors are made like disc thermistors, except that a hole is formed in the centre in order to make them suitable for mounting on a bolt. Rod thermistors are extruded through dies to make long cylindrical units of 1.25, 2.75, and 4.25 mm in diameter and 12.5 – 50 mm long. Leads are attached to the end of the rods. Their resistance usually varies from 1 – 50 kΩ.

The advantage of rod thermistors over other configurations is the ability to produce high resistance units with moderately high power handling capability.

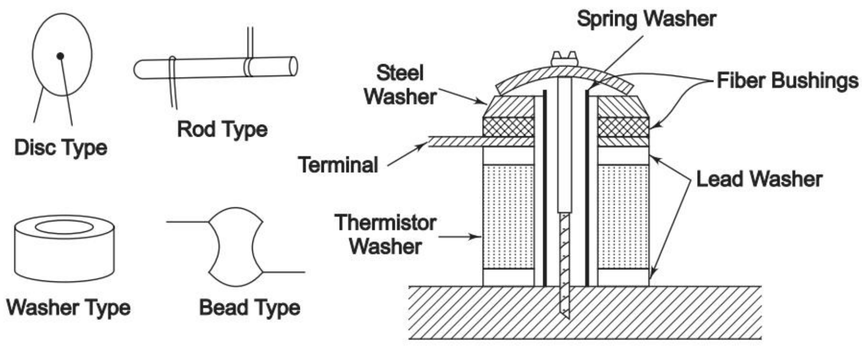
Thermistors can be connected in series/parallel combinations for applications requiring increased power handling capability. High resistance units find application in measurements that employ low lead wires or cables.

Thermistors are chemically stable and can be used in nuclear environments. Their wide range of characteristics also permits them to be used in limiting and regulation circuits, as time delays, for integration of power pulses, and as memory units.

Typical thermistor configurations are as shown in Fig. 13.13(a). Figure 13.13(b) shows a bush type thermistor.

A thermistor in one arm of a Wheatstone bridge provides precise temperature information. Accuracy is limited, in most applications, only by the readout devices.

Thermistors are non-linear devices over a temperature range, although now units with better than 0.2% linearity over the 0–100°C temperature range are available. The typical sensitivity of a thermistor is approximately 3 mV/°C at 200°C.



(a) (b)

Fig. 13.13 (a) Various configurations of thermistor (b) Bush-type thermistor

Advantages of Thermistor

1. Small size and low cost.
2. Fast response over narrow temperature range.
3. Good sensitivity in the NTC region.
4. Cold junction compensation not required due to dependence of resistance on absolute temperature.
5. Contact and lead resistance problems not encountered due to large R_{th} (resistance).

Limitations of Thermistor

1. Non-linearity in resistance vs temperature characteristics.
2. Unsuitable for wide temperature range.
3. Very low excitation current to avoid self-heating.
4. Need of shielded power lines, filters, etc. due to high resistance.

Example 13.4 *The circuit of Fig. Ex. 13.2 (a) is to be used for temperature measurement. The thermistor is a 4 kΩ type. The meter is a 50 mA meter with a resistance of 3 Ω, R_c is set to 17 Ω, and supply V_t is 15 V. What will be the meter reading at 77°F (25°C) and at 150°F.*

Solution From the graph of temperature versus resistance, the resistance at 25°C is 4 kΩ. Therefore the current at 25°C is

$$I = \frac{V_t}{R_t} = \frac{15}{4000 + 17 + 3} = \frac{15}{4020} = 3.73 \text{ mA.}$$

At 150°F, the graph shows that the thermistor resistance is approximately 950 Ω. The meter reading will then be

$$I = \frac{V_t}{R_t} = \frac{15}{950 + 17 + 3} = \frac{15}{970} = 15.5 \text{ mA.}$$

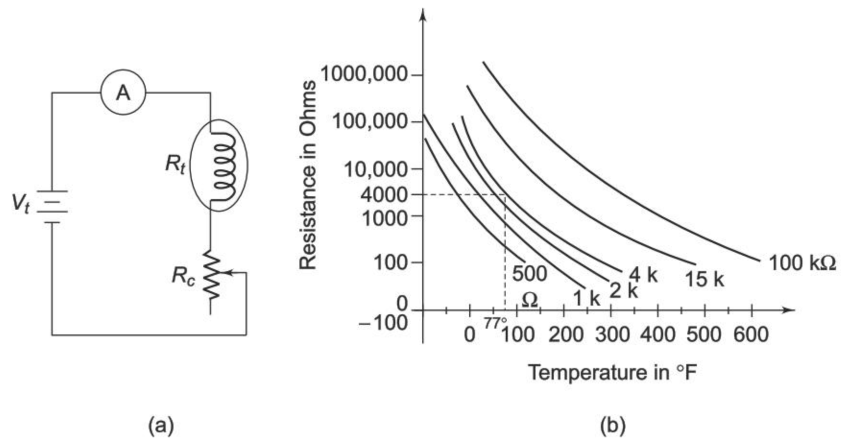


Fig. Ex. 13.2

INDUCTIVE TRANSDUCER 13.9

Inductive transducers may be either of the self generating or the passive type. The self generating type utilises the basic electrical generator principle, i.e. a motion between a conductor and magnetic field induces a voltage in the conductor (generator action). This relative motion between the field and the conductor is supplied by changes in the measured.

An inductive electromechanical transducer is a device that converts physical motion (position change) into a change in inductance.

Transducers of the variable inductance type work upon one of the following principles.

1. Variation of self inductance
2. Variation of mutual inductance

Inductive transducers are mainly used for the measurement of displacement. The displacement to be measured is arranged to cause variation in any of three variables

1. Number of turns
2. Geometric configuration
3. Permeability of the magnetic material or magnetic circuits

For example, let us consider the case of a general inductive transducer. The inductive transducer has N turns and a reluctance R. When a current i is passed through it, the flux is

$$\phi = \frac{Ni}{R}$$

Therefore $$\frac{d\phi}{dt} = \frac{N}{2} \times \frac{di}{dt} - \frac{Ni}{R^2} \times \frac{dR}{dt}$$

If the current varies very rapidly,

$$\frac{d\phi}{dt} = \frac{N}{2} \times \frac{di}{dt}$$

But emf induced in the coil is given by $e = N \times d\phi/dt$

Therefore $$e = N \times \frac{N}{2} \times \frac{di}{dt} = \frac{N^2}{R} \times \frac{di}{dt} \tag{13.9}$$

Also the self inductance is given by

$$L = \frac{e}{di/dt} = \frac{N^2}{R} \tag{13.10}$$

Therefore, the output from an inductive transducer can be in the form of either a change in voltage or a change in inductance.

13.9.1 Change in Self Inductance with Numbers of Turns

The output may be caused by a change in the number of turns.

Figures 13.14(a) and (b) are transducers used for the measurement of displacement of linear and angular movement respectively.

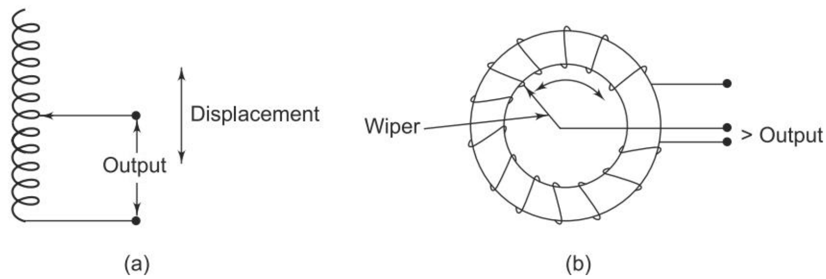


Fig. 13.14 (a) Linear inductive transducer (using air core)
(b) Angular inductive transducer (using ferrite core)

Figure 13.14(a) is an air cored transducer for measurement of linear displacement.

Figure 13.14(b) is an iron cored coil used for the measurement of angular displacement.

In both cases, as the number of turns are changed, the self inductance and the output also changes.

13.9.2 Transducer Working on the Principle of Change in Self Inductance with Change in Permeability

Figure 13.15 shows an inductive transducer which works on the principle of the variation of permeability causing a change in self inductance. The iron core is surrounded by a winding. If the iron core is inside the winding, its permeability is increased, and so is the inductance. When the iron core is moved out of the winding, the permeability decreases, resulting in a reduction of the self inductance of the coil. This transducer can be used for measuring displacement.

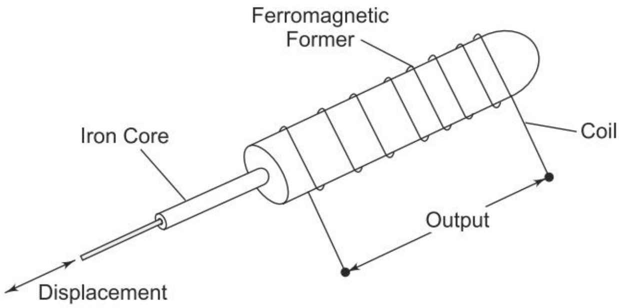


Fig. 13.15 Inductive transducer working on the principle of variation of permeability

13.9.3 Variable Reluctance Type Transducer

A transducer of the variable type consists of a coil wound on a ferromagnetic core. The displacement which is to be measured is applied to a ferromagnetic target. The target does not have any physical contact with the core on which it is mounted. The core and the target are separated by an air gap, as shown in Fig. 13.16(a).

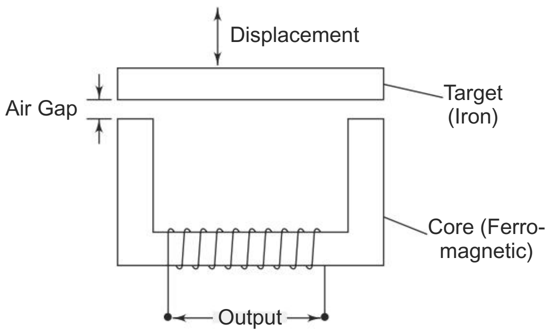


Fig. 13.16 (a) Variable reluctance transducer

The reluctance of the magnetic path is determined by the size of the air gap. The inductance of the coil depends upon the reluctance of the magnetic circuits.

The self inductance of the coil is given by

$$L = \frac{N^2}{R_i + R_g} \tag{13.11}$$

where N = number of turns
R_i = reluctance of iron parts
R_g = reluctance of air gap

The reluctance of the iron part is negligible compared to that of the air gap.

Therefore $$L = N^2/R_g \tag{13.12}$$

But reluctance of the air gap is given by

$$R_g = \frac{l_g}{\mu_o \times A_g} \tag{13.13}$$

where l_g = length of the air gap
A_g = area of the flux path through air
μ_o = permeability

R_g is proportional to l_g, as μ_o and A_g are constants.

Hence L is proportional to $1/lg$, i.e. the self inductance of the coil is inversely proportional to the length of the air gap.

When the target is near the core, the length is small and therefore the self inductance large. But when the target is away from the core the reluctance is large, resulting in a smaller self inductance value. Hence the inductance of the coil is a function of the distance of the target from the core, i.e. the length of the air gap.

Since it is the displacement which changes the length of the air gap, the self inductance is a function of displacement, albeit a non-linear one.

A variable reluctance bridge is shown in Fig. 13.16(b).

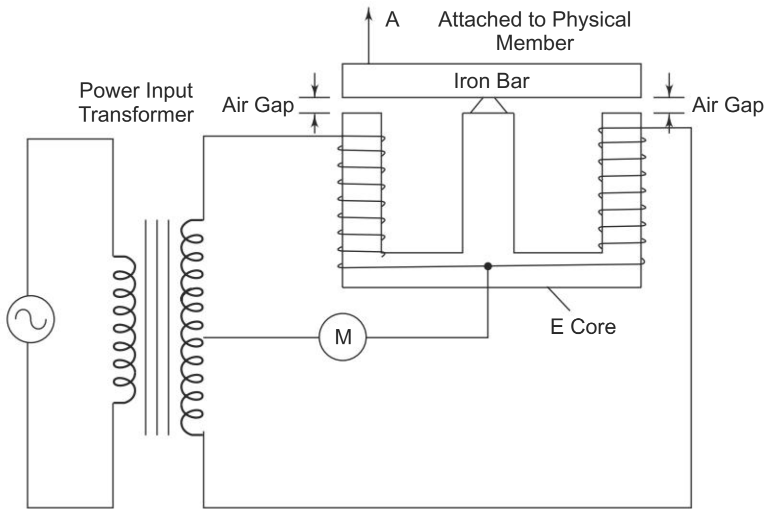


Fig. 13.16 (b) Variable reluctance bridge circuit

A separate coil is wound on each outside leg of an E core and an iron bar is pivoted on the centre leg. A magnet extends from each outside leg through an air gap and through the iron bar to the centre leg.

The moving member is attached to one end of the iron bar and causes the bar to wobble back and forth, thereby varying the size of each air gap.

The bridge consists of two transducer coils and a tapped secondary of the input power transformers. It is balanced only when the inductance of the two transducer coils are equal, i.e. when the iron bar is in a nearly exact horizontal position and the air gaps are equal.

Whenever the iron bar at point A moves and alters the air gap, the bridge becomes unbalanced by an amount proportional to the change in inductance, which in turn is proportional to the displacement of the moving member.

The increase and decrease of the inductance with varying air gap sizes is non-linear, and so is the output. Also, the flux density within the air gaps is easily affected by external fields.

Example 13.5 *A variable reluctance type inductive transducer has a coil of inductance of 2500 μH When the target made of ferromagnetic material is 1mm away from the core. Calculate the value of inductance when a displacement of 0.04 mm is applied to the target in a direction moving it towards the core .*

Solution Given inductance with gap length of 1 mm is $L = 2500$ μH

Step 1: Length of air gap when a displacement is applied to the target

$$= 1.00 - 0.04 = 0.96 \text{ mm}$$

Step 2: Now inductance is inversely proportional to the length of air gap

Therefore 'L' with gap length of 0.96 mm

$$= L + \Delta L = 2500\ \mu\text{H} \times \frac{1}{0.96\ \text{mm}} = 2604\ \mu\text{H}$$

Step 3: Therefore, change in inductance

$$\Delta L = 2604\ \mu\text{H} - 2500\ \mu\text{H} = 104\ \mu\text{H}$$

DIFFERENTIAL OUTPUT TRANSDUCERS 13.10

The differential output transducer consists of a coil which is divided into two parts, as shown in Figs. 13.17(a) and (b).

(Inductive transducers using self inductance as a variable use one coil, while those using mutual inductance as a variable use multiple coils.)

Normally the change in self inductance, ΔL, for inductive transducers, (working on the principle of change of self inductance) is not sufficient for detection of subsequent stages of the instrumentation system.

However, if successive stages of the instrument respond to ΔL or ΔM, rather than $L + \Delta L$, or $M + \Delta M$, the sensitivity and accuracy will be much higher.

The transducers can be designed to provide two outputs, one of which represents inductance (self or mutual) and the other the decrease in inductance (self or mutual). The succeeding stages of the instrumentation system measure the difference between these outputs. This is known as differential output.

Advantages of Differential Output

1. Sensitivity and accuracy are increased.
2. Output is less affected by external magnetic fields.
3. Effective variations due to temperature changes are reduced.
4. Effects of change in supply voltages and frequency are reduced.

In response to a physical signal (which is normally displacement), the inductance of one part increases from L to $L + \Delta L$, while that of the other part decreases from L to $L - \Delta L$. The change is measured as the difference of the two, resulting in an output of $2\ \Delta L$ instead of ΔL, when one winding is used. This increases the sensitivity and also eliminates error.

Inductive transducers using the change in the number of turns to cause a change in the self inductance are shown in Fig. 13.17.

Figure 13.17(a) is used for measurement of linear displacement using an air cored coil.

Figure 13.17(b) is used for the measurement of angular displacement using an iron cored coil.

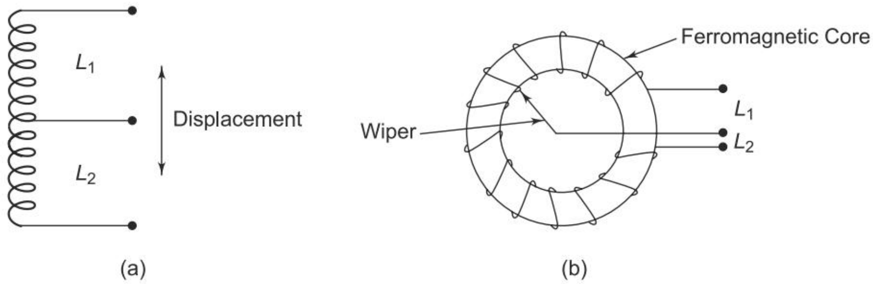


Fig. 13.17 (a) Linear differential output transducer (b) Angular differential output transducer

Figure 13.18 shows an inductive transducer giving a differential output. The output represents a change of self inductance due to change of reluctance. (This inductive transducer also works on the principle of change of self inductance of the two coils with change in reluctance of the path of the magnetic circuit. The target as well as cores on which the coil is wound are made up of iron.)

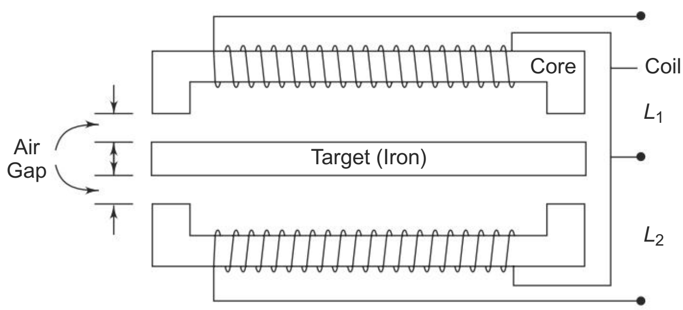


Fig. 13.18 Inductive transducer differential output (Reluctance principle)

LINEAR VARIABLE DIFFERENTIAL TRANSDUCER (LVDT) 13.11

The differential transformer is a passive inductive transformer. It is also known as a Linear Variable Differential Transformer (LVDT). The basic construction is as shown in Fig. 13.19.

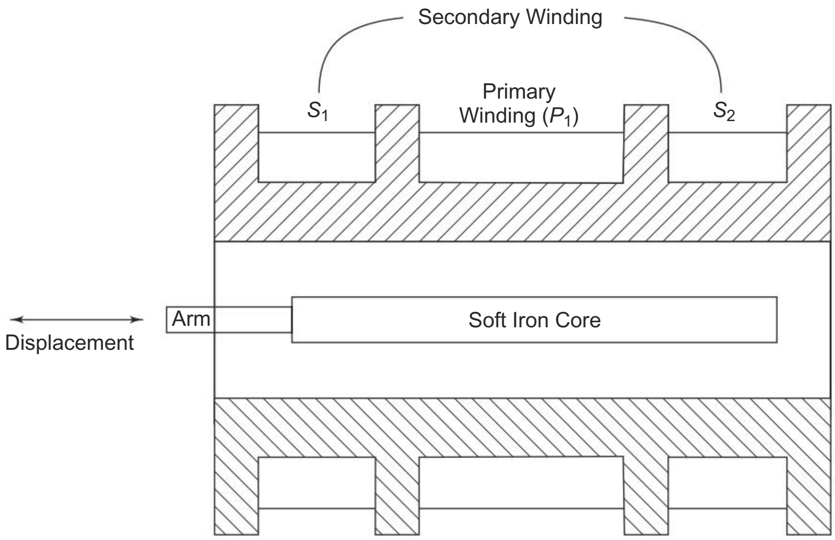


Fig. 13.19 Construction of a linear variable differential transducer (LVDT)

The transformer consists of a single primary winding P_1 and two secondary windings S_1 and S_2 wound on a hollow cylindrical former. The secondary windings have an equal number of turns and are identically placed on either side of the primary windings. The primary winding is connected to an ac source.

An movable soft iron core slides within the hollow former and therefore affects the magnetic coupling between the primary and the two secondaries.

The displacement to be measured is applied to an arm attached to the soft iron core.

(In practice, the core is made up of a nickel-iron alloy which is slotted longitudinally to reduce eddy current losses.)

When the core is in its normal (null) position, equal voltages are induced in the two secondary windings. The frequency of the ac applied to the primary winding ranges from 50 Hz to 20 kHz.

The output voltage of the secondary windings S_1 is E_{S1} and that of secondary winding S_2 is E_{S2}.

In order to convert the output from S_1 to S_2 into a single voltage signal, the two secondaries S_1 and S_2 are connected in series opposition, as shown in Fig. 13.20.

Hence the output voltage of the transducer is the difference of the two voltages. Therefore the differential output voltage $E_o = E_{S1} \sim E_{S2}$.

When the core is at its normal position, the flux linking with both secondary windings is equal, and hence equal emfs are induced in them. Hence, at null position $E_{S1} = E_{S2}$. Since the output voltage of the transducer is the difference of the two voltages, the output voltage E_o is zero at null position.

Now, if the core is moved to the left of the null position, more flux links with winding S_1 and less with winding S_2. Hence, output voltage E_{S1} of the secondary winding S_1 is greater than E_{S2}. The magnitude of the output voltage of the secondary is then $E_{S1} - E_{S2}$, in phase with E_{S1} (the output voltage of secondary winding S_1).

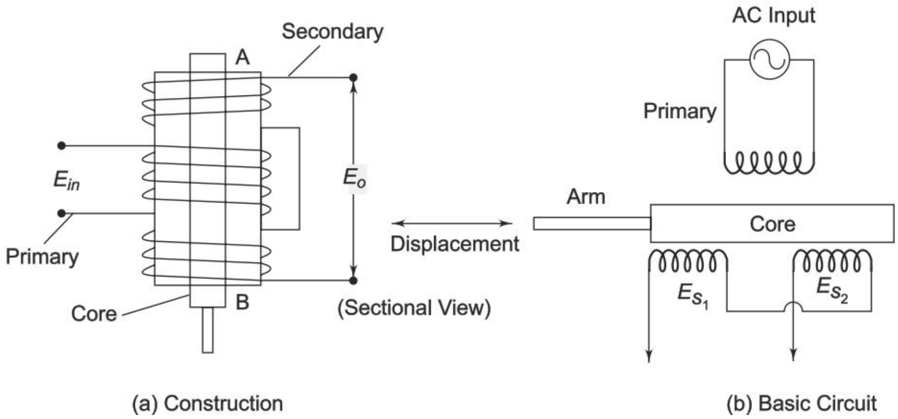


Fig. 13.20 Secondary winding connected for differential output

Similarly, if the core is moved to the right of the null position, the flux linking with winding S_2 becomes greater than that linked with winding S_1. This results in E_{S2} becoming larger than E_{S1}. The output voltage in this case is $E_o = E_{S2} - E_{S1}$ and is in phase with E_{S2}.

The amount of voltage change in either secondary winding is proportional to the amount of movement of the core. Hence, we have an indication of the amount of linear motion. By noting which output is increasing or decreasing, the direction of motion can be determined. The output ac voltage inverts as the core passes the centre position. The farther the core moves from the centre, the greater the difference in value between E_{S1} and E_{S2} and consequently the greater the value of E_o. Hence, the amplitude is function of the distance the core has moved, and the polarity or phase indicates the direction of motion, as shown in Fig. 13.21.

As the core is moved in one direction from the null position, the difference voltage, i.e. the difference of the two secondary voltages increases, while maintaining an in-phase relation with the voltage from the input source. In the other direction from the null position, the difference voltage increases but is 180° out of phase with the voltage from the source.

By comparing the magnitude and phase of the difference output voltage with that of the source, the amount and direction of the movement of the core and hence of the displacement may be determined.

The amount of output voltage may be measured to determine the displacement. The output signal may also be applied to a recorder or to a controller that can restore the moving system to its normal position.

The output voltage of an LVDT is a linear function of the core displacement within a limited range of motion (say 5 mm from the null position).

Figure 13.21(d) shows the variation of the output voltage against displacement for various position of the core. The curve is practically linear for small displacements (up to 5 mm). Beyond this range, the curve starts to deviate.

The diagram in Figs 13.21(a), (b) and (c) shows the core of an LVDT at three different positions.

In Fig. 13.21(b), the core is at O, which is the central zero or null position. Therefore, $E_{S1} = E_{S2}$ and $E_o = 0$.

When the core is moved to the left, as in Fig. 13.21(a) and is at A, E_{S1} is more than E_{S2} and E_o is positive. This movement represents a positive value and therefore the phase angle, is $\phi = 0°$.

When the core is moved to the right towards B, E_{S2} is greater than E_{S1} and hence E_o is negative. Therefore, S_2 the output voltage is 180° out of phase with the voltage which is obtained when the core is moved to the left. The characteristics are linear from $O - A$ and $O - B$, but after that they become non-linear.

One advantage of an LVDT over the inductive bridge type is that it produces higher output voltage for small changes in core position. Several commercial models that produce 50 mV/mm to 300 mV/mm are available. 300 mV/mm implies that a 1 mm displacement of the core produces a voltage output of 300 mV.

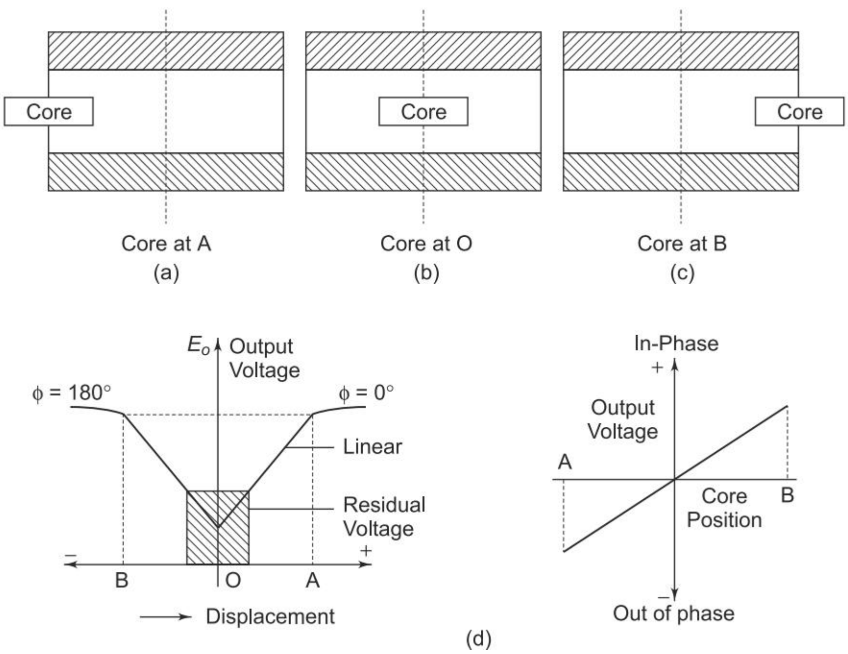


Fig. 13.21 (a), (b), (c) Various core position of LVDT
(d) Variation of output voltage vs displacement

LVDTs are available with ranges as low as ± 0.05 in. to as high as ± 25 in. and are sensitive enough to be used to measure displacements of well below 0.001

in. They can be obtained for operation at temperatures as low as – 265°C and as high as + 600°C and are also available in radiation resistance designs for nuclear operations.

Advantages of LVDT

1. ***Linearity*** The output voltage of this transducer is practically linear for displacements upto 5 mm (a linearity of 0.05% is available in commercial LVDTs).
2. ***Infinite resolution*** The change in output voltage is stepless. The effective resolution depends more on the test equipment than on the transducer.
3. ***High output*** It gives a high output (therefore there is frequently no need for intermediate amplification devices).
4. ***High sensitivity*** The transducer possesses a sensitivity as high as 40 V/mm.
5. ***Ruggedness*** These transducers can usually tolerate a high degree of vibration and shock.
6. ***Less friction*** There are no sliding contacts.
7. ***Low hysteresis*** This transducer has a low hysteresis, hence repeatability is excellent under all conditions.
8. ***Low power consumption*** Most LVDTs consume less than 1 W of power.

Disadvantages

1. Large displacements are required for appreciable differential output.
2. They are sensitive to stray magnetic fields (but shielding is possible).
3. The receiving instrument must be selected to operate on ac signals, or a demodulator network must be used if a dc output is required.
4. The dynamic response is limited mechanically by the mass of the core and electrically by the applied voltage.
5. Temperature also affects the transducer.

Example 13.6 *An ac LVDT has the following data.*

Input = 6.3 V, Output = 5.2 V, range ± 0.5 in. Determine

(i) Calculate the output voltage vs core position for a core movement going from + 0.45 in. to – 0.30 in.

(ii) The output voltage when the core is – 0.25 in. from the centre.

Solution

(i) 0.5 in. core displacement produces 5.2 V, therefore a 0.45 in. core movement produces $(0.45 \times 5.2)/0.5 = 4.68$ V.

Similarly a – 0.30 in. core movement produces

$$(-0.30 \times -5.2)/(-0.5) = -3.12 \text{ V}$$

(ii) – 0.25 in. core movement produces

$$(-0.25 \times -5.2)/(-0.5) = -2.6 \text{ V}$$

13.11.1 Rotational Variable Differential Transducer (RVDT)

As seen previously, an LVDT is used to measure linear displacement. But when angular displacement is desired, a Rotational Variable Differential Transducer (RVDT) is used.

A Rotary Variable Differential Transformer (RVDT) is a type of electrical transformer used for measuring angular displacement.

More precisely, a Rotary Variable Differential Transformer (RVDT) is an electromechanical transducer that provides a variable alternating current (ac) output voltage that is linearly proportional to the angular displacement of its input shaft. When energized with a fixed ac source, the output signal is linear within a specified range over the angular displacement.

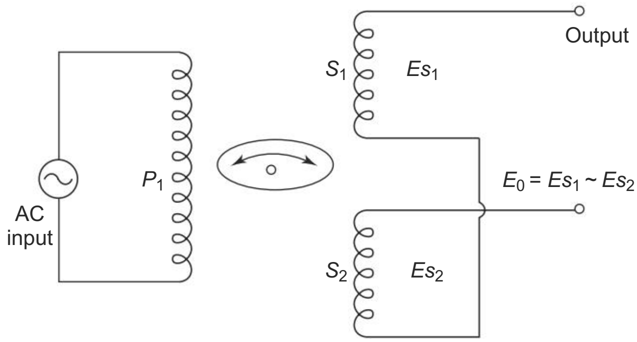


Fig 13.22 Rotational variable differential transducer

The RVDT is similar in construction to the LVDT, except that a cam-shaped core replaces the core in the LVDT as shown in Fig 13.22.

If the core is at the centre as in the case of LVDT, equal flux is linked with both secondary windings. Equal voltages are developed from the primary windings to the secondary windings. Hence the output being the differential output will be zero or null.

If the core is turned above the centre, the flux linking with one winding S_1, increases while the other S_2 decreases. Hence the output can be consider as a positive value.

If the core is turned in the other direction, the flux linking with winding S_1 reduces, while that linked with winding S_2 increases, hence producing an out of phase output that is in the opposite direction that is a negative value.

Hence the output gives the magnitude and direction of rotation. Hence angular displacement can be measured.

Most RVDT's are composed of a wound, laminated stator and a salient two-pole rotor. The stator, containing four slots, contains both the primary winding and the two secondary windings. Some secondary windings may also be connected together.

RVDT's utilize brushless, non-contacting technology to ensure long life and reliable, repeatable position sensing with infinite resolution. Such reliable and repeatable performance assures accurate position sensing under the most extreme operating conditions.

Operation of RVDTs The two induced voltages of the secondary windings, V_1 and V_2, varies linearly to the mechanical angle of the rotor, θ:

The difference $V_1 - V_2$ gives a proportional voltage:

$$\Delta V = 2 \cdot G \cdot \theta$$

and the sum of the voltages is a constant:

$$C - \Sigma V = 2 \cdot V_0$$

This constant gives the RVDT great stability of the angular information, independence of the input voltage or frequency, or temperature and enables it to also detect a malfunction.

Basic RVDT construction and operation is provided by rotating an iron-core bearing supported within a housed stator assembly. The housing is passivated stainless steel. The stator consists of a primary excitation coil and a pair of secondary output coils. A fixed alternating current excitation is applied to the primary stator coil that is electromagnetically coupled to the secondary coils. This coupling is proportional to the angle of the input shaft. The output pair is structured so that one coil is in-phase with the excitation coil, and the second is 180 degrees out-of-phase with the excitation coil. When the rotor is in a position that directs the available flux equally in both the in-phase and out-of-phase coils, the output voltages cancel and result in a zero value signal. This is referred to as the Electrical Zero position or EZ. When the rotor shaft is displaced from EZ, the resulting output signals have a magnitude and phase relationship proportional to the direction of rotation. Because RVDTs perform essentially like a transformer, excitation voltage changes will cause directly proportional changes to the output (transformation ratio). However, the voltage out to excitation voltage ratio will remain constant. Since most RVDT signal conditioning systems measure signal as a function of the Transformation Ratio (TR), excitation voltage drift beyond 7.5% typically has no effect on sensor accuracy and strict voltage regulation is not typically necessary. Excitation frequency should be controlled within +/– 1% to maintain accuracy

Although the RVDT can theoretically operate between ±45°, accuracy decreases quickly after ±35°. Thus, its operational limits lie mostly within ±30°, but some up to ±40°. Certain types can operate up to ±60°.

Advantages

- low sensitivity to temperature, primary voltage and frequency variations
- sturdiness
- relative low cost due to its popularity
- simple control electronics
- small size
- solid and robust, capable of working in a wide variety of environments

- No friction resistance, since the iron core does not contact the transformer coils, resulting in an very long service life
- High signal to noise ratio and low output impedance
- Negligible hysteresis
- Infinitesimal theoretical resolution; in reality, angle resolution is limited by the resolution of the amplifiers and voltage meters used to process the output signal
- No permanent damage to the RVDT if measurements exceed the designed range

Disadvantages The core must be in contact (directly or indirectly) with the measured surface which is not always possible or desirable.

Common Specifications Common specifications for commercially available RVDT's are listed below:

Input: Power input is a 3 to 15 V (rms) sine wave with a frequency between 60 to 20,000 Hz.

Angle: Capable of continuous rotational measurement. However, most RVDTs have effective angle limits of up to ±60°.

Nonlinearity: Higher accuracy in the smaller angle range: 0.25% @ ±30°, 0.50% @ ±40°, 1.50% @ ±60°.

PRESSURE INDUCTIVE TRANSDUCER 13.12

A simple, arrangement, wherein a change in the inductance of a sensing element is produced by a pressure change, is given in Fig. 13.23.

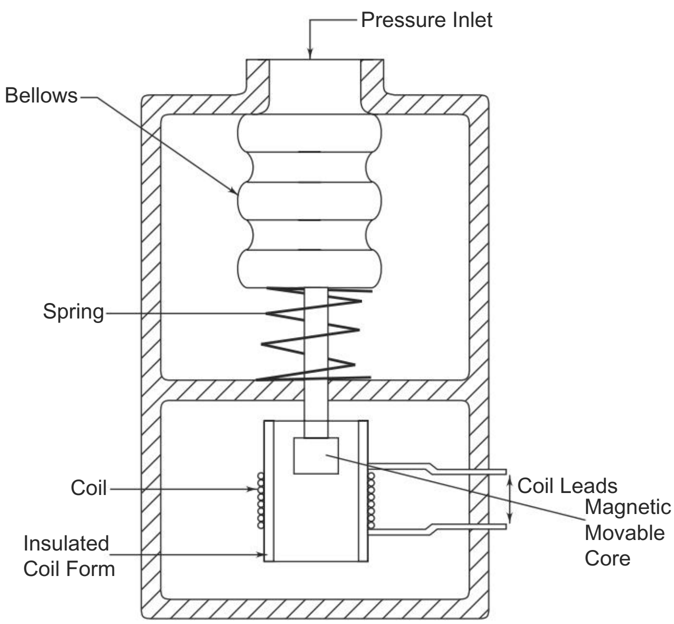


Fig. 13.23 Pressure inductive transducer

Here the pressure acting on a movable magnetic core causes an increase in the coil inductance corresponding to the acting pressure. The change in inductance can again be made on the basis of an electrical signal, using an ac bridge.

An advantage of the inductive type over the resistive type is that no moving contacts are present, thereby providing continuous resolution of the change, with no extra friction load imposed on the measuring system.

In a slightly modified form, this principle is used to obtain a change in mutual inductance between magnetically coupled coils, rather than in the self inductance of a single coil. When a change in an induced voltage is involved, the transducer is sometimes called a variable reluctance sensor or magnetic pickup. A very important example of the mutual type is the LVDT.

13.12.1 Inductive Position Transducers (Synchro's)

Synchro's is a generic name for a faculty of inductive devices which can be connected in various ways to form shaft angle measurement. All these devices work essentially on the same principle, that is of a rotating transformer. A Synchro appears like an AC motor consisting of a rotor and a stator.

Synchro's are normally used in control system, but have properties that can be used in instrumentation also.

A Synchro can be an angular position transducer working on inductive principle, wherein a variable coupling between primary and secondary winding is obtained by changing the relative orientation of the windings.

Internally, most synchro's are similar in construction. They have a rotor with one or three windings capable of revolving inside a fixed stator. There are two common types of rotors, the *salient pole* and the *wound rotor*.

The primary winding is a single phase winding wound on a rotor made of laminations. The connection to the rotor windings are made through precision slip rings shown in Fig. 13.24.

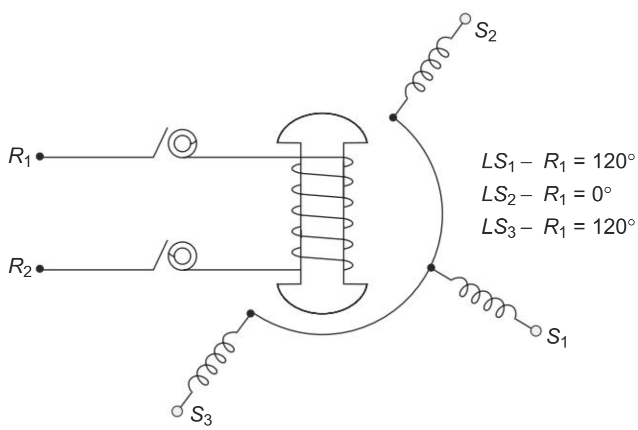


Fig. 13.24 Basic synchro

The stator has a 3-phase winding with the windings of the 3-phase displaced by 120°. The synchro may be viewed as a variable coupling transformer. A synchro is also called as *Selsyn*.

The rotor is energized by an ac voltage and coupling between rotor and stator windings varies as a trigonometric or linear function of the rotor position.

Synchro systems consists of two or more interconnected synchros. They are grouped or connected together according to the purpose to be used.

A Synchro system formed by interconnection of the devices called the Synchro transmitter and Synchro control transmitter is perhaps the most widely used error detector in feedback control system. It measures and compares two angular displacements and its output voltage is approximately linear with angular displacement.

The conventional Synchro transmitter (TX) uses a salient pole rotor with sleeved slot. When an ac excitation voltage is applied to the rotor, the resultant current produces a magnetic field and by transformer action induces voltages in the stator coils. The effective voltage induced in any stator coil depends upon the angular position of the coil axis with respect to the rotor axis (when the coil voltage is known, the induced voltage at any angular displacement can be determined).

Initially winding S_2 of the stator of transmitter is positioned for maximum coupling with the rotor winding as shown in Fig. 13.25(a). Suppose the voltage is V, the coupling between S_1 and S_2 of the stator and primary (rotor) winding is a cosine function. In general if the rotor is excited by 50 Hz ac, also called *reference voltage*, the voltage induced in any stator winding will be proportional to the cosine of the angle between the rotor axis and the stator axis. The voltages induced across any pair of stator terminals ($S_1 - S_2$, $S_1 - S_3$, or $S_2 - S_3$) will be sum or difference, depending on the phase of the voltage measured across the coils.

For example, if a reference voltage $V \sin \omega t$ excites the rotor of a synchro ($R_1 - R_2$), the stator terminals will have a voltage of the following form:

$$V(S_1 - S_2) = V \sin \omega t \sin \theta$$

$$V(S_1 - S_2) = V \sin \omega t (\sin \theta + 120°)$$

$$V(S_2 - S_3) = V \sin \omega t (\sin \theta + 240°)$$

where θ is the shaft angle.

These voltages are known as *Synchro format voltages*.

Therefore, the effective voltages in these windings are proportional to cos 60° or they are $V/2$ each. So long as the rotors of the transmitter and receiver remains in this position, no current will flow between the stator windings because of the voltage balance.

When the rotor of the transmitter is moved to a new position, the voltage balance is disturbed or changed. Assuming that the rotor of the transmitter is moved through 30° as shown in Fig. 13.25(b), the stator winding voltages of the transmitter will be changed to 0, $\sqrt{3/2}\,V$ and $\sqrt{3/2}\,V$ respectively.

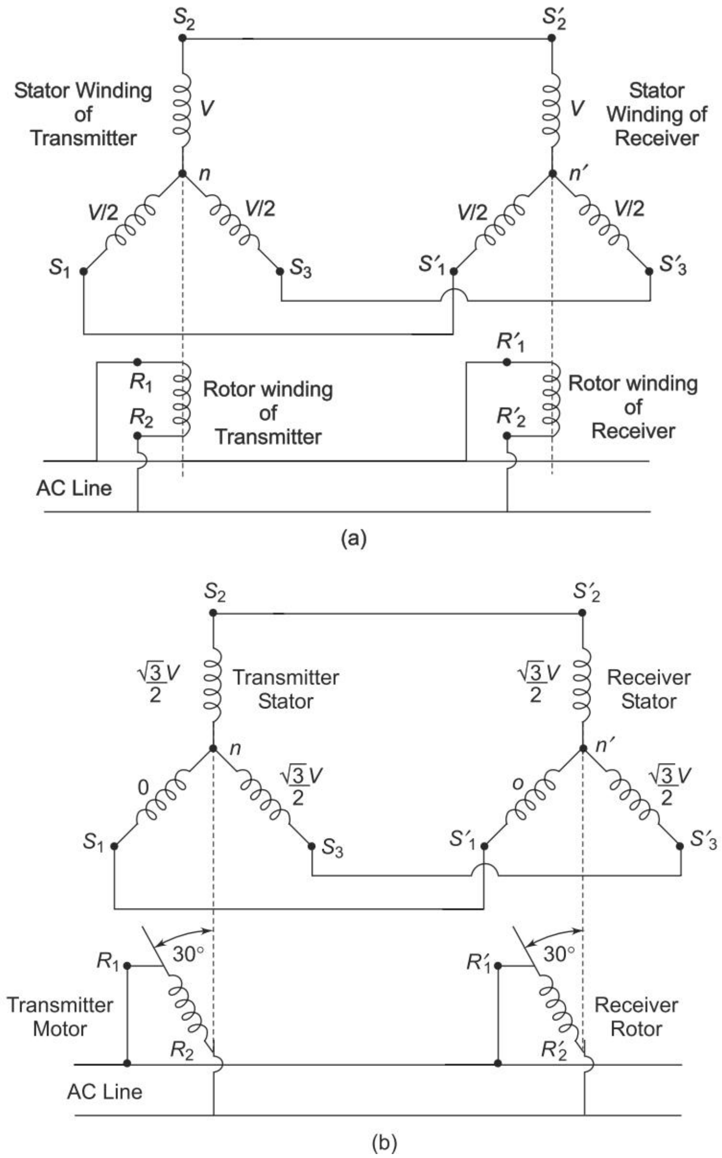


Fig. 13.25 (a) Torque transmission using synchro trans
(b) Follow up conditions of transmitter-receiver system

Hence, a voltage imbalance occurs between the stator windings of the transmitter and receiver. This voltage imbalance between the windings causes current to flow between the windings producing a torque that tends to rotate the rotor of the receiver to a new position where the voltage balance is again restored. This balance is restored only if the receiver turns through the same angle as the transmitter and also the direction of rotation is the same as that of the transmitter. Hence a Synchro can be used to determine the magnitude and direction of angular displacement.

Torque synchros are required if it is necessary to transmit angular displacement information from a shaft of one Synchro to the shaft of another Synchro without using any additional amplifiers or gears.

Torque synchro are commonly connected in repeater system consisting of torque transmitter and torque receiver. These repeater system is accurate to ± 1° and is used in systems in which a rotating devices output is required to position a remote pointer.

Today, most of these units have been replaced by a Syncrho-to-digital converter driving a LED display of the digital readout of angular position.

13.12.2 Selsyn

Synchros also called as Selsyn are motor like devices used to form position sensing and indicating system. Its principle is also based on the variation of mutual coupling between transmitter windings.

There are two types of Selsyn:

1. Torque type
2. Control type

Torque type of Selsyn consists of two rotors with a single winding and a stator with 3 windings distributed 120° apart. This section acts as a transmitter. The generator stator windings are connected to the other remotely located 3 phase windings. This section acts as a receiver and also has receiver rotor windings. The torque Selsyn is shown in Fig. 13.26.

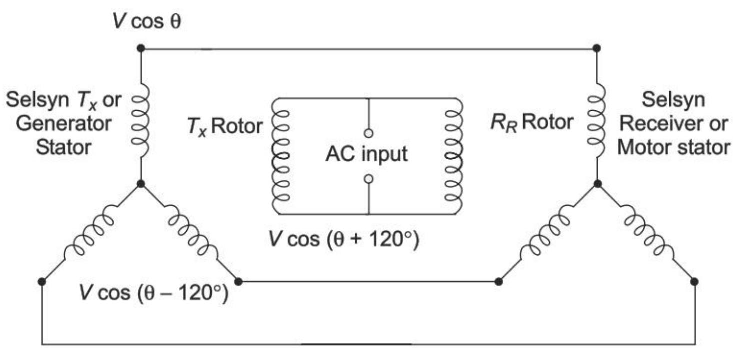


Fig. 13.26 Torque selsyn

The rotor of both units are excited through slip rings by an ac source at a convenient frequency.

If the receiver rotor is in the same relative position with reference to stator windings as the transmitter, the inductive coupling is identical and no stator current flows between the two units.

If transmitter rotor is shifted from this null position, voltages induced in the two stator will be the difference and a current starts to flow. This current flow produces a torque in the receiver which causes the receiver rotor to move into alignment with the new transmitter rotor position and the previous null position is obtained. Hence the receiver always follow the angular motion in accordance with that of the transmitter. Its accuracy is limited by the friction of the receiver bearings and calibration of its dial face.

13.12.3 Control Synchro

When larger amounts of power or torque and greater accuracy are required control synchros are used.

These devices are used for providing and handling control signals to a servo amplifier when more power and accurate angular displacement of a large load are required. The control synchros are not designed to handle any mechanical load.

The Control Transformer (CT) develops an ac rotor output voltage that is proportional to the relative shaft angles between synchro transmitter and control transformer. The devices are normally connected as shown in Fig. 13.27.

The two most common control synchro are Control Transmitter (CX) and Control Transformer (CT).

The output of the CX (transmitter) is fed to the stator of CT (transformer). The CT is a high impedance version of the torque receiver with its rotor aligned at 90° to that of transmitter (TR).

In a control system, when the shaft angle of the CX equals that of CT shaft angle, a minimum and null voltage will appear on the rotor terminals R_1 and R_2 of the CT. Any variation from this null will produce a signal in CT rotor whose phase will depend in which direction it is moved off the null.

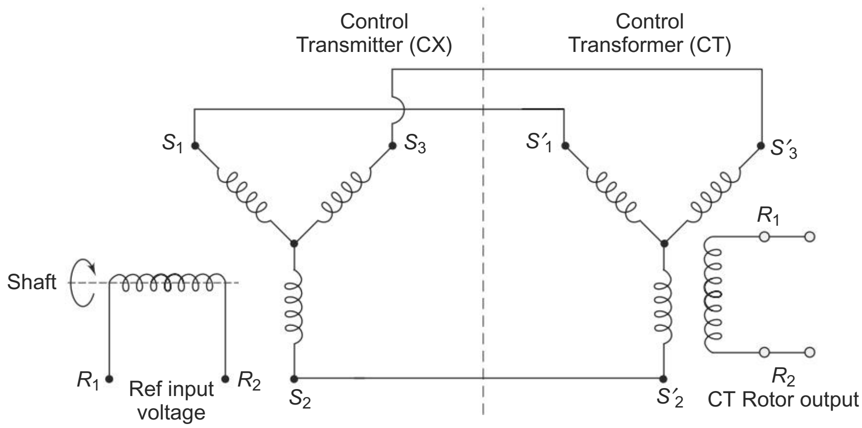


Fig. 13.27 Synchro control system

A simple closed loop servo system using a CX and CT control system is shown in Fig. 13.28.

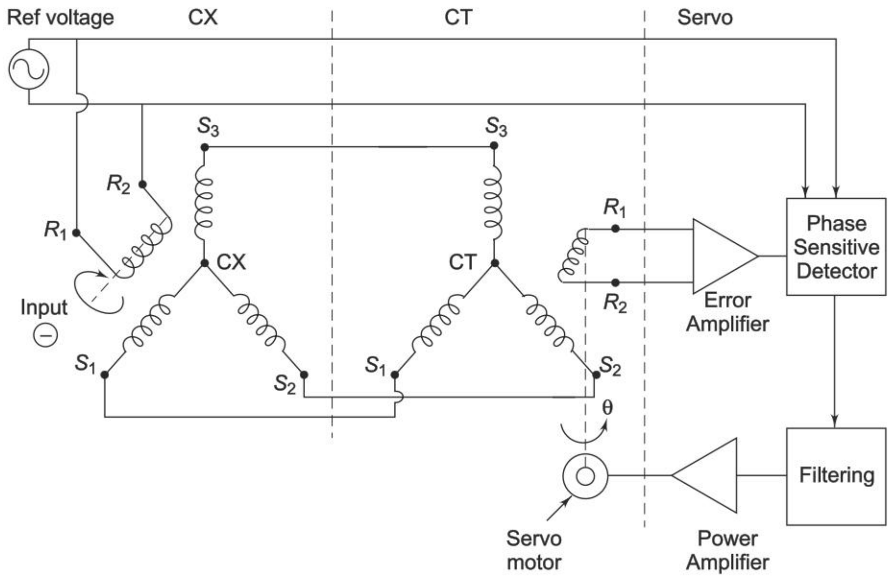


Fig. 13.28 Simple servo system using control synchro

When the shaft of the CX is turned to some angle θ, S_1, S_2 and S_3 output provide synchro format voltage as discussed previously. These voltages are transmitted to the CT stators S_1, S_2, S_3. If the CT is not at the input angle θ, a voltage will be produced at the output of the CT rotor winding. This signal (error) is amplified, phase detected and fed to a servo amplifier to cause a servo motor to position a load and the CT shaft to a position where the CT rotor output is minimum (null). The direction, in which the motor turns towards the angle θ, is determined by the phase of the CT rotor with respect to the reference voltage. Hence servo action takes place.

CAPACITIVE TRANSDUCER (PRESSURE) 13.13

A linear change in capacitance with changes in the physical position of the moving element may be used to provide an electrical indication of the element's position.

The capacitance is given by $C = KA/d$ (13.14)

where K = the dielectric constant

A = the total area of the capacitor surfaces

d = distance between two capacitive surfaces

C = the resultant capacitance.

From this equation, it is seen that capacitance increases (i) if the effective area of the plate is increased, and (ii) if the material has a high dielectric constant.

The capacitance is reduced if the spacing between the plates is increased.

Transducers which make use of these three methods of varying capacitance have been developed.

With proper calibration, each type yields a high degree of accuracy. Stray magnetic and capacitive effects may cause errors in the measurement produced, which can be avoided by proper shielding. Some capacitive dielectrics are temperature sensitive, so temperature variations should be minimised for accurate measurements.

A variable plate area transducer is made up of a fixed plate called Stator and a movable plate called the Rotor.

The rotor is mechanically coupled to the member under test. As the member moves, the rotor changes its position relative to the stator, thereby changing the effective area between the plates. A transducer of this type is shown in Fig. 13.29.

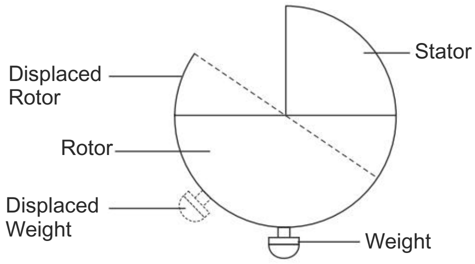


Fig. 13.29 Capacitive transducer

Such a device is used to detect the amount of roll in an aircraft. As the aircraft rolls to the left, the plates moves to the relative position shown by dashed lines in Fig. 13.29 and the capacitance decreases by an amount proportional to the degree of roll. Similarly to the right. In this case the stator, securely attached to the aircraft, is the moving element. The weight on the rotor keeps its position fixed with reference to the surface of the earth, but the relative position of the plates changes and this is the factor that determines the capacitance of the unit.

Figure 13.30 shows a transducer that makes use of the variation in capacitance resulting from a change in spacing between the plates. This particular transducer is designed to measure pressure (in vacuum).

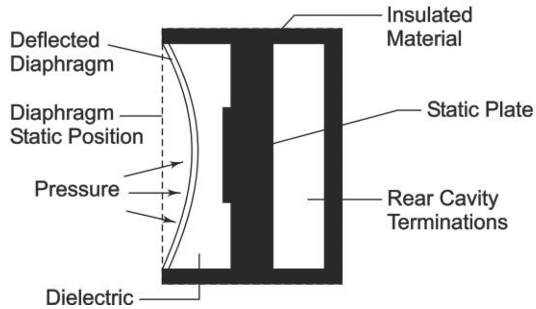


Fig. 13.30 Capacitive pressure transducer

Enclosed in an airtight container is a metallic diaphragm which moves to the left when pressure is applied to the chamber and to the right when vacuum is applied. This diaphragm is used as one plate of a variable capacitor. Its distance from the stationary plate to its left, as determined by the pressure applied to the unit, determines the capacitance between the two plates. The monitor indicates the pressure equivalent of the unit's capacitance by measuring the capacitor's reactance to the ac source voltage.

(The portion of the chamber to the left of the moving plate is isolated from the side into which the pressurised gas or vapour is introduced. Hence, the dielectric constant of the unit does not change for different types of pressurised gas or vapour. The capacity is purely a function of the diaphragm position.) This device is not linear.

Changes in pressure may be easily detected by the variation of capacity between a fixed plate and another plate free to move as the pressure changes. The resulting variation follows the basic capacity formula.

$$C = 0.885\,\frac{K(n-1)A}{t}\ \text{pf} \tag{13.15}$$

where A = area of one side of one plate in cm^2
n = number of plates
t = thickness of dielectric in cm
K = dielectric constant

The capacitive transducer, as in the capacitive microphone, is simple to construct and inexpensive to produce. It is particularly effective for HF variations.

However, when the varying capacitance is made part of an ac bridge to produce an ac output signal, the conditions for resistive and reactive balance generally require much care to be taken against unwanted signal pickup in the high impedance circuit, and also compensation for temperature changes. As a result, the receiving instrument for the capacitive sensor usually calls for more advanced and complex design than is needed for other transducers.

LOAD CELL (PRESSURE CELL) 13.14

The load cell is used to weigh extremely heavy loads. A length of bar, usually steel, is used as the active element. The weight of the load applies a particular stress to the bar. The amount of strain which results in the bar for different values of applied stress is determined, so that the strain may be used as a direct measure of the stress causing it.

The load cell shown in Fig. 13.31 is a good example of the use of strain gauges in weighing operations.

As the stress is applied along the direction of S (shown by the arrow in Fig. 13.31), the steel bar experiences a compression along that axis and an expansion along the X and Y axes. As a result, gauge A experiences a decrease in resistance, while gauge B undergoes an increase in resistance. When these two gauges and the gauges on the two remaining sides of the steel are connected to form a bridge circuit, four times the sensitivity of a simple gauge bridge is obtained. This makes the load cell sensitive to very small values of applied stress, as well as to extremely heavy loads.

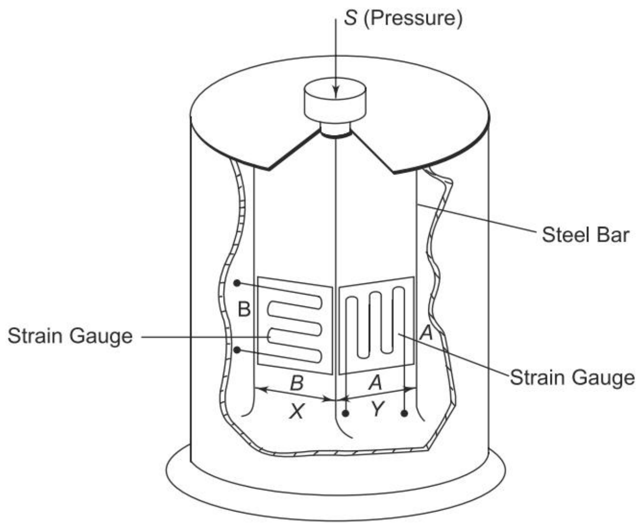


Fig. 13.31 Strain gauge load cell

PIEZO ELECTRICAL TRANSDUCER 13.15

A symmetrical crystalline materials such as Quartz, Rochelle salt and Barium titanate produce an emf when they are placed under stress. This property is used in piezo electric transducers, where a crystal is placed between a solid base and the force-summing member, as shown in Fig. 13.32.

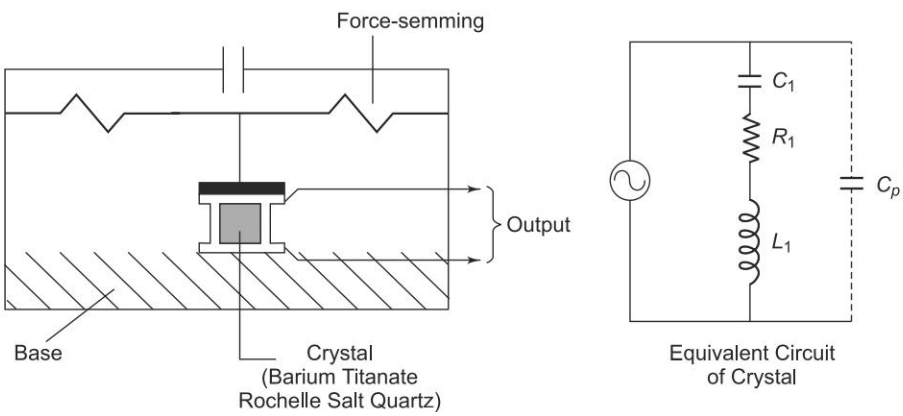


Fig. 13.32 Piezo electric transducer

An externally applied force, entering the transducer through its pressure port, applies pressure to the top of a crystal. This produces an emf across the crystal proportional to the magnitude of applied pressure.

Since the transducer has a very good HF response, its principal use is in HF accelerometers. In this application, its output voltage is typically of the order of 1 – 30 mV per gm of acceleration. The device needs no external power source and is therefore self generating. The disadvantage is that it cannot measure static conditions. The output voltage is also affected by temperature variation of the crystal. The basic expression for output voltage E is given by

$$E = \frac{Q}{C_p}$$

where Q = generated charge
C_p = shunt capacitances

This transducer is inherently a dynamic responding sensor and does not readily measure static conditions. (Since it is a high impedance element, it requires careful shielding and compensation.)

For a piezo electric element under pressure, part of the energy is converted to an electric potential that appears on opposite faces of the element, analogous to a charge on the plates of a capacitor. The rest of the applied energy is converted to mechanical energy, analogous to a compressed spring. When the pressure is removed, it returns to its original shape and loses its electric charge.

From these relationships, the following formulas have been derived for the coupling coefficient K.

$$K = \frac{\text{Mechanical energy converted to electrical energy}}{\text{Applied mechanical energy}}$$

or

$$K = \frac{\text{Electrical energy converted to mechanical energy}}{\text{Applied electrical energy}}$$

An alternating voltage applied to a crystal causes it to vibrate at its natural resonance frequency. Since the frequency is a very stable quantity, piezo electric crystals are principally used in HF accelerometers.

The principal disadvantage is that voltage will be generated as long as the pressure applied to the piezo electric element changes.

Example 13.7 *A certain crystal has a coupling coefficient of 0.32. How much electrical energy must be applied to produce an output of 1 oz.in. of mechanical energy?*

Solution

$$1 \text{ oz.in.} = 1 \text{ oz.in.} \times \frac{1 \text{ ft}}{12 \text{ in.}} \times \frac{1 \text{ lb}}{16 \text{ oz}} \times \frac{1.3561}{1 \text{ ft lb}}$$

$$= 7.06 \times 10^{-3} \text{ J}$$

$$\text{Applied Electrical energy} = \frac{\text{Electrical energy converted to mechanical energy}}{K}$$

$$= \frac{7.06 \times 10^{-3}}{0.32} = 22.19 \text{ mJ}$$

PHOTO ELECTRIC TRANSDUCER 13.16

Photo electric devices can be categorised as photo emissive, photo-conductive or photo-voltaic.

In photo emissive devices, radiation falling on a cathode causes electrons to be emitted from the cathode surface.

In photo conductive devices, the resistance of a material is changed when it is illuminated.

Photo voltaic cells generate an output voltage proportional to the radiation intensity. The incident radiation may be infrared, ultraviolet, gamma rays, X-rays, or visible light.

13.16.1 Photo Multiplier Tube

The photo multiplier tube consists of an evacuated glass envelope containing a photo cathode, an anode and several additional electrodes, termed Dynodes, each at a higher voltage, than the previous dynode.

Figure 13.33 illustrates the principle of the photo multiplier. Electrons emitted by the cathode are attracted to the first dynode. Here a phenomenon known as secondary emission takes place.

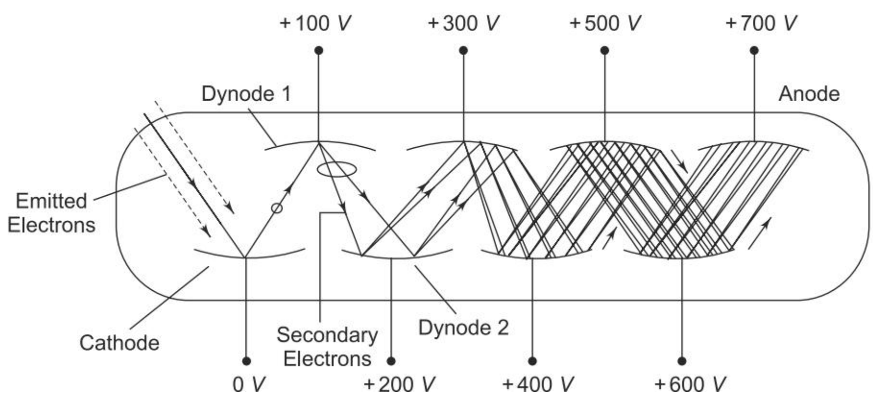


Fig. 13.33 Principle of a photo multiplier tube

When electrons moving at a high velocity strike an appropriate material, the material emits a greater number of electrons than it was struck with.

In this device, the high velocity is achieved by the use of a high voltage between the anode and the cathode. The electrons emitted by the first dynode are then attracted to the second dynode, where the same action takes place again. Each dynode is at a higher voltage, in order to achieve the requisite electron velocity each time. Hence, secondary emission, and a resulting electron multiplication, occurs at each step, with an overall increase in electron flow that may be very great. Amplification of the original current by much as $10^5 - 10^9$ is common. Luminous sensitivities range from 1A per lumen or less, to over 2000 A per lumen. Typical anode current ratings range from a minimum of 100 μA to a maximum of 1 mA.

The extreme luminous sensitivity possible with these devices is such that for a sensitivity of 100 A per lumen, only 10^{-5} lumen is needed to produce 1 mA of output current.

Magnetic fields affect the photo multiplier because some electrons may be deflected from their normal path between stages and therefore never reach a dynode or anode. Hence, the gain falls. To minimise this effect μ-metal magnetic shields are often placed around the photo multiplier tube.

13.16.2 Photo Conductive Cells or Photo Cells

The devices discussed above achieve an electrical output by photo emission. Another photo electric effect that is very useful is the photo conductive effect. In this effect, the electrical resistance of the material varies with the amount of incident light, as shown in Fig. 13.34(b).

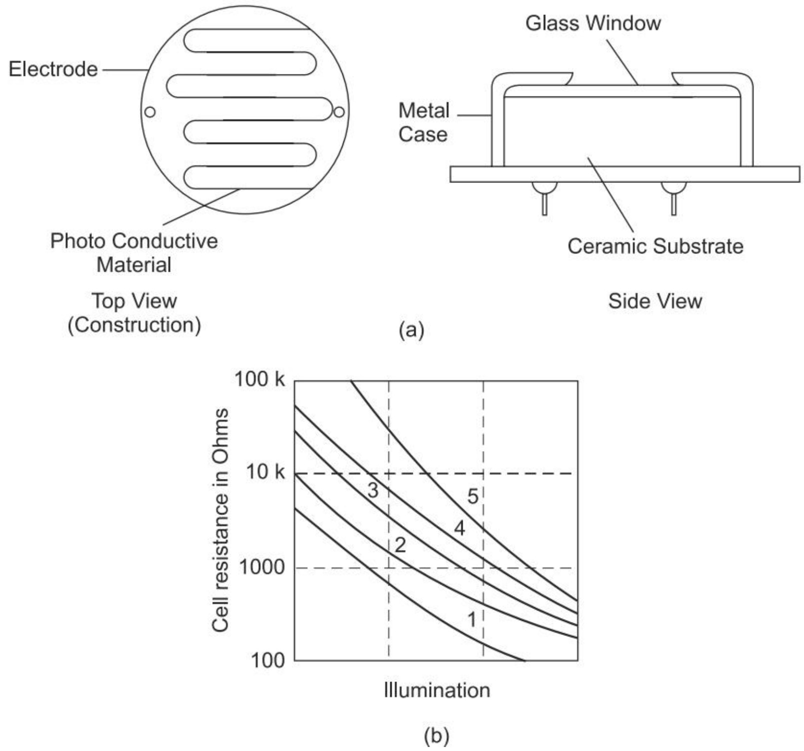


Fig. 13.34 The photo conductive cell (a) construction (b) Typical curves of resistance vs illumination

A typical construction is as shown in Fig. 13.34(a). The photo conductive material, typically Cadium sulphide, Cadium selenide or Cadium sulpho-selenide, is deposited in a zig zag pattern (to obtain a desired resistance value and power rating) separating two metal coated areas acting as electrodes, all on an insulating base such as ceramic. The assembly is enclosed in a metal case with a glass window over the photo conductive material.

Photocells of these types are made in a wide range of sizes, from 1/8 in. in diameter to over 1 in. The small sizes are suitable where space is critical, as in punched card reading equipment. However, very small units have low power dissipation ratings.

A typical control circuit utilising a photo conductive cell is illustrated in Fig. 13.35. The potentiometer is used to make adjustments to compensate for manufacturing tolerances in photocells sensitivity and relay operating sensitivity.

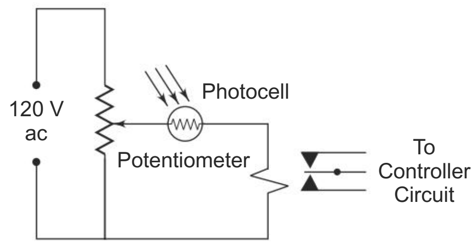


Fig. 13.35 Photo cell and relay control circuit

When the photocell has the appropriate light shining on it, its resistance is low and the current through the relay is consequently high enough to operate the relay. When the light is interrupted, the resistance rises, causing the relay current to decrease enough to de-energise the relay.

Example 13.8 (a) *The relay of Fig. Ex. 13.7(a) is to be controlled by a photo-conductive cell with the characteristics shown in Fig. Ex. 13.7(b). The potentiometer delivers 10 mA at a 30 V setting when the cell is illuminated with about 400 l/m^2 and is required to be de-energised when the cell is dark. Calculate (i) the required series resistance, and the (ii) dark current level.*

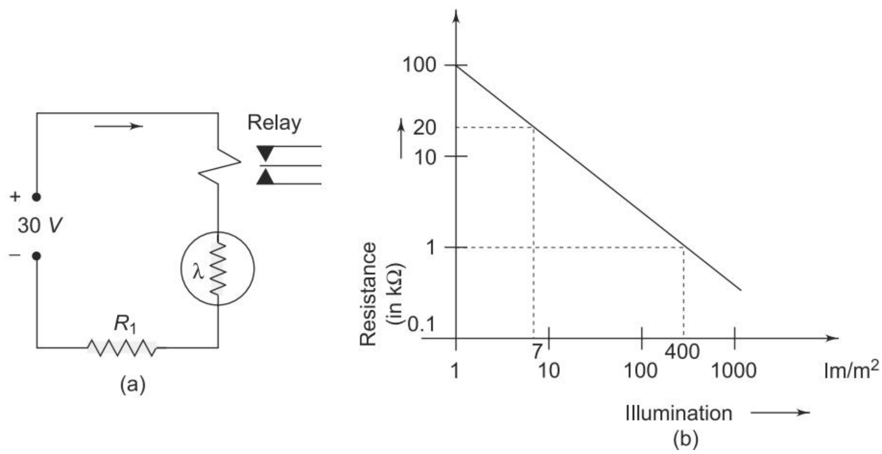


Fig. Ex. 13.7

Solution

(i) The cell resistance at 400 l/m^2 is 1 kΩ.

Therefore, $$I = \frac{30\,V}{R_1 + R_{cell}} \quad \text{i.e.} \quad R_1 = \frac{30\,V}{I} - R_{cell}$$

$$= \frac{30\,V}{10\text{ mA}} - 1\text{ k} = 3\text{ k} - 1\text{ k} = 2\text{ k}\Omega$$

(ii) The cells dark resistance is 100 kΩ

$$\text{Dark current} = \frac{30\,V}{2\text{ k} + 100\text{ k}} = \frac{30\,V}{102\text{ k}} = 0.3\text{ mA}$$

A typical use of this type of circuit is in counting objects, as on the conveyor belt of a production line. In this type of a circuit a beam of light traverses the path of the moving object that is on a conveyor belt. These objects have to be counted. Therefore, each time an object passes the beam is interrupted, and the relay actuates a counter.

PHOTO-VOLTAIC CELL 13.17

The photo-voltaic or solar cell, produces an electrical current when connected to a load. Both silicon (Si) and selenium (Se) types are known for these purposes.

Multiple unit silicon photo-voltaic devices may be used for sensing light in applications such as reading punched cards in the data processing industry.

Gold-doped germanium cells with controlled spectral response characteristics act as photo-voltaic devices in the infra-red region of the spectrum and may be used as infra-red detectors.

The silicon solar cell converts the radiant energy of the sun into electrical power. The solar cell consists of a thin slice of single crystal P-type silicon, up to 2 cm^2 into which a very thin (0.5 micron) layer of N-type material is diffused. The conversion efficiency depends on the spectral content and intensity of illumination.

SEMICONDUCTOR PHOTO DIODE 13.18

A reverse biased semiconductor diode passes only a very small leakage current (a fraction of 1 μA in silicon diodes), if the junction is exposed to light. Under illumination, however, the current rises almost in direct proportion to the light intensity. Hence, the photo-diode can be used for the same purposes as a photo-conductive cell.

This device, when operated with a reverse voltage applied, functions as a photo-conductive cell. When operated without reverse voltage it operates as a photo-voltaic cell.

A photo-diode can also be arranged to change from photo-conductive to photo-voltaic mode.

The response time of a photo-diode is very fast, so that it may be used in applications where light fluctuations occur at high frequency, but a photo-diode cell is useful only at very low frequencies.

The symbol and illumination characteristics are as shown in Figs. 13.36(a) and (b).

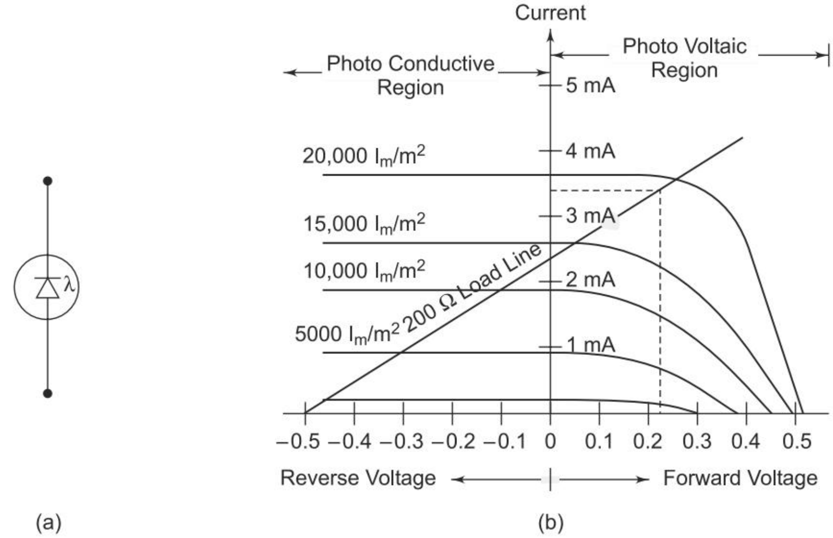


Fig. 13.36 (a) Photo diode symbol (b) Illumination characteristics

THE PHOTO-TRANSISTOR 13.19

The sensitivity of a photo diode can be increased by as much as 100 times by adding a junction, resulting in an NPN device. A simple representation of the construction is shown in Fig. 13.37.

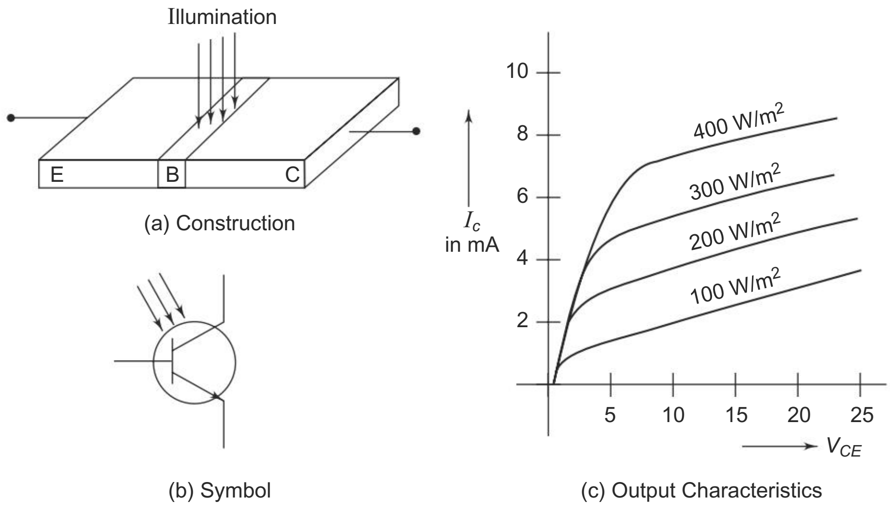


Fig. 13.37 Photo-transistor

Illumination of the central region causes the release of electron hole pairs. This lowers the barrier potential across both junctions, causing an increase in the flow of electrons from the left region into the centre region and on to the right region.

For a given amount of illumination on a very small area, the photo-transistor provides a much larger output current than that available from a photo diode, i.e. a photo-transistor is more sensitive.

Arrays of transistors and low current photo diodes are widely used as photo detectors for such applications as punched card and tape readouts. Photo-transistors are more sensitive than photo-diodes, but the latter have a faster switching time.

One application of a photo-transistor is shown in Fig. 13.38.

The light incident on the photo-transistor causes its current to increase and therefore increases both the voltage drop across 50 kΩ and the input to the transistor which drives the relay. This raises the current to the operational level.

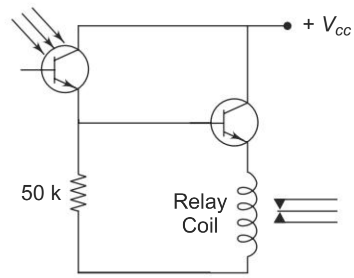


Fig. 13.38 Photo transistor and relay circuit

TEMPERATURE TRANSDUCERS 13.20

13.20.1 Introduction to Temperature Transducers

Temperature is one of the most widely measured and controlled variable in industry, as a lot of products during manufacturing requires controlled temperature at various stages of processing.

A wide variety of temperature transducers and temperature measurement systems have been developed for different applications requirements.

Most of the temperature transducers are of Resistance Temperature Detectors (RTD), Thermistors and Thermocouples. Of these RTD's and Thermistor are passive devices whose resistance changes with temperature hence need an electrical supply to give a voltage output. On the other hand thermocouples are active transducers and are based on the principle of generation of thermoelectricity, when two dissimilar metals are connected together to form a junction called the *sensing junction*, an emf is generated proportional to the temperature of the junction. Thermocouple operate on the principle of *seeback effect*. Thermocouple introduces errors and can be overcomed by using a reference junction compensation called as a *cold junction compensation*.

Thermocouples are available that span cryogenic to 2000°C temperature range. They have the highest speed of response. Thermocouples can be connected in series/parallel to obtain greater sensitivity called a *Thermopile*.

RTD commonly use platinum, Nickel or any resistance wire whose resistance varies with temperature and has a high intrinsic accuracy. Platinum is the most widely used RTD because of its high stability and large operating range. RTD's are usually connected in a Wheatstones bridge circuit. The lead wire used for connecting the RTD's introduces error, hence compensation is required. This is obtained by using three-wire or four wire compensation, but 3-wire compensation is mostly used in the industry.

Another form of temperature measurement is by the use of thermistor. A thermistor is a thermally sensitive resistor that exhibits change in electrical resistance with change in temperature. Thermistors made up of oxides exhibit a negative temperature coefficient (NTC), that is, their resistance decreases with increase in temperature. Thermistor are also available with positive temperature coefficient (PTC), but PTC thermistor are seldom used for measurement since they have poor sensitivity.

Thermistors are available in various sizes and shapes such as beads, rods, discs, washers and in the form of probes.

Radiation pyrometer are used where non-contact temperature is required to be measured. It measures the radiant (energy) heat emitted or reflected by a hot object. Radiation pyrometers are of two types total radiation pyrometer and infrared pyrometer.

Total radiation pyrometer virtually receives all the radiation from a heated body and measures temperature in the range around 1200°C–3500°C. Infrared pyrometers are partial or selective radiation pyrometers and are used in the range of 1000°C–1200°C.

Optical pyrometers are used in the visible wavelength. The most common type of optical pyrometer is the disappearing filament type and is used in temperature range of 1400°C and can be extended up to 3000°C. Optical pyrometers are widely used for accurate measurement of temperatures of furnaces, molten metals etc.

13.20.2 Resistance Temperature Detector (RTD)

Resistance temperature detector* commonly use platinum, nickel or any resistance wire whose resistance varies with temperature and which has a high intrinsic accuracy. They are available in many configuration and sizes; as shielded or open units for both immersion and surface applications.

The relationship between temperature and resistance of conductors in the temperature range near 0°C can be calculated using the equation

$$R_t = R_{ref}\,(1 + \alpha\Delta\, t) \tag{13.16}$$

where R_t = resistance of conductor at temperature t°C

R_{ref} = resistance of the reference temperature, usually 0°C

α = temperature coefficient of resistance

$\Delta\, t$ = difference between operating and reference temperature

Almost all metals have a positive temperature coefficient (PTC) of resistance, so that their resistances increases with increase in temperature. Some materials,

* Refer to section 13.7.

such as Carbon and Germanium have a negative temperature coefficient (NTC) of resistance.

A high value of 'α' is desired in a temperature sensing element, so that sufficient change in resistance occurs for a relatively small change in temperature. This change in resistance (ΔR) can be measured with a Wheatstone's bridge which can be calibrated to indicate the temperature, that caused the resistance change rather than the resistance itself. The sensing element of the RTD is selected according to the intended applications.

RTD's are wire-wound resistance with moderate resistance and a PTC of resistance. Platinum is the most widely used resistance wire type because of its high stability and large operating range. However, Nickel and Copper are also used in RTDs. The temperature ranges for various resistance wire are given in Table 13.1.

Table 13.1

Platinum	– 200°C – 850°C
Copper	– 200°C – 260°C
Nickel	– 80°C – 300°C

Platinum RTDs provide high accuracy and stability. They have the following advantages:

1. Linearity over a wide operating range.
2. Wide operating range
3. Higher temperature operation
4. Better stability at high temperature

Disadvantages of RTD

1. Low sensitivity
2. It can be affected by contact resistance, shock and vibration
3. Requires no point sensing
4. Higher cost than other temperature transducers
5. Requires 3 or 4 wire for its operation and associated instrumentation to eliminate errors due to lead resistance

RTD's are not adaptable to applications requiring fast response or small area temperature sensing. Measurement of temperature using RTD is done after proper calibration that involves conversion of resistance value to temperature.

Most RTD instruments use a Wheatstone's Bridge or its modified version. The RTD and its leads are connected in one of its arms. This bridge is essentially a resistance measuring device which converts the resistance of the RTD into an electrical signal that is used for

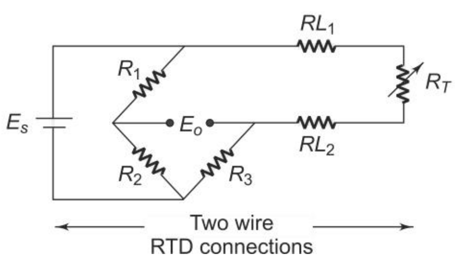


Fig. 13.39 Two wire RTD Connections

monitoring or controlling temperature. The basic Wheatstone's Bridge with a two wire RTD connected is as shown in Fig.13.39.

where E_s = Supply voltage

E_o = output voltage used for monitoring or controlling temperature

$R_1 - R_2$ and R_3 = fixed value resistors

R_T = resistance of temperature sensing element in RTD

RL_1 and RL_2 = resistance of the two leads connected in the RTD element

The value of R_T at the control point or at the mid point of the temperature range to be monitored will influence the arm resistors R_1, R_2 and R_3. Also, these values must be selected to limit bridge currents to avoid self-heating of RTD or bridge resistors. For achieving high accuracy, the bridge must be stable and insensitive to ambient temperature variations.

If the value of the lead resistance is small and also the variation of the lead resistance is small over the temperature range as compared to that of the RTD values, then the errors introduced by the lead resistance are not significant.

However, the errors introduced by long leads are significant and can be reduced by using thick lead wire. As an effective method to obtain a high degree of accuracy, it is desirable to use lead resistance compensation techniques or lead error elimination technique such as three wire and four wire RTD connections.

Two lead RTD which is lower in cost, is normally used when the lead resistance is low in comparison with Ohms/°C resistance change of the RTD or when lead wire resistance compensation is provided in the instrumentation.

Three Lead Wire RTD Three lead wire RTD, offers a practicable method for lead wire compensation that is sufficiently accurate for most industrial applications. The bridge circuit automatically compensates resistance change due to ambient temperature change, which is the input to the instrumentation. A three lead RTD connected to a Wheatstone Bridge circuit is as shown in Fig. 13.40.

This bridge circuit is used whenever lead wire resistance is significant in comparison with Ohms/°C sensitivity of the RTD elements, for example, the 10 Ω copper RTD element used in the industry should always have three leads. Other commonly used RTD elements such as 100 Ω platinum and 120 Ω Nickel may also need such compensation when lead wire resistance is significant.

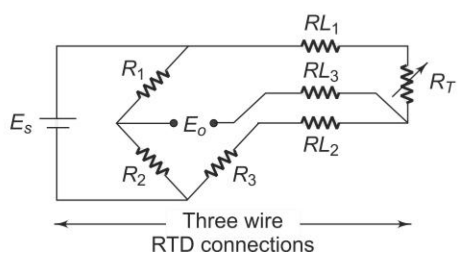


Fig. 13.40 Three wire RTD connections

Referring to Fig. 13.40, it can be seen that one of the RTD leads L_1 is in the arm of the bridge with R_T and a second lead L_2 is in the adjacent arm with R_3.

When the bridge is balanced, all the bridge arm resistance are equal, hence same current flows in all the bridge arms. The same current flows through both these leads L_1 and L_2 under balance conditions, therefore the voltage drop across them will be identical and being in the adjacent arms will effectively cancel out. Hence the effect of the lead wire is eliminated.

The third lead L_3, is connected in the output circuit of the bridge and has no effect on the bridge ratios or balance. Hence when the bridge is balanced no current flows through L_3 and therefore has no effect on the bridge balance. This method gives very good accuracy if the lead resistances are matched.

Four Lead RTD Connection The three lead RTD gives sufficient accuracy for most of the industrial applications. However, when a higher degree of accuracy and precision is required, a four lead RTD connection is used.

The four lead RTD is the most expensive type, especially when long four wire extension leads are needed to connect to the instrumentation. However the four lead RTD can offer the greatest accuracy if the instrumentation is properly designed. Four lead RTDs are widely used in laboratory work, where highest precision is required.

To select a 100 Ω Platinum RTD, the information given in Table 13.2 is helpful.

Table 13.2

1.	Temperature range	200°C–850°C
2.	Temperature coefficient in %°C at 25°C	$\alpha = 0.39$
3.	Construction	Wire wound or thin film deposited platinum
4.	Self heating	0.02°C–0.75°C/mV (typical)
5.	Lead wire	Copper two, three or four depending on the system
6.	Lead resistance compensation	Use three or four wire lead systems
7.	Accuracy	± 0.6 at 100°C
8.	Resolution	0.29–0.39 Ω/°C
9.	Drift	Approximately 0.01–0.1°C per year

Example 13.8 (b) *A platinum resistance thermometer has a resistance of 180 Ω at 20ºC. Calculate its resistance at 60ºC (α_{20} = 0.00392)*

Solution Given $R = R_o\,(1 - \alpha\,\Delta T)$

$$R = 180\,[1 - 0.00392\,(60ºC - 20ºC\,)]$$
$$R = 180\,[1 - 0.00392 \times 40ºC\,)]$$
$$R = 180\,[1 - 0.1568]$$
$$R = 180 \times 0.8432$$
$$R = 151.78\ \Omega$$

Example 13.9 *A platinum resistance thermometer has a resistance of 100 Ω at 25ºC. Find its resistance at 50ºC. The resistance temperature coefficient of platinum is 0.00392 Ω/ΩºC.*

If the thermometer has a resistance of 200 Ω, calculate the value of temperature.

Solution

Step 1: Using the linear approximations, the value of resistance at any temperature is given by,

$$R = R_0 (1 + \alpha_0 \, \Delta t)$$
$$R = 100(1 + [(0.00392) \times (50 - 25)°C])$$
$$= 100(1 + [(0.00392 \times 25)\ °C])$$
$$= 109.8\ \Omega$$

Step 2: Suppose t_2 is the unknown temperature then

$$200 = 100(1 + [(0.00392) \times (t_2 - 25)°C])$$
$$2 = (1 + [(0.00392) \times (t_2 - 25)\ °C])$$
$$2 - 1 = [(0.00392) \times (t_2 - 25)°C])$$
$$(t_2 - 25)°C = \frac{1}{0.00392} \quad \text{therefore, } t_2 = 280°C$$

13.20.3 Platinum Thin Film Sensors

Platinum thin film sensors are manufactured by a very thin layer of platinum in suitable pattern to achieve smaller dimension and higher resistance, on a ceramic base. The deposition of layers and introduction of patterns are obtained using different methods.

Normally, a number of such chips are manufactured on a single large substrate and the chips are properly cut and proper contacts are made. This platinum layer is usually coated with suitable material depending on the operating range to provide protection against mechanical and chemical damages.

The advantage of thin film sensors is reduction in the size and simultaneous increase in the nominal resistance. The response time is reduced by nearly 10 times or more because of the reduction in size. By using such a type of sensor, it is possible to manufacture a probe with a stainless steel sheath of 2 mm diameter.

The sensitivity of the sensor also increases with increase in nominal resistance. The small dimension of thin film platinum sensors allows temperature measurements in very small areas with a much higher accuracy as compared to the thermocouple.

Thin film sensors are cheaper to manufacture than wire wound RTDs, hence can be used in many fields of application for resistance thermometers such as automobiles, medical thermometer, etc.

13.20.4 Resistance Thermometer

The resistance of a conductor changes when its temperature is changed. This property is used for measurement of temperature. The resistance thermometer uses the change in electrical resistance of conductor to determine the temperature.

The requirements of a conductor material to be used in these thermometers are:

1. The change in resistance of material per unit change in temperature must be as large as possible.

2. The resistance of the material must have a continuous and stable relationship with temperature.

The main section of a resistance thermometer is its sensing element. The characteristics of the sensing element, determines the sensitivity and operating temperature range of the instrument.

The sensing element may be any material that exhibits a relatively large resistance change with the change in temperature. The material used should also have a stable characteristics, that is, neither its resistance nor its temperature coefficient of resistance should undergo permanent change with use or age.

It is necessary to consider stability in order to maintain the calibration of a resistance thermometer. The need for stability frequently limits the temperature range over which the sensing element may be used.

Another desirable characteristics for a sensing element is a linear change in resistance with change in temperature.

When the measured temperature is subjected to rapid variations, the speed with which a resistive element responds to changes in temperature is important. The smaller a given sensing element, less heat is required to raise its temperature, the faster is its response.

Platinum, Nickel and Copper are the metals most commonly used to measure temperature. The resistivity of platinum tends to increase less rapidly at higher temperatures than for other materials, hence it is a commonly used material for resistance thermometers. The temperature range over which Platinum has stability is 260 °C–1100 °C. An industrial Platinum resistance thermometer is as shown in Fig.13.41 (a).

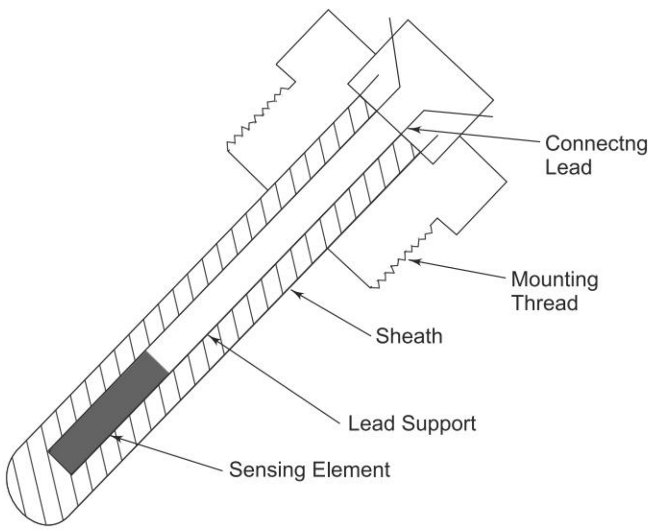


Fig. 13.41 (a) Industrial platinum ressistance thermometer

The changes in resistance caused by changes in temperature as detected by a Wheatstone's Bridge is shown in Fig. 13.41(b).

Hence, the temperature sensing element, which may be Nickel, Copper or Platinum contained in a bulb or well, along with the balancing bridge, forms the basic important components of a temperature measuring system based on this principle.

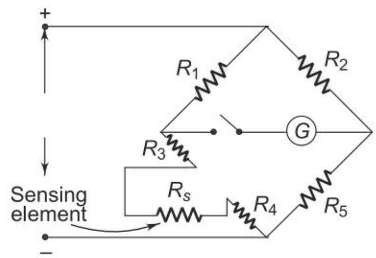


Fig.13.41 (b) Resistance thermometer connected in a bridge circuit

The sensing element R_s is made of a material having a high temperature coefficient, R_1, R_2 and R_5 are made of resistance that are practically constant under normal temperature changes.

When the sensing element is very near the bridge, and under balance conditions, the following relationship holds good.

$$\frac{R_1}{R_2} = \frac{R_s}{R_5}$$

In normal practice, the sensing element is away from the indicator and the bridge, and its leads have a resistance, say R_3, R_4.

$$\therefore \qquad \frac{R_1}{R_2} = \frac{R_s + R_3 + R_4}{R_5}$$

When resistance R_s changes, the bridge balance is upset and the galvanometer shows a deflection, which can be calibrated to give a suitable temperature scale.

Advantages of Resistance Thermometers The measurement of temperature by the electrical resistance method has the following advantages and characteristics.

1. The measurement is very accurate.
2. Indicators, recorders and controllers can also be operated.
3. More than one resistance element can be clubbed to the same indicating/ recording instrument.
4. The temperature resistance element can be easily installed and replaced.
5. The accuracy of the measuring circuit can be easily checked by substituting a standard resistor for the resistive element.
6. Resistive elements can be used to measure differential temperature.
7. Resistance thermometer have a wide working range without loss of accuracy, and can be used for temperature ranges (–200°C–650°C)
8. They are best suited for remote sensing and indication.
9. The response time of the resistive element is 2–10 s.
10. The error of the resistive element is in the range of ± 0.25% of the scale reading.
11. The size of the resistive element may be about 6–12 mm in diameter.
12. No necessity of temperature compensation.

13. Extremely accurate temperature sensing.
14. Performance stability over longer periods of time

Limitations of Resistance Thermometer

1. High cost
2. Need for bridge and power source
3. Possibility of self heating.

13.20.5 Thermistors

We have already discussed thermistors in detail in Sec. 13.8.

13.20.6 Thermocouple

One of the most commonly used methods of measurement of moderately high temperature is the thermocouple effect. When a pair of wires made up of different metals is joined together at one end, a temperature difference between the two ends of the wire produces a voltage between the two wires as illustrated in Fig. 13.42

Temperature measurement with Thermocouple is based on the Seebeck effect. A current will circulate around a loop made up of two dissimilar metal when the two junctions are at different temperatures as shown in Fig. 13.43.

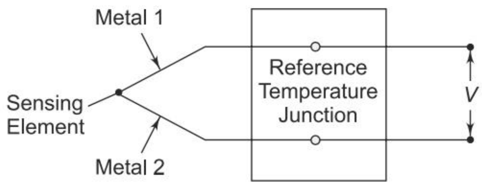


Fig. 13.42 Basic Thermocouple Connection

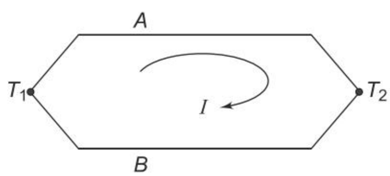


Fig. 13.43 Current through two dissimilar Metals

When this circuit is opened, a voltage appears that is proportional to the observed seebeck current.

There are four voltage sources, their sum is the observed seebeck voltage. Each junction is a voltage source, known as *Peltier emf*. Furthermore, each homogenous conductor has a self induced voltage or Thomson emf.

The Thomson and Peltier emfs originate from the fact that, within conductors, the density of free charge carriers (electrons and holes) increases with temperature.

(If the temperature of one end of a conductor is raised above that of the other end, excess electrons from the hot end will diffuse to the cold end. This results in an induced voltage, the *Thomson effect*, that makes the hot end positive with respect to the cold end.

Conductors made up of different materials have different free-carriers densities even when at the same temperature. When two dissimilar conductors are joined, electrons will diffuse across the junction from the conductor with higher electron density. When this happens the conductor losing electrons acquire a positive voltage with respect to the other conductor. This voltage is called the *Peltier emf*.)

When the junction is heated a voltage is generated, this is known as seebeck effect. The seebeck voltage is linearly proportional for small changes in temperature. Various combinations of metals are used in Thermocouple's.

The magnitude of this voltage depends on the material used for the wires and the amount of temperature difference between the joined ends and the other ends. The junction of the wires of the thermocouple is called the *sensing junction*, and this junction is normally placed in or on the unit under test.

Since it is the temperature difference between the sensing junction and the other ends that is the critical factor, the other ends are either kept at a constant reference temperature, or in the case of very low cost equipment at room temperature. In the latter case, the room temperature is monitored and thermocouple output voltage readings are corrected for any changes in it.

Because the temperature at this end of the thermocouple wire is a reference temperature, this function is known as the reference, also called as the *cold junction*.

A thermocouple, therefore consists of a pair of dissimilar metal wires joined together at one end (sensing or hot junction) and terminated at the other end (reference or cold junction), which is maintained at a known constant temperature (reference temperature). When a temperature difference exists between the sensing junction and the reference junction, an emf is produced, which causes current in the circuit.

When the reference end is terminated by a meter or a recording device, the meter indication will be proportional to the temperature difference between the hot junction and the reference junction.

The magnitude of the thermal emf depends on the wire materials used and in the temperature difference between the junctions.

Figure 13.44 shows the thermal emfs for some common thermocouple materials. The values shown are based on a reference temperature of 32°F.

The thermpcouple (TC) is a temperature transducer that develops an emf that is a function of the temperature difference between its hot and cold junctions.

A thermocouple may be regarded as a thermometer based on thermo-emf and works on the principle that the potential between two dissimilar metals or metal alloys is a function of temperature.

Type 'E' Thermocouple units use Chromel alloy as the positive electrode and constantan alloy as the negative electrode.

Type 'S' Thermocouple produces the least output voltage but can be used over greatest temperature range.

Type 'T' shown in Fig. 13.45, uses copper and constantan.

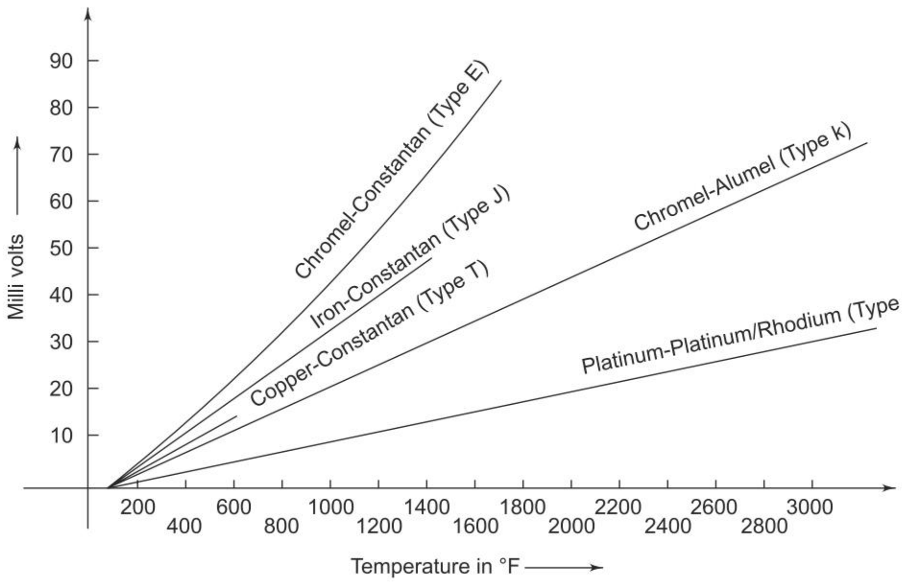


Fig. 13.44 Thermocouple output voltage as a function of temperature for various thermocouple materials

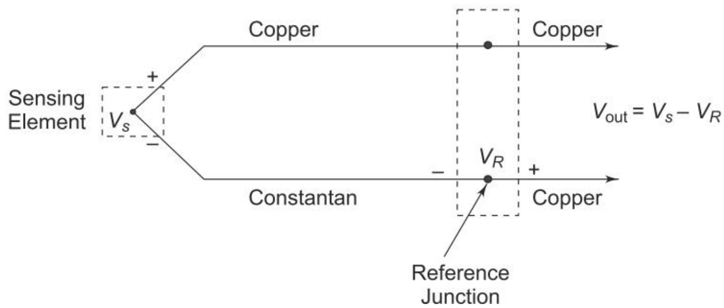


Fig. 13.45 A type t thermocouple with reference junction

Copper used, is an element and constantan used is an alloy of nickel and copper. The copper side is positive and constantan side is negative. Assuming copper wires used to connect the thermocouple to the next stage (circuit), a second Copper-Constantan junction is (formed) produced. This junction is called as the reference junction. It generates a Seebeck voltage that opposes the voltage generated by the sensing junction. If both junctions are at the same temperature, the output voltage V_{out} will be zero. If the sensing junction is at a higher temperature, V_{out} will be proportional to the difference between the two junction temperature. The temperature cannot be derived directly from the output voltage alone. It is subjected to an error caused by the voltage produced by the reference junction. This can be overcome by placing the reference junction in an ice bath to keep it at a known temperature. This process is called as *cold junction* *compensation* as shown in Fig.13.46(a). The reference voltage is maintained at 0°C. The reference voltage is now predictable from the calibration curve of the type 'T' thermocouple.

When copper is not one of thermocouple metal then four junction circuit is formed. The type 'J' thermocouple uses iron and constantan as the two elements shown in Fig.13.46(b). When it is connected to copper wires, two iron copper junctions result. These junctions present no additional difficulties because of the *isothermal block* used. This block is made of material that is a poor conductor of electricity but a good conductor of heat. Both Iron–Copper junctions will be at the same temperature and generate the same Seeback voltage and hence these two voltages will cancel. Cold junction compensation is also used as the Reference junction in this case.

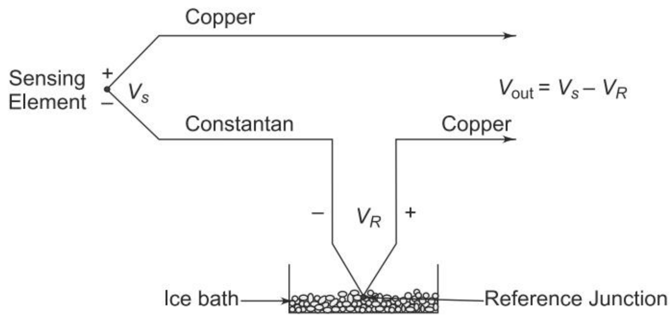


Fig. 13.46 (a) Cold junction compensation

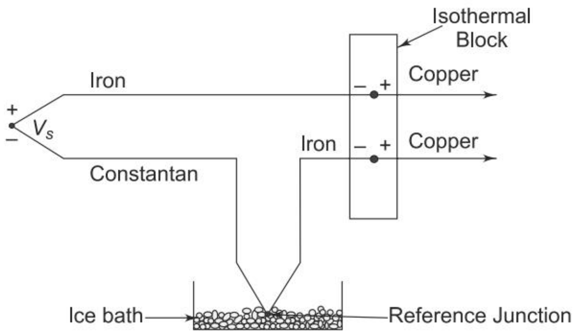


Fig. 13.46 (b) Type J thermocouple using isothermal block

The ice-baths method is not the most convenient method, to compensate the reference junction. This technique is often used in the calibration laboratory. Industry uses a different method of reference junction compensation as shown in Fig.13.47.

The isothermal block contains two reference junctions and a thermistor. The resistance of the thermistor is a function of temperature. A circuit is used to sense this resistance and to compensate for the voltage introduced by the two reference junctions. This arrangement is sometimes called as *Electronic ice point reference*.

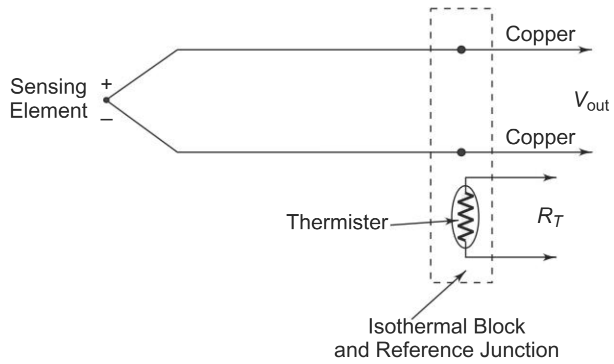


Fig. 13.47 Reference junction compensation used in industry

If the sensor is interfaced to a computer, the reference temperature will be converted to a reference voltage and then subtracted from the output voltage V_{out}. This process is known as *Software Compensation*.

An isothermal block with one temperature sensor can provide compensation for several units.

Table 13.3 gives the construction and thermoelectric properties of various thermocouples.

Table 13.3 Different Types of Thermocouples

Thermocouple type	*Materials used*	*Temperature range/ °C*	*Sensitivity µV/ °C*
Type T	Copper/Constantan	–200–400	15–60
E	Chromel/Constantan	0–850	40–55
J	Iron / Constantan	–200–900	45–57
K	Chromel/Alumel	–200–1250	40–55
R	Platinum/Platinum 13%	0–1600	5–12
	Rhodium		
S	Platinum/Platinum 10%		
	Rhodium	0–1500	5–12
B	Platinum 6% Rhodium/		
	Platinum 30% Rhodium	30–1800	0.3–0.8
G	Tungsten/Tungsten 26%		
	Rhenium	15–2800	3–20
C	Tungsten 5% Rhenium/		
	Tungsten 25% Rhenium	0–2750	10–20

For accurate measurement of hot junction temperature, the cold junction or the reference junction should be kept at 0 °C. If the reference junction is kept at the ambient temperature, then a voltage corresponding to this temperature must be added to the measurement to obtain accurate reading.

Most modern thermocouple measurement systems employ electrical cold junction (an electronic circuit which simulates the voltage that the reference junction would generate at ambient temperature) compensation. A popular technique used for reference junction compensation used in data loggers and data acquisition systems is shown in Fig. 13.48.

The measuring junction's terminals are screwed on an isothermal block (the temperature of which remains uniform within ± 0.05 °C). The temperature of the isothermal block is measured independently and compensating voltage is generated using electronic circuitry. This compensation voltage is combined with the emf from measuring junction to obtain the true temperature

Thermocouples are sometimes connected in series or parallel to provide increased voltage or current output.

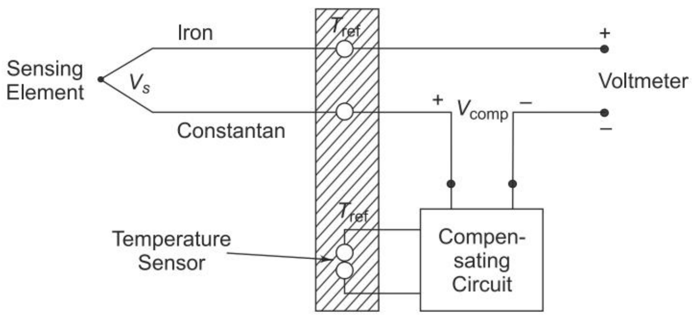


Fig. 13.48 Practical isothermal block reference junction for data loggers, etc.

In Fig.13.49(a), four thermocouples are connected in series, with wire *A* being positive and *B* being negative in each thermocouple.

The total emf between points 1 to 5 is the sum of individual thermocouple emf. An arrangement of this type is called a *Thermopile* and is used to obtain increased sensitivity and greater absolute emf from a thermocouple installation.

Figure 13.49(b) shows four thermocouples in parallel. This arrangement provides a large current but emf is same as that of any one thermocouple.

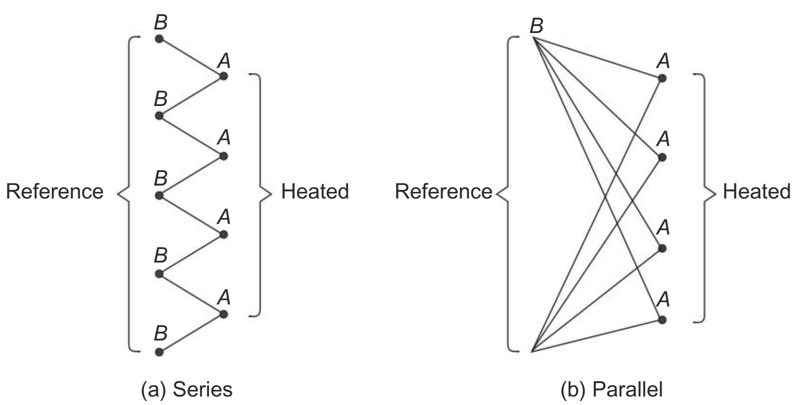


Fig. 13.49 Thermocouples in series and parallel (Thermopile)

Thermocouples must be protected from mechanical damage and isolated from corrosive or contaminating effect that most gases and liquids have at high temperature. The device used for this purpose are called *wells* or *tubes* depending upon their physical construction or *thermowells*.

Thermocouples are made from a number of different metal alloys, covering a wide range of temperature from as low as –270 °C (–418 °F) to as high as 2700 °C (about 5000 °F). They may be obtained in a simple uninsulated wire form, in insulated form or inside protective sheaths or probes (sheath diameter as small as 0.25 mm).

The thermo-junction is protected from contamination from the process materials by enclosing it in a protective sheath. For example, a cupro-nickel sheath for copper/chromel thermocouple and mild sheath for iron/chromel thermocouples.

The temperature ranges covered by thermocouples make them appropriate for use in industrial furnaces as well as for measurement in the cryogenic range. Different types of thermocouples are as shown in Fig. 13.50.

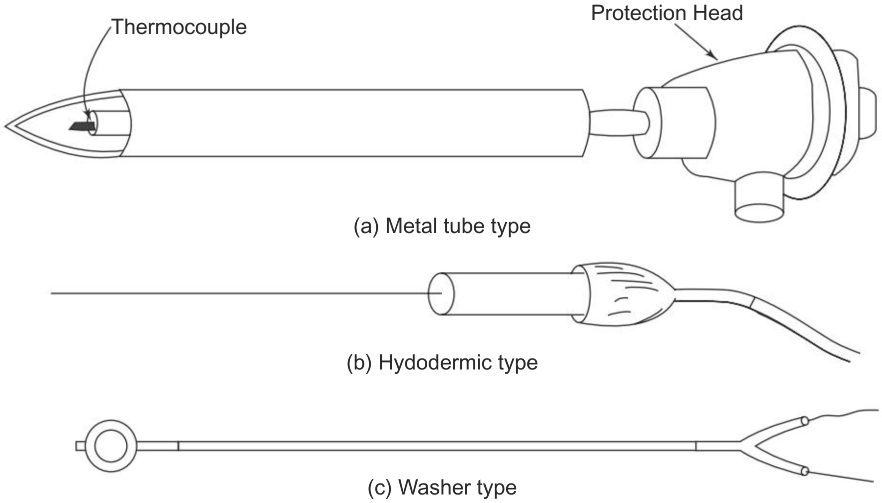


Fig. 13.50 Different type of thermocouples

Advantages of Thermocouple

1. It has rugged construction.
2. It has a temperature range from –270 °C–2700 °C.
3. Using extension leads and compensating cables, long distances transmission for temperature measurement is possible.
4. Bridge circuits are not required for temperature measurement.
5. Comparatively cheaper in cost.
6. Calibration checks can be easily performed.
7. Thermocouples offer good reproducibility.
8. Speed of response is high compared to the filled system thermometer.

9. Measurement accuracy is quite good.

Disadvantages of Thermocouple

1. Cold junction and other compensation is essential for accurate measurements.
2. They exhibit non-linearity in the emf versus temperature characteristics.
3. To avoid stray electrical signal pickup, proper separation of extension leads from thermocouple wire is essential.
4. Stray voltage pick-up are possible.
5. In many applications, the signals need to be amplified.

13.20.7 Semiconductor Diode Temperature Sensor

Semiconductor diode is a versatile device and finds use in many applications. Two of its parameter, that is forward voltage drop (V_f) and reverse saturation current (I_s) are temperature sensitive. The sensitivity of I_s with temperature is non-linear, but V_f has a linear temperature coefficient over a wide temperature range. Hence, V_f can serve as the basis for electronic thermometers

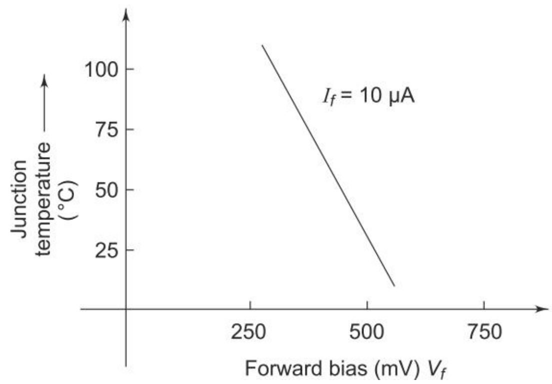


Fig.13.51 Characteristics of forward bias (V_f) versus temperature

Figure 13.51 shows the characteristics of V_f versus temperature for a typical silicon pn junction diode. The linearity of this characteristics at high values of temperature is affected by the following factors.

1. Dependence of V_f on I_s which is also temperature sensitive.
2. Presence of finite surface leakage component across the pn junction.
3. Finite resistance of the bulk semiconductor used for the diode structure.

The limitations introduced by point 1 can be overcomed by using a pair of well-matched junction transistors in a differential configuration to serve as the temperature sensors.

If the two transistors are operated at widely different emitter current values, the resulting V_{BE}, Base emitter differential voltage can serve as an excellent index of temperature.

The association of differential mode voltage gain of this stage with V_{BE} provides means of adjusting the temperature coefficient to a desired value. This is the basis of IC type temperature transducer.

13.20.8 IC Type Sensor

IC sensor produces a voltage or current signal that increases with increase in temperature. IC sensors eliminate the linearity errors associated with thermistor. However, being semiconductor devices, they are available in both voltage and current output configuration.

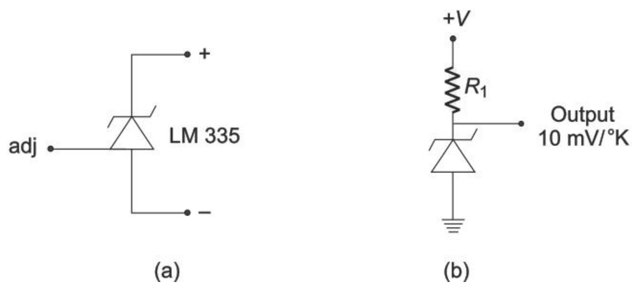


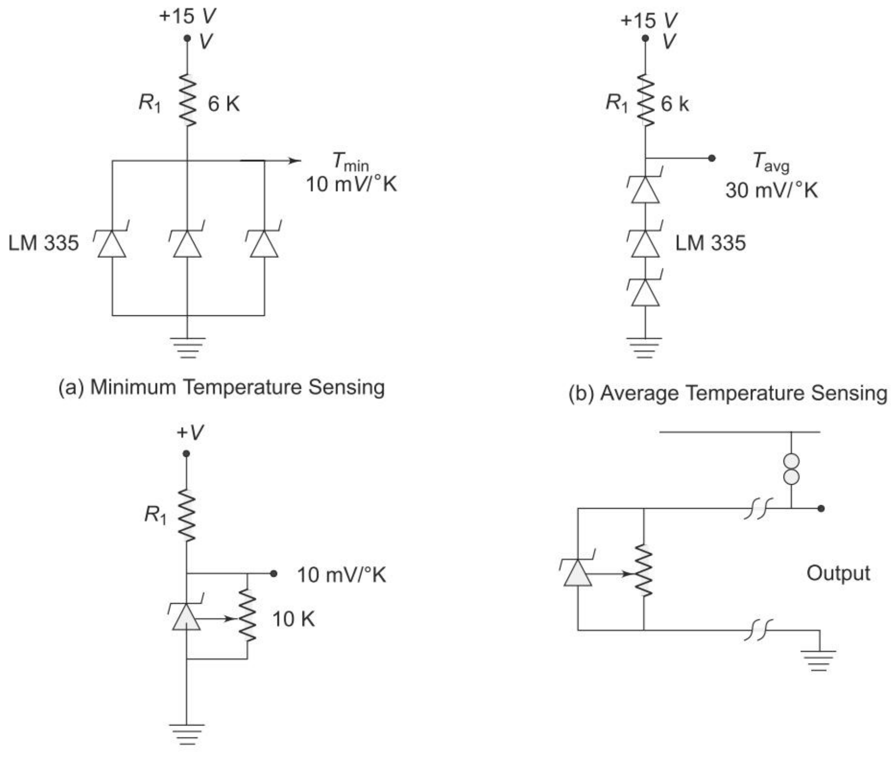
Fig. 13.52 National semiconductor LM 335 IC temperature sensor.

Figure 13.52 shows a National semiconductor LM335 IC temperature sensor. It provides a proportional output of 10 mV/°K. It operates as a two terminal zener. It has a dynamic impedance of less than 1Ω and operates over a current range of 400 μA to 5 mA with virtually no change in performance.

When calibrated at 25 °C, it typically shows less than 1 °C error over a 100° range. Its usable range is –10 °C to +100 °C and LM135 is also available with a range of –55 to 150°C.

The IC sensor can be used as a minimum temperature sensing circuit, as shown in Fig.13.53(a), because of low dynamic impedance of the sensors. The coolest sensor will set the output voltage. The average circuit using IC sensor shown in Fig.13.53(b) will simply add the individual output voltage. A simple potentiometer circuit used along with the IC sensor provides one point calibration as shown in Fig.13.53(c). The simple point calibration is usually done at the midpoint of the temperature scale. As the output of the IC Sensor LM335 is also available in the form of current source, remote sensing of the temperature using long wire length is also possible as shown in Fig. 13.53(d).

Fabricated on a single monolithic chip, they include a temperature sensor (in the form of a differential pair of transistor), a stable voltage reference and an operational amplifier.



(a) Minimum Temperature Sensing

(b) Average Temperature Sensing

(c) One point calibration (Calibration for 2.982 *V* at 25°C)

(d) Remote temperature sensor

Figs. 13.53 IC sensor

Using the internal Operational Amplifier (op-amp) with external resistors any temperature scale factor is usually obtained. By connecting the op-amp as a voltage comparator, the output will switch as the temperature crosses the set point making the device useful as an Basic on-off temperature controller as shown in Fig.13.54.

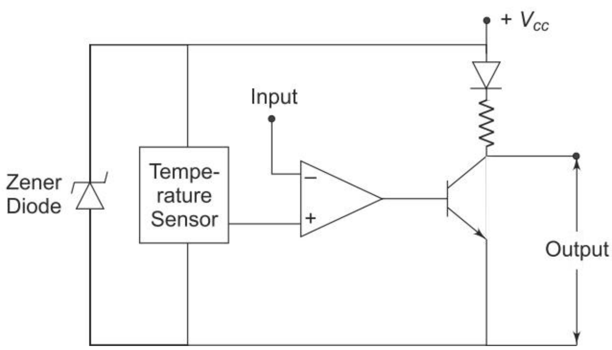


Fig. 13.54 Basic level

13.20.9 Pyrometers

When temperature being measured is very high and physical contact with the medium to be measured is impossible or impractical, optical pyrometers based on the principle of thermal radiation are used. These pyrometers are used under condition where corrosive vapours or liquids would destroy thermocouples, resistance thermometer and thermister, if made to come in contact with the measured medium.

Radiation pyrometers measures the radiant (energy) heat emitted or reflected by a hot object. Thermal radiation is an electro magnetic radiation emitted as a result of temperature and lies in the wavelength of 0.1 – 100 µm.

According to the principle of thermal radiation, the energy radiated from a hot body is a function of its temperature. Referring to Fig. 13.55, the heat radiated by the hot body is focused on a radiation detector. The radiation detector is blackened and it absorbs all or almost all radiation falling on it (if the temperature is very small compared with that of hot body, then

$$q = 5.72 \times 10^{-8} \times T^4 \text{ w/m}^2$$

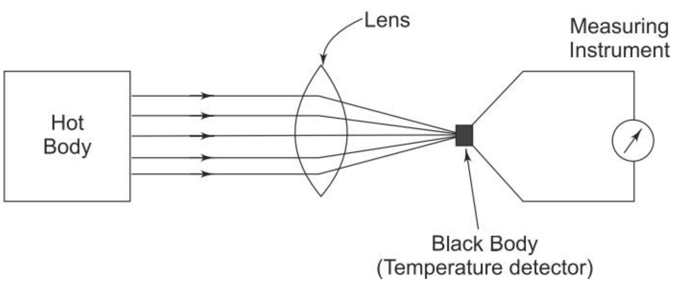


Fig. 13.55 Basic principle of pyrometer

Therefore, the heat received by the detector is proportional to the fourth power of the absolute temperature of the hot body.

Radiation pyrometers are of two types.

1. Total Radiation Pyrometers
2. Infrared Pyrometers

Both of these are discussed in the following sections.

13.20.10 Total Radiation Pyrometer (TRP)

The total radiation pyrometer receives virtually all the radiation from a hot body and focuses on a hot body and focuses on a sensitive temperature transducer such as thermocouple, bolometer, thermopile, etc. Total radiation includes both visible and infrared radiation.

The total radiation pyrometer consists of a radiation receiving element and a measuring device to indicate the temperature directly. Figure 13.56 shows a mirror type radiation pyrometer.

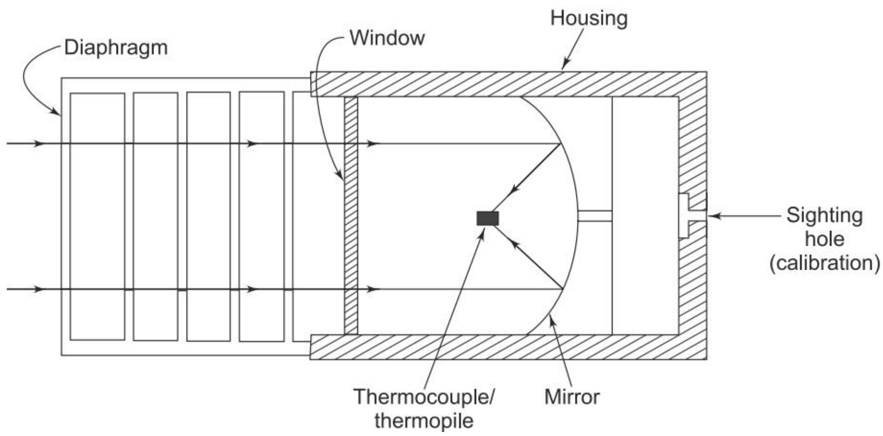


Fig. 13.56 Total radiation pyrometer

In this type of pyrometer, a diaphragm unit along with a mirror is used to focus the radiation on a radiant energy sensing transducers. The lens (mirror) to the transducer distance is adjusted for proper focus. The mirror arrangement has an advantage that since there is no lens, both absorption and reflection are absent.

Presence of any absorbing media between the target and the transducers, reduces the radiation received and the pyrometer reads low.

Due to the fourth Power Law (q is proportional to T^4) the characteristics of total radiation pyrometer are non-linear and has poor sensitivity in lower temperature ranges. Therefore, total radiation pyrometers cannot be used for measurement of temperature lower than 600 °C, since errors are introduced at lower temperatures.

Hence, total radiation pyrometers are used mostly in the temperature range of 1200 °C – 3500 °C.

The output from a total radiation pyrometers whether amplified or not, is usually taken to a PMMC instrument or to a self-balancing potentiometer. The output may be fed to a recorder or controller.

13.20.11 Infrared Pyrometers

Infrared pyrometers are partial or selective radiation pyrometers. Above temperatures of 550 °C, a surface starts to radiate visible light energy and simultaneously there is a proportional increase in the infrared energy.

Infrared principles using thermocouples, thermopile and bolometers are used. Also various types of photo-electric transducers are most commonly used for infrared transducers. The most useful transducers used for industrial application are the Photo-voltaic cells. These cells used in radiation pyrometers, respond to wavelength in infrared region and may be used to measure temperature down to 400 °C.

The infrared radiation is focused on a photo-voltaic cell as shown in Fig. 13.57. It is necessary to ensure that the cell does not become overheated. The core of radiation passing to the cell is defined by the area of the first diaphragm.

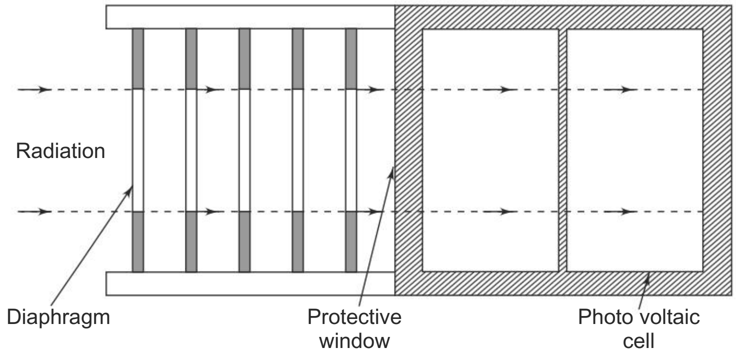


Fig. 13.57 Infrared pyrometer

The protective window is made of thin glass and serves to protect the cell and filter from physical damage. The filter is used on the range of 1000 °C to 1200 °C in order to reduce the infrared radiation passed to the photo cell. This help in preventing the photo cell from being overheated.

All infrared systems depend on the transmission of the infrared radiant energy being emitted by a heated body to a detector in the measuring system. The sensor head is focused on the object whose temperature is being measured and/or controlled.

The infrared energy falling on the detector either changes the detector resistance in proportion to the temperature as in the case of thermister or generates an emf in the detector such as a thermopile. The change in resistance or generated emf is then indicated on a meter.

13.20.12 Optical Pyrometer

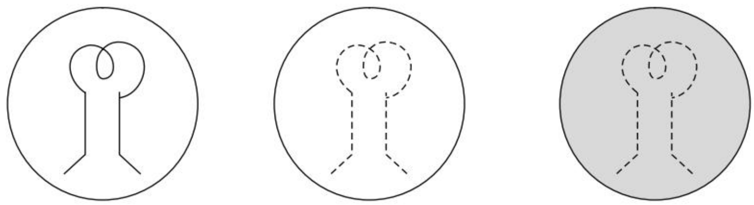
Any metallic surface when heated emits radiation of different wavelengths which are not visible at low temperature but at about 550 °C, radiations in shorter wavelength are visible to eye and from the colour approximate temperature is measured. The approximate values of temperature for colour (colour scale) is given in Table 13.4.

Table 13.4 Colour scale

Dark Red	540°C
Medium Cherry Red	680°C
Orange	900°C
Yellow	1010°C
White	1205°C

The radiations from a heated body at high temperature fall within the visible region of the EM spectrum. For a given wavelength in the visible region the energy radiated is greater than at higher temperature.

Within a visible range, a given wavelength has a fixed colour and the energy of radiation is interpreted as Intensity or Brightness. Hence if we measure the brightness of the light of a given colour emitted by a hot source, we have an indication of temperature. This is the principle of optical pyrometer.



(a) Filament colder than background (b) Filament invisible against background (c) Filament and background having equal brightness

Fig. 13.58

In an optical pyrometer, the wavelength of radiation accepted is restricted by means of a colour filter and brightness is measured by comparison with a standard lamp.

The most common type of optical pyrometer used is the disappearing filament pyrometer. The schematic is shown in Fig.13.59.

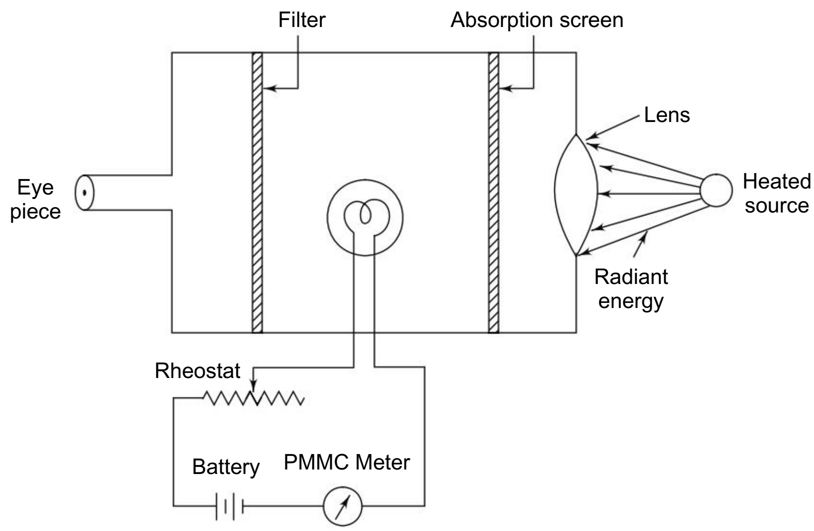


Fig. 13.59 Optical pyrometer

An image of the radiating source is produced by a lens and made to coincide with the filament of an electric lamp.

The current through the lamp filament is made variable so that the lamp intensity can be adjusted. The filament is viewed through an eye piece and filters. The current through the filament is adjusted until the filament and the images are of equal brightness.

When brightness of image produced by the source and brightness produced by the filament are equal, the outline of the filament disappears as shown in Fig.13.58(c).

However, if the temperature of the filament is higher than that required for equality of brightness, filament becomes too bright as shown in Fig.13.58(b).

On the other hand if the temperature of filament is lower, the filament becomes dark as shown in Fig.13.58(a).

Since the intensity of light of any wavelength depends upon the temperature of the radiating body and the temperature of filament depends upon the current flowing through the lamp. The instrument may be directly calibrated in terms of the filament current. However, the filament current depends upon the resistance of the filament, modern pyrometers are calibrated in terms of resistance directly.

The range of temperature, which can be measured by an instrument of this type depends on the maximum allowable temperature of the lamp which is around 1400 °C.

The range can be extended by using an absorption type screen placed near the objective lens. Hence a known fraction of radiant energy enters the pyrometer for comparison. The range can be extended to 3000 °C by this technique.

Optical pyrometers is widely used for accurate measurement of temperature of furnaces, molten metals and other heated materials.

13.20.13 Ultrasonic Temperature Transducer

Ultrasonic transducers (sound vibrations above 20 kHz) are useful when we are concerned with rapid temperature fluctuations, temperature extremes, limited access, nuclear or other severe environmental requirements and measurement of temperature distribution inside solid bodies.

They may be used to measure the distribution of parameters other than temperature (example flow). Ultrasonics also offers possibilities of remote sensing.

Ultrasonic thermometer sensors permit one to measure an extremely wide range of temperatures, from cryogenic to plasma levels and to achieve micro to milli-second response time, milli-degree resolution, greater choice of materials for sensors, operation in nuclear or corrosive environment and remote location of transducer.

Using ultrasonics for profiling permits one to obtain 2 – 10 or more temperatures using a single transmission line. This feature simplifies installation and provides reliable, accurate data at a reasonable price.

FREQUENCY GENERATING TRANSDUCER 13.21

Unlike variable parameter analog transducers, frequency and pulse generating transducers measure physical variables in terms of a pulse repetition rate.

Since frequency is an analog quantity, these can be treated as either analog or digital devices. Therefore, any phenomenon to be measured is converted into a corresponding rate of pulse generation which undergoes counting process and is made digital in nature.

These transducers produce a simple series of voltage or current pulses or cycles in proportion to the change in the physical parameter being measured. The series of pulses can be counted over a certain precisely determined period of time and the input magnitude can be known with considerable accuracy.

Examples of such devices are pulse generating pick ups operating on the variable reluctance principle, photo-tubes, etc.

RELUCTANCE PULSE PICK-UPS 13.22

This transducer is very suitable for the measurement of shaft speed and liquid flow. It is based on the principle that if the field of any magnet is varied momentarily by the motion of an external magnetic body near it, a voltage pulse is generated at the coil of the magnet because of the change in flux surrounding the coil. The transducer consists of a permanent magnet on which a coil is wound. The output voltage depends upon the rate of change of magnetic flux and the number of turns. Flux depends upon the clearance between the pick up and the actuating medium, the rate of movement and size of the actuating medium. The output voltage is inversely proportional to the distance between the head of the pick up and the actuating medium.

The pick up is actuated by the teeth of a gear or blades of turbines. In rpm measurement, the pick up is placed near the teeth of a gear. The motion of the gear tooth distorts the magnetic field around the core magnet.

A voltage pulse is produced every time a tooth enters or leaves the area of the pick up coil. The frequency of pulse is then proportional to the speed of the gear. This can be used for a tachometer.

Similarly, a pick up can be used to determine the flow, as the speed of the turbine through which the liquid flows is proportional to the flow. The total pulses counted are proportional to the total flow.

FLOW MEASUREMENT (MECHANICAL TRANSDUCERS) 13.23

The measurement of flow rate and quantity is the oldest of all measurements of process variables in the field of instrumentation. It is used to determine the amount of materials flowing in or out of a process.

Without flow measurements, plant material balancing, quality control and the operation of any continuous process would be impossible. Flow velocities are also measured by inductive transducers. The measurement of liquids containing suspended solids, such as sewage or feed to paper mills, present considerable problems. This is overcomed by the use of a flowmeter.

The transducer can be used to measure the flow of any flowing material that is electrically conductive. (The meter can be regarded as a section of pipe that is lined with an insulating material.)

Two saddle coils are arranged opposite each other and electrodes diametrically opposite are arranged flush with the inside of the lining. If the coils are energised, the moving liquid, (as a length of conductor) cuts the lines of force, resulting in the generation of an electromotive force that is picked up by the electrodes. By suitable circuitry and amplification, an electrical signal proportional to the flow can be obtained.

Many accurate and reliable methods are available for measuring flow, some of which are applicable only to liquids, some only to gases and some others to both. Fluids measured may be clear or opaque, clean or dirty, wet or dry, erosive or corrosive. Fluid streams may be multiphase, vapour, liquid or slurries. The flow may be turbulent or laminar, and viscosity and pressure may vary from vacuum to many atmospheres. Temperature may range from cryogenic to hundreds of °C.

Flow rate may vary from a few drops per hour to thousands of gallons per minute.

MECHANICAL FLOW METER 13.24

In mechanical flow meters, there is a mechanism in the path of the flow which moves continuously at a speed which is proportional to the flow rate. These are generally used for metering liquids however with certain modifications, they can be designed to meter gases also. They can be divided into two main categories, with further subgroups.

1. Displacement type
2. Inferential type

The displacement type are volumetric in operation, the cyclic displacement of the detecting element, e.g. piston, being directly proportional to the volume of the fluid passing through the meter during each cycle.

Inferential type flow meters are current type flow meters are measure the velocity of flow, from which the volume of flow is inferred.

MAGNETIC FLOW METERS 13.25

Magnetic flow meters are the first type of flowmeters to be considered for high corrosive applications and for applications involving measurement of erosive slurries. These meters work on the principle of Faraday's law of electromagnetic induction, which states that whenever a conductor moves through a magnetic field of given field strength, a voltage is induced in the conductor proportional to the relative velocity between the conductor and the magnetic field. This concept is used in electric generators. In case of flowmeters, electrically conductive flowing liquids work as the other conductor. The induced voltage is given by the equation

$$E = C \times B \times L \times V \tag{13.17}$$

where E = Induced voltage in volts
C = Dimensional constants
B = Magnetic flux in Wb/sq.m
L = Length of the conductor (fluid) in m
V = Velocity of the conductor (fluid) in m/s

The equation to convert a velocity measurement to volumetric flow rate is

$$Q = V \times A \tag{13.18}$$

where Q = Volumetric flow rate
V = Fluid velocity
A = Cross-sectional area of the flowmeter

From (13.17) $$V = \frac{E}{C\,B\,L}$$

Therefore $$Q = \frac{E\,A}{C\,B\,L}$$

For a given flowmeter, A, C, B, L are constants.

Therefore $$\text{Q} = K \times E \qquad \text{where } K = \frac{A}{C\,B\,L} \tag{13.19}$$

Therefore the induced voltage is directly proportional and linear to the volumetric flow rate.

Construction The magnetic flow meter consists of an electrically insulated or non conducting pipe, such as fibre glass, with a pair of electrodes mounted opposite each other and flush with the inside walls of the pipe, and with the magnetic coil mounted around the pipe so that a magnetic field is generated in a plane mutually perpendicular to the axis of the flow meter body and to the plane of the electrodes.

If a metal pipe is used, an electrically insulating liner is provided on the inside of the pipe.

The basic operating principle of a magnetic flowmeter in which the flowing liquid acts as a conductor is shown in Fig. 13.60.

The length L is the distance between the electrodes and equals the pipe diameter. As liquid passes through the pipe section, it also passes through the magnetic field set up by the magnet coils, thus inducing a voltage in the liquid, which is detected by a pair of electrodes mounted on the pipe wall. The amplitude of the induced voltage is proportional to the velocity of the flowing liquid. The magnetic coils may be excited by either ac or dc voltage. Currently, pulsating dc in which magnetic coils are periodically energised is used.

Magnetic flow meters are available in sizes from 2.54 – 2540 mm in diameter, with an accuracy range of ± 0.5 to ± 2%. The measurement taken by these meters are independent of viscosity, density, temperature and pressure. (The range of such meters may be 30 :1, but normally a 20:1 range is accepted.)

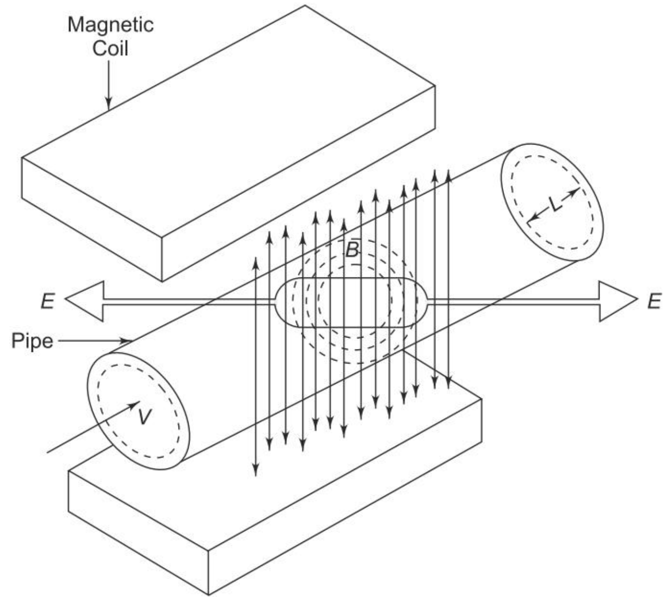


Fig. 13.60 Working principle of magnetic flowmeter

A modern design of magnetic flowmeter is one which can be inserted into the line through couplings. It consists of electrodes mounted on each side of a probe and magnetic coils which are also integral to the probe. The probe can be mounted on pipes of diameters 152.4 mm and above can easily be mounted for open channel flow.

Advantages of Magnetic Flowmeter

1. It can handle slurries and greasy materials.
2. It can handle corrosive fluids.
3. It has very low pressure drop.
4. It is totally obstructionless.
5. It is available in large pipe sizes and capacity as well as in several construction materials.
6. It is capable of handling low flows (with minimum size, less than 3.175 mm inside diameter) and very high volume flow rate (with sizes as large as 3.04 m).
7. It can be used as bidirectional meter.

Disadvantage of Magnetic Flowmeter

1. It is relatively expensive.
2. It works only with fluids which are adequate electrical conductors.
3. It is relatively heavy, especially in larger sizes.
4. It must be full at all times.
5. It must be explosion proof when installed in hazardous electrical areas.

TURBINE FLOWMETER 13.26

The turbine flowmeter is used for the measurement of liquid gas and gases of very low flow rate. It works on the principle of turbine. It consists of a multibladed rotor (called turbine wheel) which is mounted 90° to the axis of the flowing liquid as shown in Fig. 13.61.

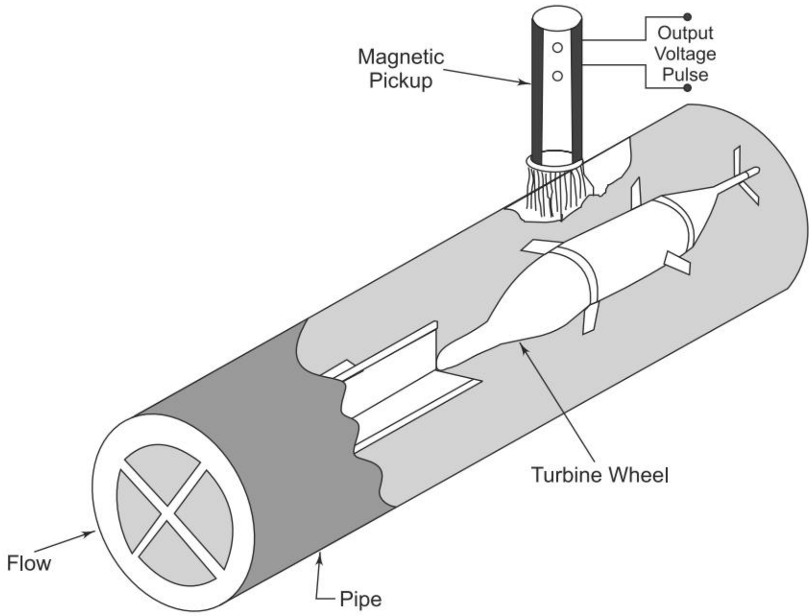


Fig. 13.61 Turbine flowmeter

The rotor is supported by the ball or sleeve bearings on a shaft which is retained in the flow meter housing by a shaft support section. The rotor is free to rotate about its axis.

The flowing liquid strikes the turbine blades (rotor) imparting a force to the blade surface which causes the rotation of the rotor. At a steady rotational speed, the speed of the rotor is proportional to the fluid velocity and hence to the volumetric flow rate. The speed of rotation is monitored by a magnetic pick up which is fitted to the outside of the meter housing.

The magnetic pick-up coil consists of a permanent magnet with coil windings which is mounted in close proximity to the rotor but internal to the fluid channel. As each rotor blade passes the magnetic pick-up coil, it generates a voltage pulse which is a measure of the flow rate. The total number of pulses gives a measure of the total flow.

The electrical voltage pulses produced can be totalled, differenced or manipulated by digital techniques, so that a zero error characteristic, using this technique, is provided from the pulse generator to the final read.

The number of pulses generated per gallon of flow, called the *K* factor is given by

$$K = \frac{T_k f}{Q} \tag{13.20}$$

where K = pulses per volume unit
T_k = time constant in minutes
f = frequency in Hz
Q = volumetric flow rate is gpm (gallon per minute)

Turbine flow meters provide very accurate flow measurement over a wide flow range. The accuracy range is ± 0.25 to ± 0.5% with excellent repeatability, i.e. precision ranging from ± 0.25% to as good as ± 0.02%.

The range of turbine meters is generally between 10 : 1 and 20 : 1; however, in a low flow rate it is less than 10 : 1.

Military turbine meters have a range greater than 100 : 1.

Turbine meters are available in different sizes, ranging from 6.35 – 650 mm and liquid flow ranges from 0.1 – 50 000 gallons per minute.

These turbine flow meters are mostly used for military applications. They are also used in blending systems in the petroleum industry. They are effective in aerospace and airborne applications for energy fuel and cryogenic (liquid oxygen and nitrogen) flow measurements.

Advantages of Turbine Flowmeter

1. Good accuracy
2. Excellent repeatability and range
3. Fairly low pressure drop
4. Easy to install and maintain
5. Good temperature and pressure ratings
6. Can be compensated for viscosity variations

Disadvantages

1. High cost (expensive)
2. Limited use for slurry applications
3. Problems caused by non-lubricating fluids

MEASUREMENTS OF THICKNESS USING BETA GAUGE 13.27

A common and characteristic feature of a radioactive element is that they disintegrate spontaneously to produce fresh radioactive elements called daughter elements. This activity of parent elements is termed as radioactivity.

During this disintegration, high energy radiations are emitted from the nuclear of the radioactive element. These high energy radiations are mostly of three types: the positively charged rays called alpha (α) rays, the negatively charged rays called beta (β) rays, and the neutral rays called gamma (γ) rays. β rays are usually, but not necessarily accompanied by (γ) rays.

Some of the basic and important properties of α, β, and γ-rays are listed in Table 13.5.

Table 13.5

1. Alpha (α) rays	(a)	They consists of positively charged particles.
	(b)	They are emitted from the nucleus of a radioactive atom with a high velocity.
	(c)	They possess high energies.
	(d)	Since α-particles produce intense ionisation, they lose energy quickly and are therefore, stopped within a short distance.
	(e)	Since they carry a positive charge, they can be deflected by electric and magnetic fields.
2. Beta (β) rays	(a)	They consists of negatively charged particles.
	(b)	They are emitted from the nucleus of radioactive atoms with very high velocities.
	(c)	They possess less energies as compared to α-rays.
	(d)	Since they produce only moderate ionisation, they do not lose energy very rapidly.
	(e)	As they carry a negative charge, they can be easily deflected by electric and magnetic fields.
3. Gamma (γ) rays	(a)	They are not made up of any charged particles at all.
	(b)	They are emitted from the nucleus of radioactive atom as high energy photons, i.e. they are electromagnetic waves of very short wavelength.
	(c)	Since they produce only feeble ionisation, they do not loose energy very fast.
	(d)	Since they do not carry any charge, they cannot be deflected by electric or magnetic fields.

Alpha rays are not used in radiometrical measurement techniques because their range is only some mm, even in air. Hence this leaves only beta and gamma radiations.

Measuring Principle The measurement of the $\dfrac{\text{weight}}{\text{unit area}}$ of the sheet material is accomplished by measuring the absorption of radiation emanating from a radio isotope when passing through the material. If the material is homogeneous and its density is known, the wt/unit area measurement can be transformed into a thickness measurement.

Mode of Operation Figure 13.62 shows a schematic block diagram of a beta gauge.

The source (1), enclosed in a source holder, is positioned to one side of the material to be measured (2) The measuring chamber (3), containing an ionisation chamber as the detector, is located exactly opposite the source, on the other side of the material. Part of the radiation from the source is absorbed during the passage through the material, and the remaining radiation ionises the gas in the ionisation chambers. A dc converter (5) supplies the anode of the ionisation chamber with high voltage. The resulting field causes an ionisation current I_X, proportional to the intensity of the received radiation. This current is collected by the measuring electrode.

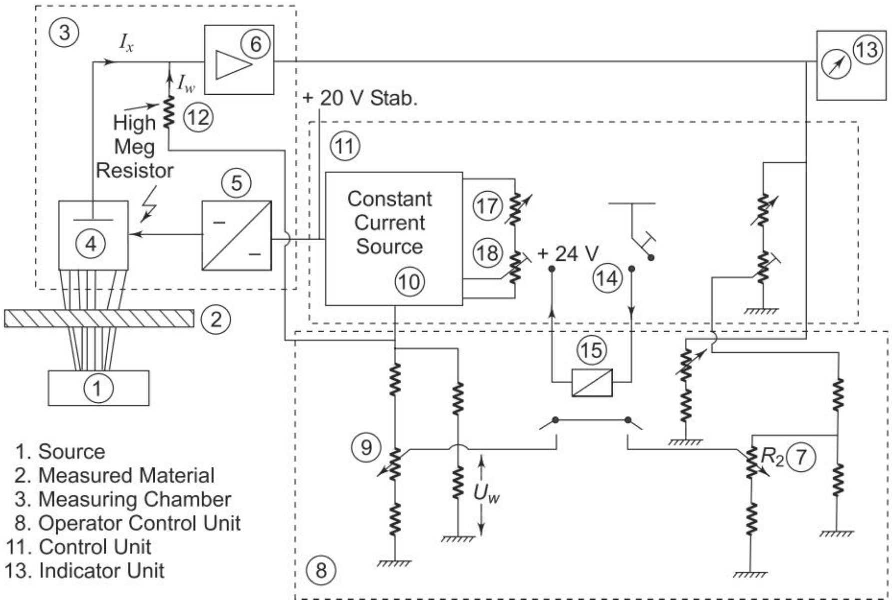


Fig. 13.62 Schematic block diagram of a beta gauge

The capacitor-diode amplifier (6), is a dc amplifier with high input impedance and low leakage. In its feedback loop, the voltages from the feedback divider (7), located in the operator control unit (8) and the target value adjust (9), are effective. The target value adjust is calibrated such that the voltage applied to the high-meg resistor of the measuring chamber causes a current I_w equal and opposite in polarity to the ionisation current I_x. If the measured wt/unit area or thickness of the material in the measuring gap is on target, the voltage at the amplifier output connected to the deviation indicator input goes to zero.

Deviations of the measured product from the target cause the indicator to move towards plus or minus proportional to the magnitude of the error.

Full radiation is used for balancing the system, i.e. no material in the measuring gap. By pushing the full radiation button (14) in the control unit (11), the feedback loop is connected via a relay (15) to a compensation voltage equivalent to zero weight or no material (product) in the measuring gap. The deviation indicator (13) should be moved to zero by actuating the coarse switch and adjusting the fine potentiometer. This adjusts both the feedback voltage and current I_w to match the ionisation current.

Measuring Chamber The measuring chamber is utilised for measuring base weight or thickness by absorption of radioactive radiation.

It is dust tight, with ambient temperature ranges from + 10°C – +50 °C. The influence of the temperature of the air in the measuring gap on the measurement is compensated for by the temperature compensation circuit. The ionisation current causes a voltage drop across the high-meg resistors. The difference between this voltage drop and the compensation voltage is amplified by a two-stage capacitor-diode amplifier.

Control Unit The control unit supplies the voltage for the measuring chamber. It contains the master switch and the elements for balancing are equipped with automatic temperature compensation, i.e. compensation of wt/unit area caused by temperature changes of the air in the measuring gap.

This system requires the following voltages: 24 V ac and +24 V dc stabilised. 24 V ac is tapped from a separate winding of the power transformer. A separate winding also supplies the circuit protected stabilised power supply + 24 V dc and 0.7 A.

A constant current source supplies the operator control unit. The target value potentiometer in this unit sets the comparator voltage for the measuring chamber. Since the constant current source must be floating, an electronic chopper is employed to supply its regulator.

The temperature compensation circuit, if supplied, adjusts for temperature related changes of wt/unit area of the air in the measuring gap. It does this by feeding an error voltage to the constant current source. This voltage is generated by an IC amplifier in a bridge circuit one of whose branches contains temperature sensors (NTC) of the measuring chamber and the source holder.

Balancing, i.e. compensation for a source decay, is accomplished by adjusting the constant current and the feedback voltage derived from the measured value to match the compensation voltage of the measuring chamber proportional to the source decay.

13.27.1 Control Sequence of the Motion of the Measuring Heads

The control of the measuring head motion is possible in various program sequences such as (i) automatic profile scanning, (ii) traversing scanning, (iii) jugging, (iv) off-sheet position, and (v) test position.

1. Automatic Profile Scanning After pressing the profile push button, the measuring head leaves the off-sheet position E_a, as shown in Fig. 13.63, passes the limit switch at the left side sheet edge E_b and stops at the permanent measuring position E_d. Presetting this button again, or after making contact with the preset time switch mechanism initiates the automatic profile scanning. The measuring head moves at high speed to the right side E_r, then back to the left side edge at low speed and returns at high speed to the center positions E_d. It records only at low speed.

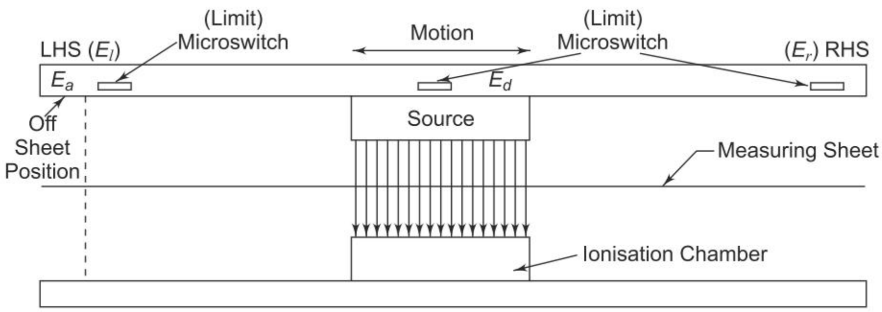


Fig. 13.63 Control sequence of movement of the head

2. Traversing This program constitutes a pendular movement of the measuring head, between two limit switches E_l and E_r, whereby the traversing speed from the right side to the left side can be determined by converting the bridges, released by pressing the traverse button.

3. Jugging At jugging, an optimal positioning of the measuring head between the limit switches E_a and E_r is possible.

4. Off-sheet Movement Off-sheet movement to the limit switch E_a can be effected by various command inputs, either by pressing the push-button off-sheet movement at the control panel, or by means of the adjustment command from the power supply unit.

5. Test Position This limit switch is situated outside the right side turning point E_r and can be reached by the Test Position button.

Review Questions

1. Define a transducer.
2. Explain the difference between primary sensors and transducers with the help of examples.
3. With the help of examples, explain the procedure of converting displacement response of some primary sensors into electrical output.
4. What do you understand by electrical transducers? State the advantages of an Electrical transducer.
5. State the various parameters of electrical transducers.
6. List five physical quantities that the transducer measures.
7. List different types of transducers.
8. List the factors to be considered while selecting a transducer.
9. Explain the difference between primary sensors and transducers with the help of examples.
10. What is the difference between passive and active transducers.
11. Differentiate between primary and secondary transducers.
12. Explain with diagram the functions of a resistive transducer.
13. Explain with diagram potentiometer used as a transducer.
14. Describe with the help of a diagram, the measurement of displacement using potentiometers. State the

advantages and disadvantages of a potentiometer.

15. Explain with a diagram the operation of a resistive pressure transducer.
16. Explain with a diagram the operation of a resistive position transducer.
17. Define a strain gauge. Define gauge factor.
18. Describe the operation and construction of strain gauge. State it's limitations.
19. Explain how temperature compensation can be achieved in strain gauge.
20. Explain with diagram the working of a resistance wire gauge.
21. Explain with diagram the working of an unbonded strain gauge.
22. Explain in brief bonded strain gauge.
23. List different types of strain gauges.
24. Explain with the help of diagram the various bonded strain gauges.
25. Distinguish between unbonded and bonded strain gauges.
26. Why are strain gauges used in bridge arrangement?
27. State the various desirable characteristics of resistance wire strain gauge.
28. Describe with a diagram and construction a foil type strain gauge.
29. What is the difference between resistance wire and foil type strain gauge.
30. Explain with a diagram the operation of a semiconductor strain gauge. State the advantages and disadvantages of a semiconductor strain gauge.
31. Under what condition is a dummy strain gauge used? What is the function of the gauge?
32. Describe with the help of a diagram the working of a load cell. State its applications
33. What is the effect of temperature changes on a strain gauge.
34. State the most common method of temperature compensation overcoming the above difficulty.
35. State the main advantages and disadvantages of semiconductor strain gauge compared to the metallic wire strain gauge.
36. Define inductance, self-inductance and mutual inductance.
37. How can inductors be used as a transducer.
38. State the advantages of using differential output rather than a single output for measuring displacement.
39. Define permeability and reluctance.
40. State the difference between a self-generating and a passive inductive transducer.
41. List different methods of varying self-inductance.
42. Describe the operation of change in self-inductance due to change in number of turns.
43. Describe the operation of change in self-inductance due to change in permeability.
44. Describe the operation of change in self-inductance due to change in reluctance.
45. Explain with the help of a diagram the method of measurements of displacement using change in self-inductance due to
    (a) change in number of turns
    (b) change in permeability
    (c) change in reluctance.
46. Describe the operation of a differential output. State the advantages of differential output.
47. Describe with the help of a diagram the construction of an LVDT.
48. Explain with the help of a diagram and characteristics the operation of LVDT.
49. Explain the method of measuring displacement using LVDT. State the advantages and disadvantages of LVDT.
50. Explain the principle of RVDT.
51. Describe the working and construction of a rotational variable differential transducer (RVDT).

52. List four types of electrical pressure transducers and describe one application of each one.
53. Describe the principle of operation of a pressure transducer employing each of the following principles:
    (a) Resistive transducer
    (b) Inductive transducer
    (c) Capacitive transducer
54. State different types of photoelectric transducers.
55. Describe with the diagram the operations of a piezo-electric transducer.
56. State the differences between photoemissive, photo conductive and photovoltaic transducers.
57. Explain with a diagram the operation of a photo multiplier tube.
58. Explain with a diagram and construction the operation of a photo cell.
59. Explain in brief the working of a solar cell. State its applications.
60. Explain the operation of a photodiode and phototransistor. State its applications
61. Define temperature.
62. List various types of temperature transducers and describe the applications of each.
63. Differentiate between thermistor and thermocouple.
64. Explain the working principle of a resistance temperature detector.
65. State the different elements used as a sensor in RTD. Explain each in brief.
66. Explain with diagram the operation of 2 wire, 3 wire and 4 wire RTD.
67. Describe the different techniques used to eliminate the lead wire effect of RTD connection in a bridge circuit.
68. Explain with diagram the operation of a 3 wire RTD.
69. Describe in brief the operation of a 4 wire RTD.
70. Explain the operation of thin film RTD.
71. Compare a thin film RTD with wire wound RTD.
72. Compare RTD with thermistor.
73. State the advantages and limitations of resistance thermometers. Describe the operation of a resistance thermometer.
74. Explain the working principle of thermistors.
75. Describe different types of thermistor. State the advantages and disadvantages of thermistors.
76. State various applications of a thermistor.
77. Explain the working principle of thermocouple
78. Explain Seebeck effect and Peltier emf.
79. Describe the construction of thermocouples. State the limitations of a thermocouple.
80. Explain the term reference junction.
81. State the error introduced in a thermocouple. Describe the various methods to compensate these errors.
82. Explain electronic ice point reference junction.
83. List various types of thermocouples.
84. Explain thermopiles and thermowell.
85. State the advantages and disadvantages of thermocouples.
86. Explain with a diagram, semiconductor temperature sensor and IC sensor.
87. State the various of applications of IC sensor.
88. Define pyrometers.
89. Explain the working principle of pyrometers.
90. Describe with a diagram the operation of total radiation pyrometers.
91. Describe with a diagram the operation of infrared pyrometers.
92. Explain the working principle of an optical pyrometer.
93. Describe with a diagram the operation of optical pyrometer.

94. It is proposed to measure the temperature of an incandescent lamp, suggest a suitable sensor for the same. Justify your answer.
95. Choose the most suitable temperature transducer for measuring the temperature in each of the following:
    a. Rapidly changing temperature
    b. Very small temperature changes about 40°C
    c. Very high temperature (>1500°C)
    d. Highly accurate temperature measurement
    e. Wide temperature variations
96. Explain in brief an ultrasonic temperature transducer.
97. Explain the operation of a reluctance pulse pick-up.
98. Describe the principle of operation of an accelerator.
99. Define flow.
100. Explain the working principle of a variable area flow meter.
101. Describe the operation of a variable area flow meter.
102. Explain the term orifice.
103. Explain the working principle of a magnetic flow meter.
104. Describe with a diagram the operation of a magnetic flowmeter. State the advantages and disadvantage of a magnetic flowmeter.
105. State the applications of magnetic flowmeter.
106. Explain the operation of a impeller type flowmeter.
107. Compare magnetic flowmeter and impeller flow meter.
108. Describe with a diagram the operation of a turbine flowmeter.
109. Compare magnetic flowmeter and turbine flow meter.
110. How can the thickness of a sheet material be measured?
111. Explain with a diagram the operating principle of a beta gauge.
112. Describe with a diagram the operation of a beta gauge.
113. Explain the various scanning modes of a beta gauge.
114. List the series of steps that you would take if you were required to select a transducer for particular measurement applications.

Multiple Choice Questions

1. RTD is a/an
   (a) active transducer
   (b) passive transducer
   (c) inductive transducer
   (d) capacitive transducer
2. Compared with thermistor, the RTD is
   (a) less sensitive but more stable
   (b) more sensitive but less stable
   (c) less sensitive but less stable
3. In three-wire RTD bridge,
   (a) one lead each is in the two adjacent arms and the third lead goes to detector
   (b) one lead each in two opposite arms and the third goes to the detector
   (c) one lead is in the RTD arm while the other two go to source and detector respectively
4. The metal mostly used in RTD is
   (a) copper (b) iron
   (c) platinum (d) silver
5. A thermistor is inherently suitable for temperature measurement involving
   (a) high span but low sensitivity
   (b) high span and high sensitivity
   (c) low span but high sensitivity
6. A thermocouple is
   (a) two similar metals connected together
   (b) two dissimilar metals connected together

(c) two wire wound resistors connected together
(d) two inductive coils connected together

7. The transducer suited for high temperature and fast response time is a
(a) thermistor (b) RTD
(c) thermocouple

8. For high temperature near to 1500°C, the suitable thermocouples used is
(a) copper/constantan
(b) iron/constantan
(c) chromel/constantan
(d) platinum-rhodium/platinum

9. To measure temperature in the range of 3000°C, the following type of sensor is used:
(a) RTD (b) thermocouple
(c) thermistor (d) pyrometers

10. Semiconductor diode parameter sensitive to temperature is
(a) forward voltage drop
(b) reverse voltage drop
(c) forward saturation current

11. Which of the following should be incorporated in an RTD to make a temperature sensing bridge most sensitive?
(a) platinum (b) nickel
(c) copper (d) thermister

12. In a thermocouple element, heat energy transferred to the hot junction is converted into electrical energy by
(a) Johnson's effect
(b) Seebeck effect
(c) Hall effect
(d) Faraday's effect

13. A thermocouple is to be used to measure temperature in the range of 700°C to 800°C .Select the pair that would be most suitable for the applications.
(a) copper–constantan
(b) iron–constantan
(c) chromel–alumel
(d) platinum–platinum + rhodium

14. Rotametre is a flowmetre based on
(a) variable area
(b) variable length
(c) variable pressure

15. Magnetic flowmeters work on the principle of
(a) electrostatics
(b) pressure
(c) electro-magnetism

16. A magnetic flowmeter measures the flow of
(a) petroleum production
(b) gases
(c) solids
(d) slurries

17. Impeller flow meter measures
(a) liquids
(b) gases
(c) petroleum products
(d) slurries

Further Reading

1. Alexander Schure, *Industrial Electronic Measurements*, John F. Rider Publisher, 1964.
2. Rufus P. Turner, *ABC's of Thermistor*, Foul Shams Tech Books, W Foulsham & Co., 1970.
3. S.K. Singh, *Industrial Instrumentation & Control*, Tata McGraw-Hill, 1987.
4. Larry Jones and A. Foster Chin, *Electronic Instruments and Measurements*, John Wiley & Sons, 1987.
5. B.S. Sonde, Transducers and Display System, TMH, 1979.
6. R.K. Jain, Mechanical Industrial Measurements, Khanna Publishers, 1983.
7. Sol. D. Prensky, Advanced Electronic Instruments and Their Use, Hayden Book Co. 1970.

Chapter 14
Signal Conditioning

INTRODUCTION 14.1

The measurand, which is basically a physical quantity, is detected by the first stage of the instrumentation or measurement system. The first stage is the detector transducer stage. The quantity is detected and converted or transduced into an electrical form in most cases. The output of the first stage has to be modified before it becomes usable and sufficient to drive the signal indicating stage which is the last stage. This last stage may consist of indicating, recording, displaying and data processing elements, or may consist of control elements.

In an electronic-aided measurement, the quantity to be measured is converted into an electrical signal and then amplified or otherwise modified to operate a device which displays the numerical value of the measured quantity. This process is illustrated for a typical case in the pictorial block diagram shown in Fig. 14.1.

An input device such as a photo detector, thermistor bridge, glass pH electrode or strain guage circuit is used to convert the quantity to be measured into an electrical signal. Some characteristic of this signal is related in a known way to the quantity to be measured. The measured quantity is now encoded as an electrical signal. The electrical signal from the input device is then modified by suitable electronic circuit to make it suitable for operating a read-out device. The electronic circuit is frequently an amplifier with the appropriate adjustable parameters (zero, standardisation, position, etc.) and sometimes with automatic compensation for non-linearities, temperature variation, etc. of the input device. It is a truism that the end result of any measurement is a number. (An electrical signal is an electrical quantity or a variation in an electrical quantity that represents information).

The output number is obtained from a readout device such as a meter or chart recorder, where the position of a marker against a numbered scale is observed, or from lights in the shape of numbers which are turned ON.

In Fig. 14.1 the measured quantity was encoded in at least three different ways. First, as a physical or chemical quantity or property, second, as some characteristic of electrical signal, and finally as a number. Each way in which data can be encoded is called a data domain. (Measurement data are represented in an instrument at any instant by a physical quantity, a chemical quantity, or some characteristic of an electrical signal. Each different characteristic or property used to represent data is called a data domain).

Measurement of dynamic physical quantities requires faithful representation of their analog or digital output obtained from the intermediate stage (signal conditioning stage) and this places a severe strain on the signal conditioning equipment.

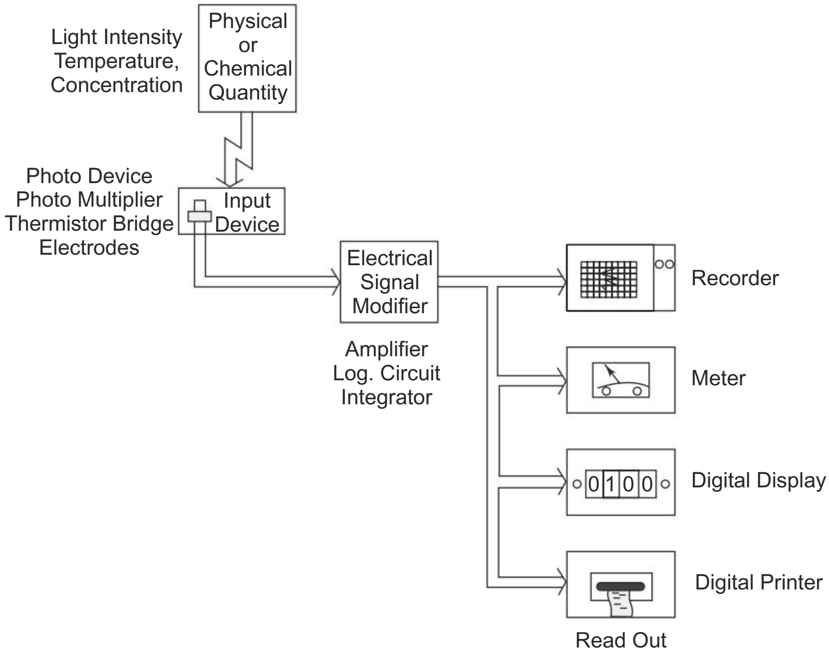


Fig. 14.1 Block diagram of an electronic-aided measurement

The signal conditioning equipment may be required to perform linear processes such as amplification, attenuation, integration, differentiation, addition or subtraction. They are also required to do non-linear processes such as modulation, demodulation, sampling, filtering, clipping and clamping, squaring and linearising or multiplication by another function. These functions require proper selection of components and faithful reproduction of the final output for the presentation stage.

The signal conditioning or data acquisition equipment is in many cases an excitation and amplification system for passive transducer, or an amplification system for active transducer. In both cases, the transducer's output is brought to the required level to make it useful for conversion, processing, indicating and recording.

Excitation is needed only for passive transducers, because they do not generate their own voltage or current. It is essential for passive transducers like strain guage, potentiometers, resistance thermometers, and inductive or capactive transducers to be excited from an external source.

Active transducers such as thermocouples, piezo-electric crystal and inductive pickups etc. do not require an external excitation because they produce their own voltages only by the application of physical quantities. However these signals are at a low level, and require amplification.

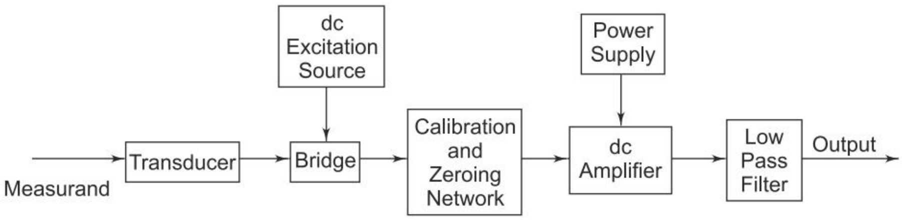


Fig. 14.2 DC signal conditioning system

The excitation source may be ac or dc. A simple dc system is shown in Fig. 14.2.

The strain guage (resistance transducer) constitutes one or more arm of the Wheatstone's bridge, which is excited by the dc source. The bridge can be balanced by using a potentiometer and can also be calibrated to indicate the unblanced condition. The dc amplifier should have the following characteritics.

1. Its input stage may be a balanced differential inputs giving a high CMRR.
2. It should have extremely good thermal and long term stability.
3. Easy to calibrate at low frequency.
4. Able to recover from an over load condition, unlike an ac system.

The main disadvantage of a dc amplifier is the problem of drift. Hence low frequency spurious signals are available as data at the output and to avoid this low drift of a dc amplifier, special low drift dc amplifiers are used.

The dc amplifiers is followed by a low-pass filter, which is used to eliminate high frequency components or noise from the data signals.

In order to overcome the problem of drift in the dc systems, ac systems are used. Figure 14.3 shows a circuit of an ac system using carrier type ac signal conditioning system.

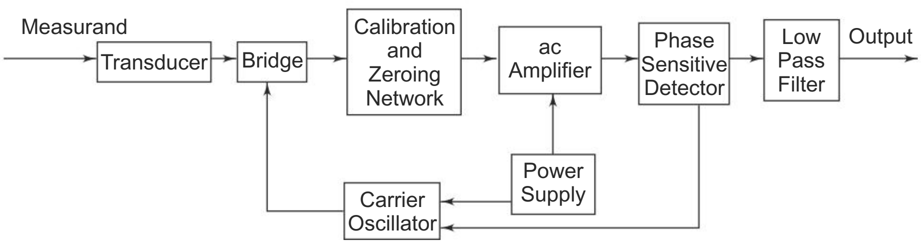


Fig. 14.3 AC signal conditioning system

The transducers used are of the variable resistance or variable inductance type. They are employed between carrier frequencies of 50 kHz and 200 kHz. The carrier frequencies are much higher, at least 5 to 10 times the signal frequencies.

The output of a transducer is applied to the bridge circuit, whose output is an amplitude modulated carrier signal. This waveform is amplified by an ac amplifier. This amplified modulated output is then applied to a phase sensitive demodulator, the carrier signal. This produces a dc output that indicates the direction of the parameter change in the bridge output.

In a carrier system amplifier, frequency drift and presence of spurious signals are not much of a problem. However it is more difficult to obtain a stable carrier oscillator than a dc stabilised source. Active filters can be used to reject this frequency and prevent overloading of the ac amplifier. The function of the phase sensitive detector is to filter out the carrier frequency components of the data signal.

To summarise, dc systems are generally used for common resistance transducers (such as potentiometers and resistance strain gauges).

AC systems are used for variable reactance transudcers and for systems where signals have to be transmitted via long cables to connect the transducers to signal conditioning equipment.

After physical quantities like temperature, pressure, strain, acceleration, etc. have been transduced into their analogous electrical form and amplified to sufficient current and voltage levels (say 1 – 10 V), they are further processed and the amplified signal applied directly to the indicating, recording or control systems.

The signal may be applied to a Sample-Hold (S/H) circuit, as shown in Fig. 14.4. This may be fed to an analog multiplexer and A/D converter. If the signal is in digital form, it may be applied to a variety of digital systems, such as computer, digital controller, digital data loggers, or transmitters.

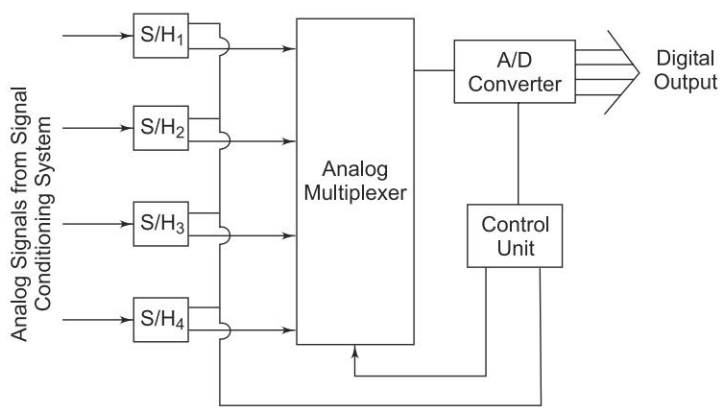


Fig. 14.4 Data acquisition and conversion system

The Sample-Hold unit shown in Fig. 14.4 samples the various inputs at a specified time and holds the voltage levels at their output while analog multiplexer performs the Time Division Multiplexing (TDM) operation between different data inputs.

TDM means each channel is sequentially multiplexed for a certain specified time, i.e. each channel is time shared. The timing of various input channels is controlled by a control unit. This unit controls the S/H circuit, multiplexer and A/D converter.

OPERATIONAL AMPLIFIER (OPAMP) 14.2

An OpAmp is a dc amplifier having a high gain, of the order of $10^4 – 10^8$. Such an amplifier can perform summing, integration, differentiation, or act as comparator with suitable feedback networks. Hence these are known as opamps. (In such circuits the required response is obtained by the application of negative feedback to a high gain dc amplifier by means of a component connected between input and output terminals, referred to as operational feedback).

An Opamp is a dc high gain amplifier usually consisting of one or more cascaded differential amplifier stages.

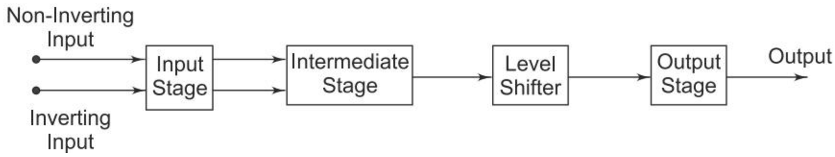


Fig. 14.5 Block diagram of a typical OpAmp

The block diagram of an Opamp shown in Fig. 14.5 consists of a four stage direct coupled amplifier in cascade. In summarized form, it can be stated as follows.

1. The first stage is double ended high gain (60) differential amplifier, i.e. dual input balanced output differential amplifier with a constant current source to increase the common mode rejection. Also this stage geneally determines the input resistance of the Opamp and provides most of the voltage gain.
2. The intermediate stage is a single ended differential amplifier, i.e. dual input unbalanced output differential amplifier in order to increase the gain (gain of intermediate stage = 30).
3. Since direct coupling is used, the output of the intermediate stage, which is the dc output voltage, is a higher than ground potential when no input is applied. Therefore to bring this dc level to ground potential, a level ttranslator (shifter) circuit is used after the intermediate stage. The third stage used in an emitter follower circuit.
4. The final stage is a push-pull complementary amplifier. With low output impedance and minimum offset voltage and currents, the power gain provided is between 5 and 10.

Hence the overall gain of the Opamp is ($60 \times 30 \times 5 = 9000$) which is very large.

The ideal opamp exhibits the following electrical characteristics:

1. Infinite voltage gain A_v.
2. Infinite input resistance R_i, so that almost any signal source can drive it and there is no loading of the preceding stage.
3. Zero output resistance R_o, so that output can drive an infinite number of devices.
4. Zero output voltage when input voltage is zero.
5. Infinite Bandwidth, so that any frequency signal from zero to infinity Hz can be amplified without attenuation.
6. Infinite CMRR, so that common-mode noise voltage is zero.
7. Infinite slew rate, so that output voltage changes occur simultaneously with input voltage changes.

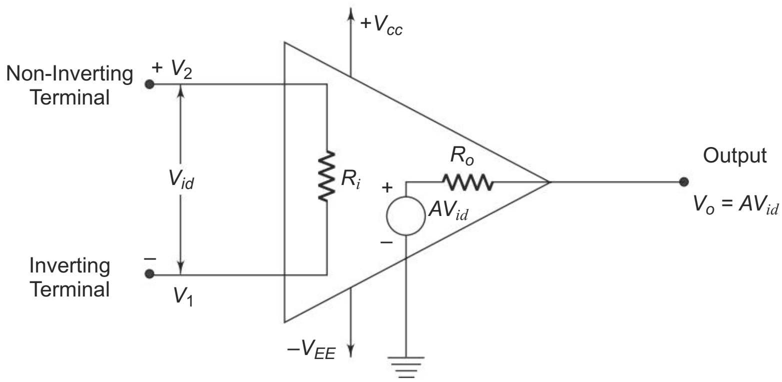


Fig. 14.6 Equivalent circuit of OpAmp

By the use of negative feedback a practical opamp can be brought close to some of the above characteristics, such as input resistance, output resistance and bandwidth. The equivalent circuit of an opamp is shown in Fig. 14.6.

The basic operating principles of an opamp can be analysed with the help of the equivalent circuit shown in Fig. 14.6. The output voltage for Fig. 14.6 is given by

$$V_o = A\,V_{id} = A\,(V_1 - V_2) \tag{14.1}$$

where A = large signal voltage gain.

V_{id} = differential input voltage

V_1 = voltage at Inverting terminal with respect to ground

V_2 = voltage at non-inverting terminal with respect to ground.

From Eq. (14.1), it is seen that the output voltage is directly proportional to the algebraic difference between the inputs. The opamp amplifies this difference voltage, not the input voltages themselves. For this reason the polarity of the output voltage depends on the polarity of the difference voltage.

14.2.1 Non-Inverting Amplifier (NI)

The circuit shown in Fig. 14.7 is commonly known as a Non-inverting amplifier with feedback (or closed loop Non-inverting amplifier), because it used a feedback and the input signal is applied to the non-inverting input terminal of the opamp. The closed loop gain of the non-inverting amplifier can be found as follows:

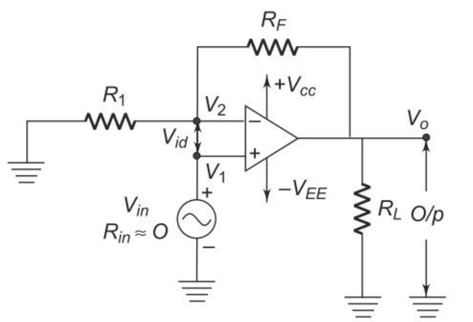


Fig. 14.7 Non-inverting amplifier with feedback

The closed loop gain $A_F = V_o / V_{in}$

From the circuit (Fig. 14.7), the output voltage is given by

$$V_o = A\,(V_1 - V_2) \tag{14.2}$$

But $V_1 = V_{\text{in}}$ and $V_2 = V_F$, V_F can be found by using the voltage divider law.

$$V_F = V_2 = \frac{R_1 V_o}{R_1 + R_F}$$

Substituting values of V_1 and V_2 Eq. (14.2) we have,

$$V_o = A\left(V_{in} - \frac{R_1 V_o}{R_1 + R_F}\right) = \frac{A(R_1 + R_F)V_{in}}{R_1 + R_F + A R_1}$$

Generally A is very large typically 10^5 hence $A\,R_1 > (R_1 + R_F)$

$$\therefore \qquad R_1 + R_F + AR_1 \approx AR_1.$$

Therefore, $$A_F = \frac{V_0}{V_{in}} = \frac{A(R, + R_F)}{AR_1} = \left(1 + \frac{R_F}{R_1}\right)$$

$\therefore$ the closed loop voltage gain for an non-inverting amplifier is

$$A_F = \left(1 + \frac{R_F}{R_1}\right) \tag{14.3}$$

14.2.2 Inverting Amplifier

The inverting amplifier configuration is shown in Fig. 14.8. The input signal is applied to the inverting terminal. The amplified inverted output is fed back to the inverting input through the resistor R_F. The signal fed back is proportional to the output voltage, feedback takes place through R_F

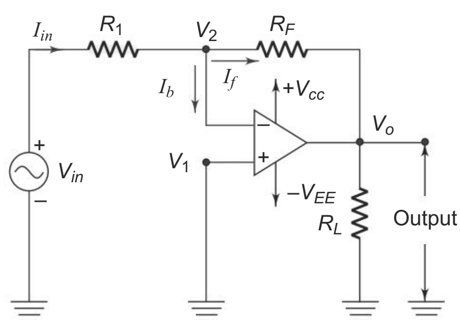


Fig. 14.8 Inverting amplifier with feedback

connected between the output and the inverting terminal of the amplifier. Phase inversion through the amplifier ensure that the feedback is negative.

In Fig. 14.8, the non-inverting terminal is grounded and the input signal is applied to the inverting terminal through a resistance R_1. The difference input voltage is ideally zero, since the voltage at the inverting terminal (V_2) is approximately equal to that at the non-inverting terminal (V_1). In other words, the inverting terminal voltage V_2 is approximately at ground potential. Therefore, the inverting terminal is said to be at virtual ground. The closed loop gain of the inverting amplifier can be computed as follows.

Referring to the circuit, since I_b is negligible $I_{in} \cong I_f$

Therefore $$\frac{V_{in} - V_2}{R_1} = \frac{V_2 - V_o}{R_F}.$$

But $$V_1 = V_2 = 0.$$

Hence $$\frac{V_{in}}{R_1} = \frac{-V_o}{R_F}$$

$$\therefore \quad \frac{V_0}{V_{in}} = -\frac{R_F}{R_1} = A_F$$

$\therefore$ closed loop voltage gain for an inverting amplifier is,

$$A_F = -\frac{R_F}{R_1} \tag{14.4}$$

14.2.3 Integrator Using OpAmp

An integrator is a circuit that performs the mathematical operation of integration because it produces an output voltage that is proportional to the integral of the input.

The integrator or integration amplifier is shown in Fig. 14.9.

A common application is to use a constant input voltage to produce a ramp voltage. (A ramp voltage is a linearly increasing or decreasing voltage.) The typical input to the integrator is a rectangular pulse. V_{in} is the constant input voltage of time T applied to the inverting terminal. Because negligible current

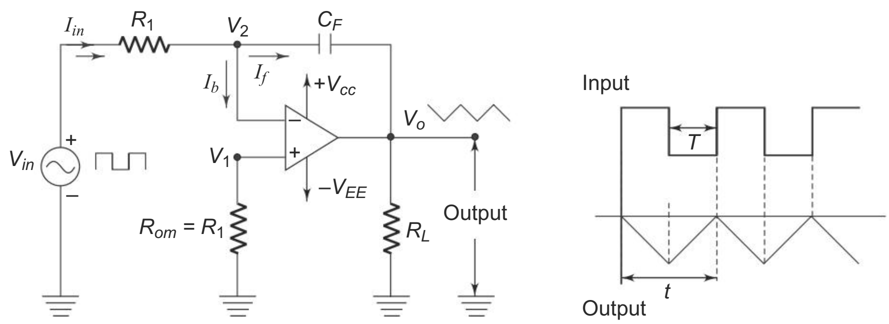


Fig. 14.9 Integrator using OpAmps

flows into the opamp, the input current is constant and equals $I_{in} = V_{in}/R_1$. Approximately all the current goes to the capacitor C_F. As $C = Q/V$ or $V = Q/C$ and since a constant current is flowing into the capacitor, the charge Q increases linearly. This means that the capacitor voltage increases linearly. Because of the phase reversal of the opamp, the output voltage is a negative ramp. At the negative pulse period of the input signal, the direction of the charging current is opposite, hence the output is an increasing ramp.

An integrator circuit is obtained by using a basic inverting amplifier configuration and by replacing the feedback resistor R_F by a capacitor C_F.The expression for the output voltage V_o can be obtained by using Kirchhoff's current equation.

$I_{in} = I_f + I_b$, since I_b is negligible, $I_{in} \cong I_f$. The current through the capacitor is given by

$$I_f = C_F \frac{dV_c}{dt}$$

But
$$I_{in} = \frac{(V_{in} - V_2)}{R_1}$$

As
$$I_{in} \cong I_f$$

$$\therefore \quad \frac{V_{in} - V_2}{R_1} = C_F \frac{dV_c}{dt} = C_F \frac{d(V_2 - V_o)}{dt}$$

But
$$V_1 \cong V_2 \cong 0$$

therefore
$$\frac{V_{in}}{R_1} = -C_F \frac{dV_0}{dt}$$

The output voltage can be obtained by integrating both sides with respect to time, hence

$$\int_0^t \frac{V_{in}}{R_1} = -C_F \int_0^t \frac{dV_o}{dt}$$

$$\therefore \quad V_o = -\frac{1}{C_F R_1} \int_0^t V_{in}\, dt \tag{14.5}$$

This equation says that the output voltage is directly proportional to the negative integral of the input voltage and inversely proportional to the time constant $C_F R_1$.

The input impedance of the integrator circuit is equal to the resistance R_1, the output impedance is low because of the negative feedback which is inherent in the circuit. CR is called the time constant of the integrator. For perfect integration, the time period T of the input signal must be longer than or equal to the time constant $C_F R_1$. In order to minimise the offset voltages, a resistor R_{om} called the "offset minimise resistor" is connected to the non-inverting terminal of the op-amp.

The integral of a sine function is a cosine function, hence, if a sine wave input is applied to an integrator, the output is a cosine wave. Also, if a square wave input is applied, the output is a triangular wave.

A practical integrator is shown in Fig. 14.10.

At low frequency and dc signals, the capacitor acts like an open circuit. Hence the closed loop voltage gain equals the open loop voltage gain. This produce too much output offset voltage. At zero frequency (DC) and without negative feedback, the circuit treats the input offsets as an valid input signal charges the capacitor. This drives the output into positive or negative saturation.

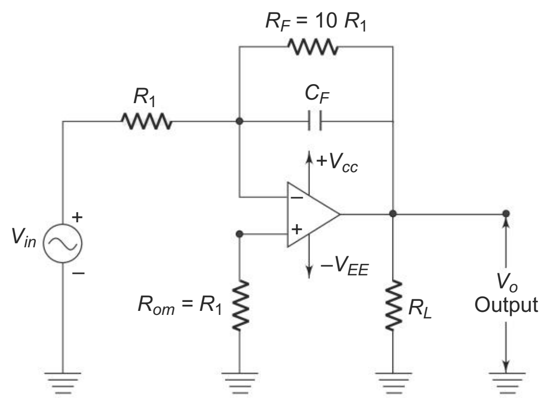


Fig. 14.10 Practical integrator using OpAmp

Hence a way to minimise the effect of these input offsets is to roll off the voltage gain at low frequencies by inserting a resistor R_F in parallel with the capacitor, as shown in Fig. 14.10. The resistor used should be at least 10 times greater than the input resistance.

The integrator is most commonly used in analog computer, A/D conversion and signal wave shaping.

14.2.4 Differentiator Using OpAmp

A differentiator is a circuit that performs the mathematical operation of differentiation. It produces an output voltage proportional to slope of the input voltage. A common application of a differentiator is the detection of the leading and trailing edges of rectangular pulse. Figure 14.11 shows a basic CR differentiator. If a rectangular pulse is applied, the output of the circuit is pos itive and negative spikes. The positive spike occurs at the same instant as the leading edge of the input and the negative spike at the same instant as the trailing edge.

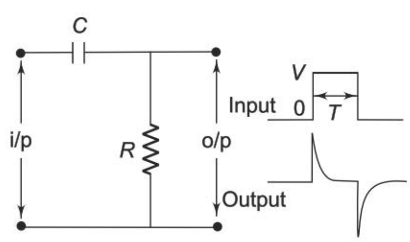


Fig. 14.11 Basic circuit of a differentiator

When the input voltage changes from 0 to V, the capacitor begins to charge exponentially, and the output voltage suddenly jumps from 0 to V (after 5 time constants, the capacitor voltage is within 1% of its final voltage V) and then decays exponentially.

Over the trailing edge of a pulse, the input voltage steps negatively and we get a negative spike.

For a CR Differentiator to produce narrow spikes the time constant must be 10 times smaller than period T.

In Fig. 14.12 (b), when the input voltage varies, the capacitor charges or discharges. Because of the virtual ground, the capacitor current passes through the feedback resistor, producing a voltage. This voltage is proportional to the slope of the input voltage.

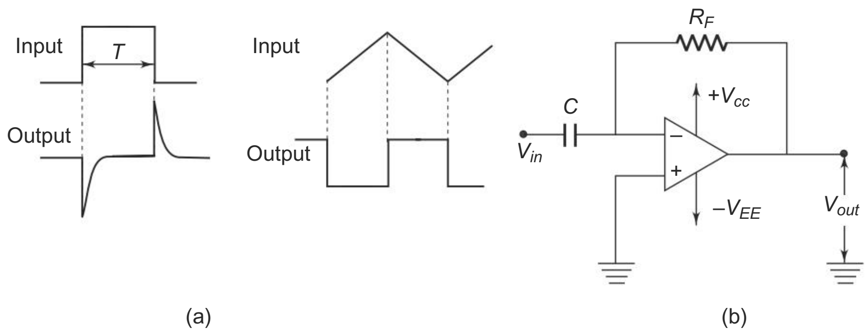


Fig. 14.12 (a) Waveforms of differentiator (b) Basic differentiator using OpAmp

A rectangular input, shown in Fig. 14.12 (a), is used to obtain positive and negative spikes. The leading edge of the pulse is approximately a positive ramp, so that the output is a negative spike of very short duration. Similarly, the trailing edge of the input pulse is approximately negative ramp, so that the output is a very narrow positive spike. Similarly, if a triangular waveform is given as input, a square waveform is obtained at the output.

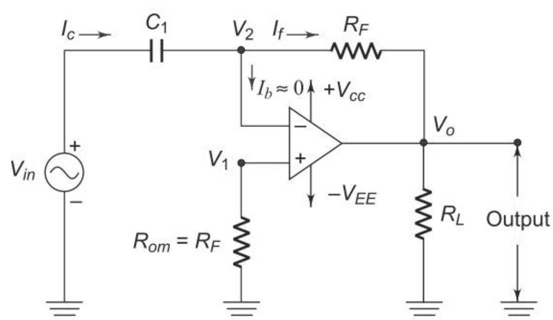


Fig. 14.13 Differentiation amplifier

A basic differentiator is constructed from a basic inverting amplifier by replacing the input resistor R_1 by a capacitor C_1 as shown in Fig. 14.13. Using

Kirchoff's current equation at node V_2 is $I_c = I_b + I_f$. Since $I_b \cong 0$,

$$\therefore \quad I_c \cong I_f$$

But
$$I_c = C_1 \frac{dV_c}{dt} = C_1 \frac{d}{dt}(V_{in} - V_2)$$

and
$$I_f = \frac{V_2 - V_o}{R_F}$$

As
$$I_c \cong I_f$$

$$\therefore \quad C_1 \frac{d}{dt}(V_{in} - V_2) = \frac{V_2 - V_o}{R_F}$$

But $V_1 = V_2 = 0$, because A is very large therefore

$$\therefore \quad C_1 \frac{dV_{in}}{dt} = -\frac{V_o}{R_F} \quad \text{or} \quad V_o = -C_1 R_F \times \frac{dV_{in}}{dt} \tag{14.6}$$

From the above equation it can be seen that a differentiator performs the opposite function of integration. If a cosine wave is applied as the input, a sine wave will be produced at the output, and a triangular wave as the input will produce a square wave at the output.

The opamp differentiator of Fig. 14.13 has a tendency to oscillate which is undesirable.

The gain R_F/X_{C1} of the circuit increases at a rate of 20 db/decade with increases in frequency, which makes the circuit unstable. As the frequency increases, the input impedance (X_{C1}) decreases, making the circuit more susceptible to high frequency noise. This noise is superimposed after amplification, on the differentiated output signal. This high frequency noise and stability can be corrected by the use of R_1 and C_F, as shown in Fig. 14.14, which is a practical differentiator.

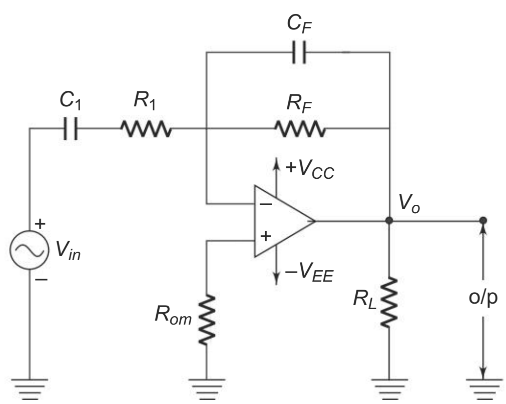


Fig. 14.14 Practical differentiator

A typical value for R_1 is between $0.01R_F$ and $0.1R_F$. With this the closed loop voltage gain is reduced (to between −10 and −100).

The differentiating circuit is most commonly used in wave shaping circuits to detect high frequency components in an input signal and as a rate of change detector in FM modulators. It is also used in triggering the time base generator in an oscilloscope.

To design a practical differentiator the following parameters are required.

(a) f_a is the highest frequency range to be differentiated, and is given by $f_a = 1/(2\pi R_F C_1)$.

(b) f_b is the gain limiting frequency at which the gain begins to decrease at a rate of 20 db/decade. This frequency is given by $f_b = 1/2\pi R_1 C_1$, where $R_1 C_1 = R_F C_F$ and $f_b = 20 f_a$.

Example 14.1

(i) Design a differentiator to differentiate an input signal that varies in frequency from 20 Hz to 800 Hz.

(ii) A sine wave of 2 V peak at 800 Hz is applied to the differentiator. Draw its output waveform.

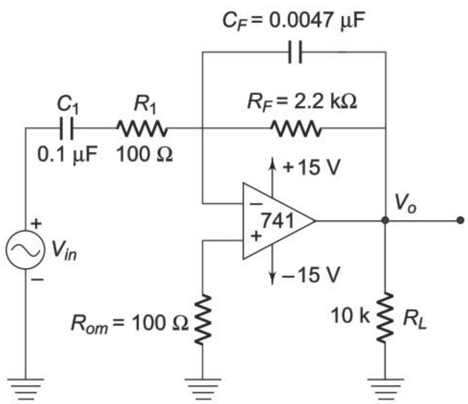


Fig. 14.15

Solution

(i) To design a differentiator, we follow the steps outlined above.

(a) f_a = 800 Hz (the highest frequency) and $f_a = 1/(2\pi C_1 R_F)$

Let $C_1 = 0.1\ \mu\text{F}$,

then $$R_F = \frac{1}{2\pi C_1 f_a} = \frac{1}{2 \times 3.14 \times 0.1\,\mu\text{F} \times 800\ \text{Hz}} = 1.98\ \text{k}\Omega$$

Let $R_F = 2.2\ \text{k}\Omega$

(b) $f_b = 20 \cdot f_a = 20 \times 800 = 16$ kHz and $f_b = \dfrac{1}{2\pi R_1 C_1}$

$$\therefore \quad R_1 = \frac{1}{2\pi C_1 f_b} = \frac{1}{1 \times 3.14 \times 0.1\,\mu\text{F} \times 16 \times 1000} = 99.5\ \Omega$$

Let $R_1 = 100\ \Omega$

Since $R_1 C_1 = R_F C_F$,

$$C_F = \frac{R_1\, C_1}{R_F}$$

$$= \frac{100 \times 0.1 \times 10^{-6}}{2200}$$

$$= 0.0045\ \mu\text{F}$$

Let $C_F = 0.0047\ \mu\text{F}$

$R_{om} = R_1 // R_F = 2.2\text{ k}//100 = 95.6\Omega$

(but 100 ohms can be used)

The complete circuit is shown in Fig 14.15.

(ii) Since $V_p = 2V$

and $f = 800$ Hz, the input voltage is $V_{in} = V_p \sin \omega t$

$V_{in} = 2\ V \sin (2\pi\, 800)t$

As $$V_o = -R_F C_1 \frac{dV_{in}}{dt} = -R_F C_1 \frac{d}{dt}\,(V_p \sin \omega t)$$

$\therefore$ $V_o = -R_F C_1 \times Vp(\omega) \cos \omega t$

$\therefore$ $V_o = -2.2\text{ k} \times 0.1\ \mu\text{F} \times 2 \times 2 \times 3.142 \times 800 \times \cos \omega t$

$\therefore$ $V_o = -2.21 \cos \omega t$

$V_o = -(2.21) \cos [2\pi\, 800\, t]$

The waveform is as shown in Fig. 14.16.

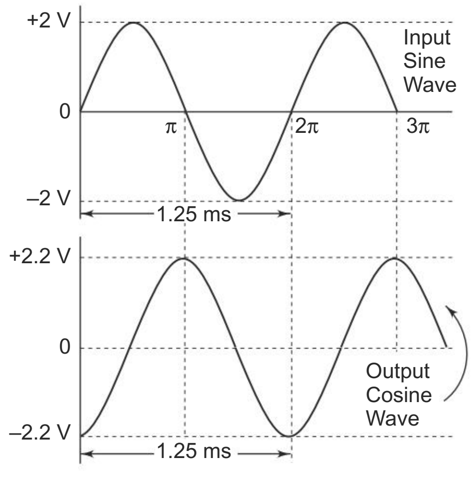


Fig. 14.16

14.2.5 Summing Amplifier

Figure 14.17 shows an summing amplifier in inverting configuration with three inputs V_a, V_b, V_c. Depending on the relation between R_a, R_b, R_c and R_F, the circuit can be used as a Summing amplifier, Scaling amplifer or Average amplifier.

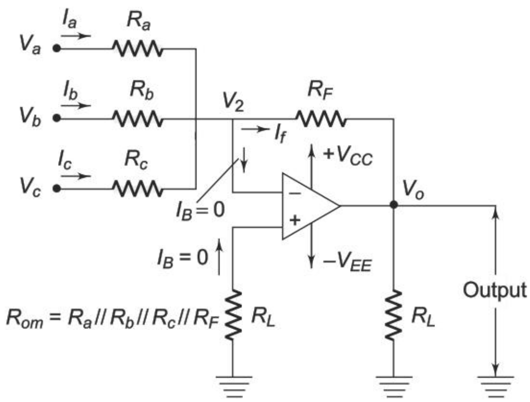


Fig. 14.17 Summing amplifier (3 input inverting configuration)

Using Kirchoff's circuit equation, we have $I_a + I_b + I_c = I_B + I_f$. But $I_B \cong 0$ and $V_1 \cong V_2 \cong 0$

Therefore $$I_a + I_b + I_c = I_f$$

i.e. $$\frac{V_a}{R_a} + \frac{V_b}{R_b} + \frac{V_c}{R_c} = -\frac{V_o}{R_F}$$

$$R_a = R_b = R_c$$

$\therefore$ $$V_o = -\frac{R_F}{R_a}(V_a + V_b + V_c)$$

In this circuit $R_a = R_b = R_c = R_F$. Therefore $V_o = -(V_a + V_b + V_c)$. (14.7)

Hence the output voltage is the negative sum of all the input voltages.

If each input voltages is amplified by a different factor, i.e. weighted differently at the output, the circuit is called a scaling or weighted amplifier (Fig. 14.17). The condition can be obtained by making R_a, R_b, and R_c, different in value. The output voltage of the scaling amplifier is then

$$V_o = -\left(\frac{R_F}{R_a}V_a + \frac{R_F}{R_b}V_b + \frac{R_F}{R_c}V_c\right) \tag{14.8}$$

where $R_F/R_a \neq R_F/R_b \neq R_F/R_c$.

Figure 14.17 can be modified to be used as an average amplifier. In this amplifier, the output voltage is the average value of the input voltages. This mofification can be obtained by making $R_a = R_b = R_c = R$. Also, the gain by which input is amplified must be equal to 1 over the number of inputs, i.e. $R_F/R = 1/n$ where n is the number of inputs. Therefore the output voltage is given by $V_o = V_a + V_b + V_c$.

Therefore the output voltage for three inputs is $R_F/R = 1/3$. The output voltage is given by

$$V_o = -\frac{V_a + V_b + V_c}{3} \tag{14. 9}$$

Example 14.2 *In the circuit of Fig. 14.18,*

$V_a = +2\text{ V}, V_b = +1\text{ V},$
$V_c = +3\text{ V}$ and
$R_a = R_b = R_c = 3\text{ k}\Omega,$
$R_F = 1\text{ k}, R_{om} = 270\ \Omega,$
Supply voltage $= \pm 15\text{ V}.$

Assuming that the opamp is initially nulled, determine the output voltage.

Solution

$$V_o = -\frac{1\text{ k}}{3\text{ k}}(2+1+3) = -\frac{1\text{ k}}{3\text{ k}} \times 6 = -2\text{ V}$$

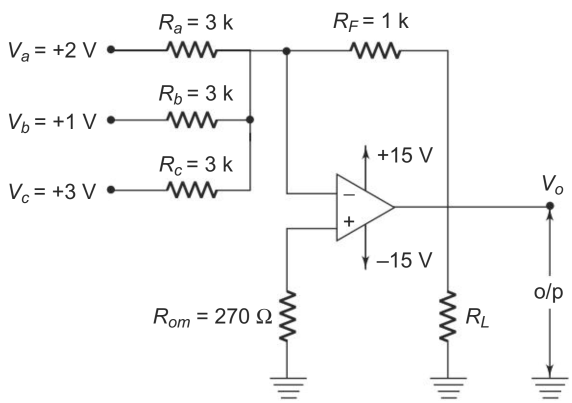


Fig. 14.18

14.2.6 Subtractor

A subtractor circuit using a basic differential amplifier is as shown in Fig. 14.19.

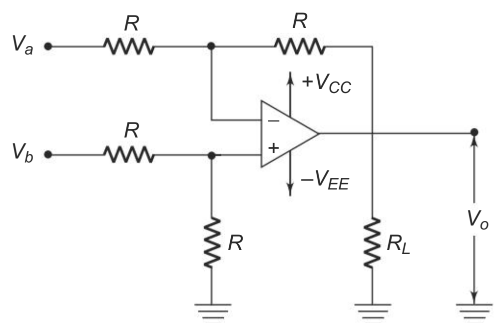


Fig. 14.19 Subtractor using OpAmp

By selecting the appropriate values for the external resistance, the input signal can be scaled (attenuated) to the desired value. If this is done, the circuit is referred to as a scaling amplifier.

As in Fig. 14.19, all values of the external resistance are equal, and the gain of the amplifier is unity.

Therefore, the output voltage of differential amplifier with unity gain is

$$V_o = -\frac{R}{R}[V_a - V_b]$$

$$V_o = -[V_a - V_b] \tag{14.10}$$

Hence the circuit is called a subtractor.

14.2.7 Comparator

In the opamp applications discussed so far, the amplifiers use negative feedback. Under normal conditions, when negative feedback is used in such a circuit, the amplifier output voltage takes on values between the positive and negative saturation limits. The amplifier has a high gain and negative feedback forces the voltage between the differential input to be small at all times.

When opamp is used without feedback (open loop operation), the amplifier output is usually in one of its saturated states. The application of a small difference input signal of appropriate polarity causes the output to switch to its other saturation states.

Therefore, a comparator is a circuit with two inputs and a single output. The two inputs can be compared with each other, i.e., one of them can be considered a reference terminal.

When the non-inverting input is higher or greater than the inverting input voltage, the output of the comparator is high and when the non-inverting voltage is less than the latter output of the comparator is low.

If the inverting terminal is grounded, even the slightest input voltage at the non-inverting terminal is enough to saturate the opamp. Therefore, the output is at positive saturation as soon as the voltage at the non-inverting terminal increases slightly above zero.

Similarly, the opamp goes to negative saturation, as soon as the voltage at the non-inverting terminal goes slightly below ground level.

The comparator circuit shown in Fig. 14.20, consists of a fixed reference voltage (V_{ref}) applied to the inverting input terminal and a sinusoidal signal V_{in} applied to the non-inverting terminal. As discussed earlier, when V_{in} is greater than V_{ref}, the output voltage goes to positive saturation, i.e. $V_{out} = +V_{sat} = +V_{cc}$ and when V_{in} is less than V_{ref}, the output goes to negative saturation, i.e. $V_{out} = -V_{sat} = -V_{EE}$.

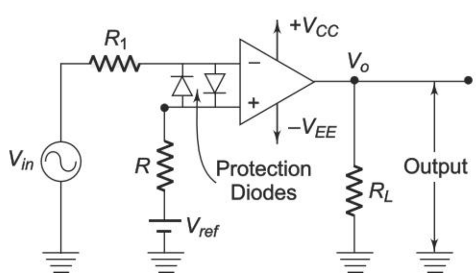


Fig. 14.20 OpAmp comparator

Hence the output changes from one saturation level to another whenever $V_{in} \cong V_{ref}$, as shown in Fig. 14.21.

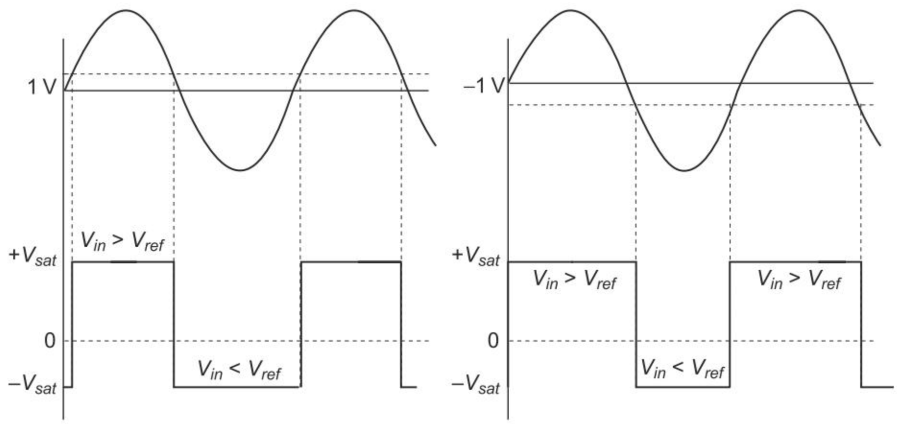


Fig. 14.21 (a) If V_{ref} is positive (b) If V_{ref} is negative

Since the sinusoidal input is applying to the non-inverting terminal, this circuit is called the non-inverting comparator.

Similarly, an inverting comparator can also be obtained by applying the sinusoidal input to the inverting terminal.

Diodes D_1 and D_2, shown in Fig. 14.20 are used to protect the opamp from damage due to excessive input voltage (V_{in}). The difference input voltage (V_{id}) of the opamp is clamped to either + 0.7 V or – 0.7 V because of the diodes D_1 and D_2. Hence the diodes are called clamping diodes.

Figure 14.22 (a) shows the circuit for an inverting comparator, the input-output waveforms are shown in Fig. 14.22. (b). In this circuit, V_{ref} is obtained by the use of a potentiometer which forms a potential divider arrangement. Comparators are used in circuits such as discriminators, voltage level detectors, oscillators, digital interfacing, Schmitt trigger, etc.

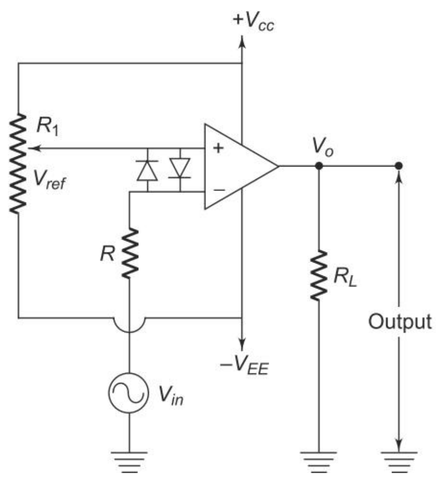


Fig. 14.22 (a) Inverting comparator

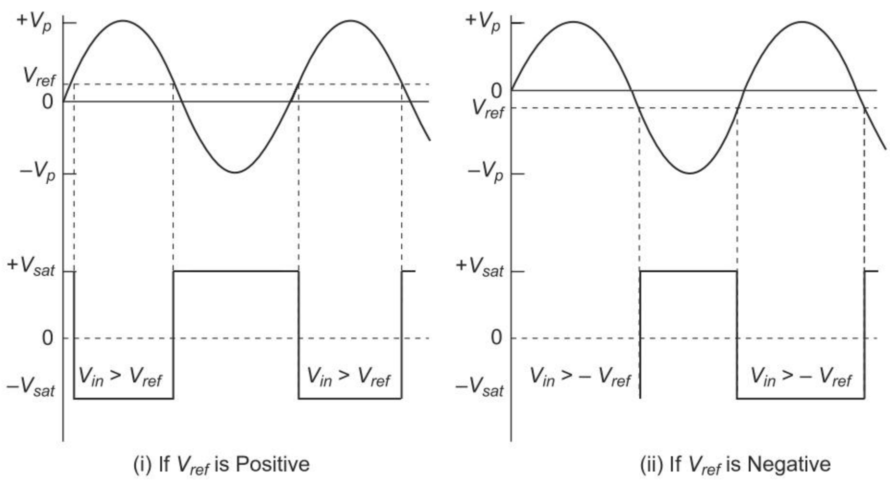


Fig. 14.22 (b) Comparator output waveform

BASIC INSTRUMENTATION AMPLIFIER 14.3

The low level signal output of electrical transducers often need to be amplified before further processing. This is done by the use of instrumentation amplifiers. The important features of Instrumentation amplifiers are as follows:

1. Selectable gain with high gain accuracy and gain linearity.
2. Differential input capability with high gain common mode rejection.
3. High stability of gain with low temperature co-efficient.
4. Low dc offset and drift errors referred to input.
5. Low output impedance.

Instrumentation amplifiers differ from the ordinary opamp in the following manner:

1. The instrumentation amplifier is often a package consisting of opamps wired up with accurate and stable resistive feedback to give a desired gain. The gain can be selected to precise value by a single external resistance.
2. The instrumentation amplifier can be used directly to amplify signals by a fixed amplification factor, because of its closed loop configuration. The gain accuracy, gain stability and drift performance are normally specified by the manufacturer, hence there is no effort required in choosing the input and feedback configuration.
3. Depending on the configuration employed, the instrumentation amplifier will often yield a high CMRR even when (a) the source impedance exceeds 1 MΩ, (b) When the impedance is unbalanced for a large value of variation 1 – 10 k. The dc operational amplifier has often a CMRR specified for it, which is only applicable for very low source impedances.

In actual applications an instrumentation amplifier is a specific combination of a suitable dc opamp wired up with feedback.

14.3.1 Instrumentation Amplifier

The schematic diagram of an instrumentation amplifier constructed with dc opamps and consisting of general features is shown in Fig. 14.23 (a).

Many of the input specifications of the opamps employed directly determine the input specification of the instrumentation amplifier.

An analysis of the circuit of Fig. 14.23 (a) gives the following equations.

Let $R_4 = R_5 = R_6 = R_7$.

Therefore,

$$e_3 = \left(1 + \frac{R_2}{R_1}\right)e_1 - \left(\frac{R_2}{R_1}\right)(e_{cm} + e_2) \tag{14.11}$$

$$e_4 = \left(1 + \frac{R_3}{R_1}\right)e_2 - \left(\frac{R_3}{R_1}\right)(e_{cm} + e_1) \tag{14.12}$$

$$e_5 = e_4 - e_3 \tag{14.13}$$

where $e_{cm} + e_1$ is the input to amplifier A_1.

and $e_{cm} + e_2$ is the input to amplifier A_2

If $R_2 = R_3$, the output voltage is given by

$$e_5 = \left(1 + \frac{2R_2}{R_1}\right) \times (e_2 - e_1) \tag{14.14}$$

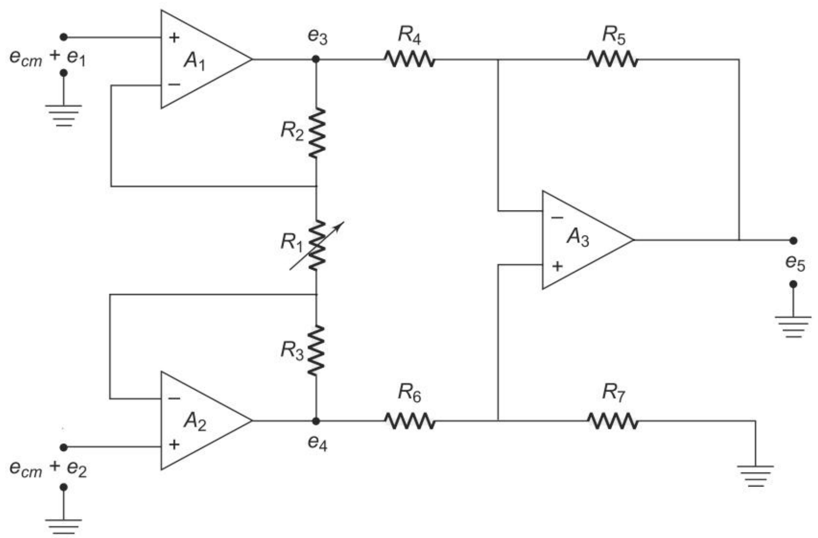


Fig. 14.23 (a) Instrumentation amplifier

The input amplifiers A_1 and A_2 act as input buffers with unity gain for common mode signals e_{cm} and with a gain of $(1 + 2R_2/R_1)$ for differential signals.

A high input impedance is ensured by the non-inverting configuration in which they operate. The common mode (CM) rejection is achieved by the following stage which is connected as a differential amplifier. The optimum common mode rejection can be obtained by adjusting R_6 or R_7 ensuring that $R_5/R_4 = R_7/R_6$.

The amplifier A_3 can also be made to have some nominal gain for the whole amplifier by an appropriate selection or R_4, R_5, R_6, and R_7.

The drift errors of the second stage add to the product of the drift errors of the first amplifier and first stage gain. Hence, it is necessary that the gain in the first stage be enough to prevent the overall drift performance from being significantly affected by the drift in the second stage. The drift problem of instrumentation amplifiers can be improved if amplifiers A_1 and A_2 have offset voltages which tends to track the temperature.

The gain of an instrumentation amplifier can be varied by changing R_1 alone. A high gain accuracy can be obtained by using precision metal film resistors for all the resistances.

Because of the large negative feedback used, the amplifier has good linearity, typically about 0.01% for a gain less than 10. The output impedance is also low, being in the range of milli ohms.

The input bias current of the instrumentation amplifier is determined by that of the amplifiers A_1 and A_2.

Figure 14.23 (b) shows a commercial unit of an instrumentation amplifier together with a circuit to program it with a microprocessor system.

Typical specifications of a commercial amplifier shown in Fig. 14.23 (b) are listed in Table 14.1.

14.3.2 Instrumentation System

The measurement and control of physical conditions is very important in many industrial and consumer applications. For example, the operator may make necessary adjustments in the measurement of temperature or humidity inside a dairy or meat plant to maintain the product quality, or to produce a particular type of plastic, precise temperature control of the plastic furnace is needed.

A transducer is generally used at the measuring site to obtain the required information easily and safely. As explained in Chapter 13, a transducer is a device that converts one form of energy into another.

For example, when a strain gauge is subjected to pressure or force (physical energy), the resistance of the strain gauge changes (electrical energy), i.e. it converts mechanical energy into electrical energy. Actually, an instrumentation system is used to measure the output signal produced by the transducer and mostly used to control the physical condition producing the output signal.

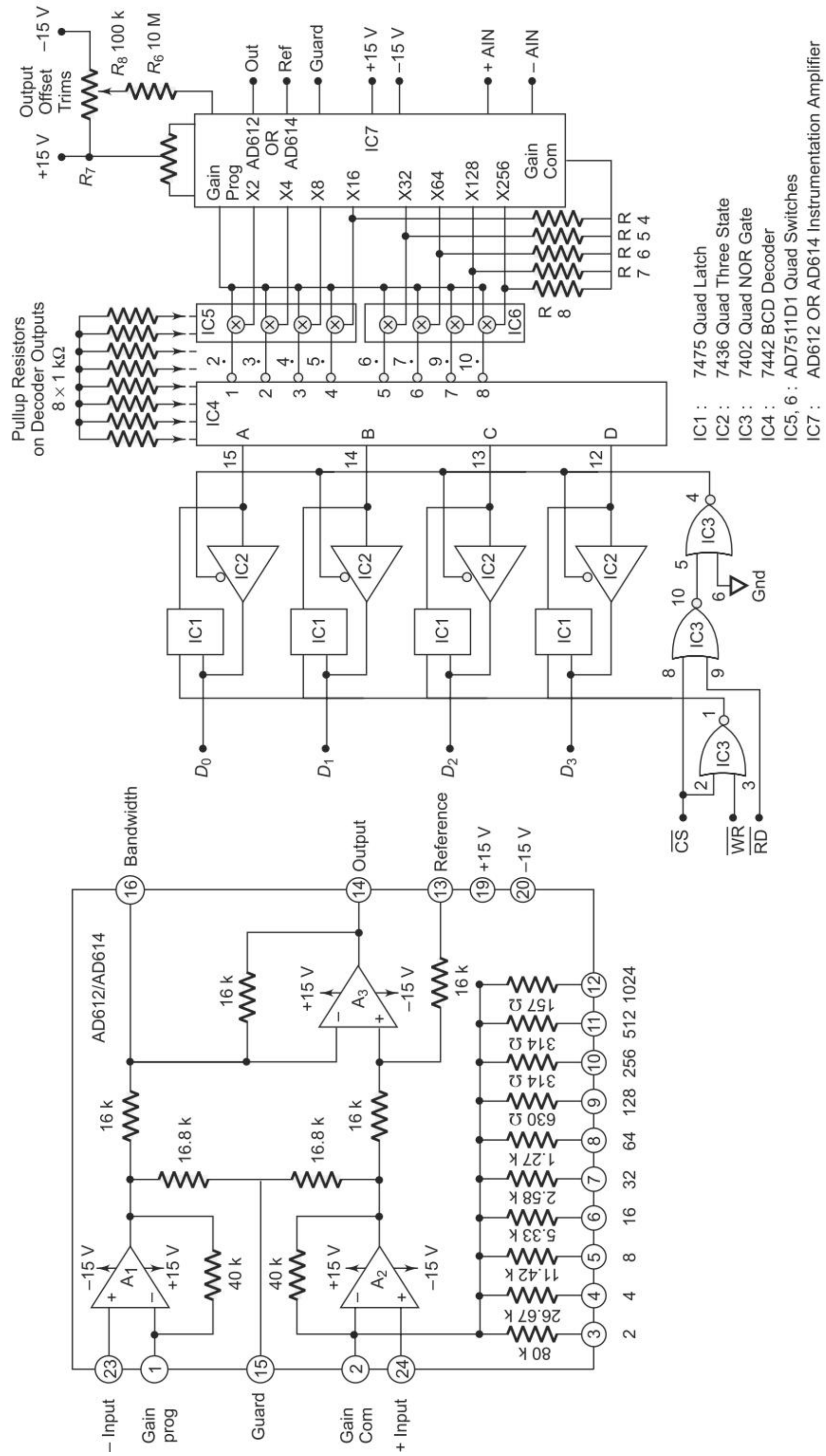


Fig. 14.23 (b) Commercial unit of instrumentation amplifier using a microprocessor system

Table 14.1 Specifications for Type 3620 Instrumentation Amplifier

Parameter	*Specification*
Gain, A	1 + 25/R (to within 0.2%); R is the gain determining resistance
Gain range	1 to 1000
Gain non-linearity	± 0.01% at A = 100
Gain stability	± 0.001% per ºC; ± 0.001% per month
Rated output	± 10 V, ± 10 mA
Input impedance	300 MΩ differential; 1000 MΩ common mode
CMRR, dc to 100 Hz;	
Gain of 10;	74 dB
(1 kΩ source unbalance)	
Gain of 1000; 1 kΩ balance	
source	100 dB
Input offset (can be zeroed)	± 1 mV
Output offset (A = 1000)	± 0.25 mV/ºC; ± 3 mV/month
Bias current	± 25 nA
Bias current variation with temp	± 0.5 nA/ºC
Input noise (A = 100)	
Voltage (0.01 Hz to 1 Hz)	1 μV_{P-P}
Current (0.01 Hz to 10 Hz)	200×10^{-12} A_{P-P}
Dynamic response (A = 100)	
Small signal ± 1%	1.5 kHz
± 3 dB	10 kHz
Full power (A = 10)	5 kHz
Settling time to within ± 10mV	
of output value	200 µs
Slew rate	0.3 V/µs
Power supply	± 12 V ± 18 V
Drain	± 24 mA
Temperature range	0 to 70ºC

The simplified form of such an instrumentation system is shown in Fig. 14.24. This instrumentation system consists of a type of transducer as the input stage, depending upon the physical quantity to be measured. The transducers output is fed to the pre-amplifier. The instrumentation amplifier is the intermediate stage. The output of the instrumentation amplifier can be connected to various devices, such as meter, oscilloscope, charts or magnetic recorders.

Advanced technology has led to use of automatic instrumentation systems. This system have an automatic process controller used at the output stage, which compensates for changes in the operating condition.

The lines connecting the various stages, as shown in Fig. 14.24, are called the transmission lines. On the system requirement and the physical quantity to be monitored, the length of these transmission lines are chosen. These transmission lines permit signal transfer from unit to unit.

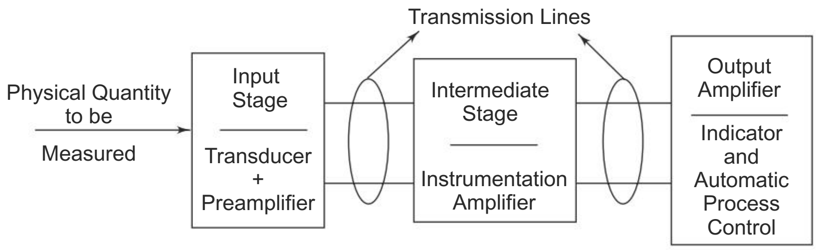


Fig. 14.24 Block diagram of an instrumentation system

The output of the transducer is the input signal source of the instrumentation amplifier. A transducer which produces sufficient strength can be used to drive the output device directly. Most do not produce sufficient output. Hence, to amplify these low level output signals of the transducer instrumentation amplifiers are used which drive the indicator or display unit.

The instrumentation amplifier is required for precise low level signal amplification. In brief, they are used, where low noise, low thermal and time drift, high input resistance and accurate closed loop gain are required.

14.3.3 Instrumentation Amplifier Using Transducer Bridge

Figure 14.25 shows a simplified circuit of a differential instrumentation amplifier using a transducer bridge.

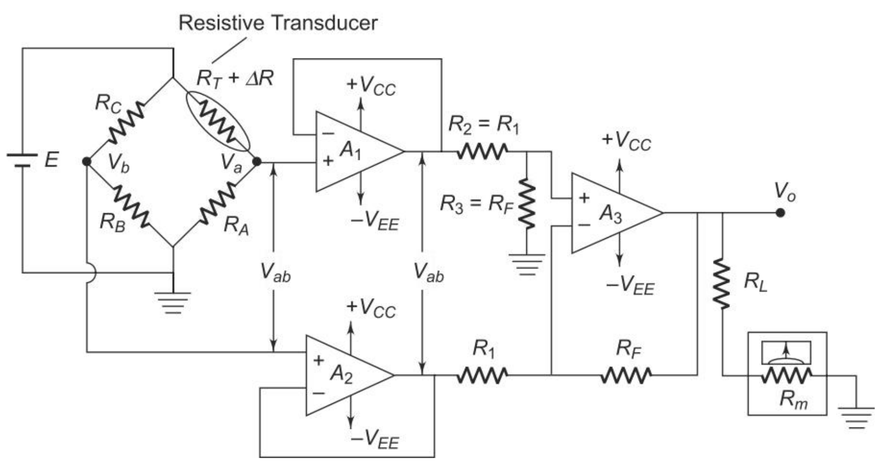


Fig. 14.25 Differential instrumentation amplifier using transducer bridge

In this circuit a resistive transducer (whose resistance changes as a function of some physical energy) is connected to one arm of the bridge.

Let R_T be the resistance of the transducer and ΔR the change in resistance of the resistive transducer. Hence the total resistance of the transducer is $(R_T \pm \Delta R)$.

The condition for bridge balance is $V_b = V_a$, i.e. the bridge is balanced when $V_b = V_a$, or when

$$\frac{R_B(E)}{R_B + R_C} = \frac{R_A(E)}{R_A + R_T}$$

Therefore,
$$\frac{R_c}{R_B} = \frac{R_T}{R_A}$$

The bridge is balanced at a desired reference condition, which depends on the specific value of the physical quantity to be measured. Under this condition, resistors R_A, R_B and R_C are so selected that they are equal in value to the transducer resistance R_T. (The value of the physical quantity normally depends on the transducers characteristics, the type of physical quantity to be measured, and the desired applications.)

Initially the bridge is balanced at a desired reference condition. As the physical quantity to be measured changes, the resistance of the transducer also changes, causing the bridge to be unbalanced ($V_b \neq V_a$). Hence, the output voltage of the bridge is a function of the change in the resistance of the transducer. The expression for the output voltage V_o, in terms of the change in resistance of the transducer is calculated as follows.

Let the change in the resistance of the transducer be ΔR. Since R_B and R_C are fixed resistors, the voltage V_b is constant, however, the voltage V_a changes as a function of the change in the transducers resistance.

Therefore, applying the voltage divider rule we have

$$V_a = \frac{R_A(E)}{R_A + (R_T + \Delta R)} \text{ and } V_b = \frac{R_B(E)}{R_B + R_C}$$

The output voltage across the bridge terminal is V_{ab} given by $V_{ab} = V_a - V_b$.

Therefore,
$$V_{ab} = \frac{R_A(E)}{R_A + (R_T + \Delta R)} - \frac{R_B(E)}{R_B + R_C}$$

However, if $R_A = R_B = R_C = R_T = R$, then

$$V_{ab} = \frac{R(E)}{2R + \Delta R} - \frac{R(E)}{2R} = E\left(\frac{R}{2R + \Delta R} - \frac{1}{2}\right)$$

$$V_{ab} = E\left(\frac{2R - 2R - \Delta R}{2(2R + \Delta R)}\right) = \frac{-\Delta R(E)}{2(2R + \Delta R)} \tag{14.15}$$

The output voltage V_{ab} of the bridge is applied to the differential amplifier through the voltage followers to eliminate the loading effect of the bridge circuit.

The gain of the basic amplifier is (R_F/R_1) and therefore the output voltage V_o of the circuit is given by

$$V_o = V_{ab}\left(\frac{R_F}{R_1}\right) = \frac{-\Delta R\,(E)}{2\,(2\,R + \Delta R)} \times \frac{R_F}{R_1} \tag{14.16}$$

It can be seen from the Eq. (14.16) that V_o is a function of the change in resistance ΔR of the transducer. Since the change is caused by the change in a physical quantity, a meter connected at the output can be calibrated in terms of the units of the physical quantity.

APPLICATIONS OF INSTRUMENTATION AMPLIFIERS (SPECIFIC BRIDGE) 14.4

We shall now consider some important applications of instrumentation amplifiers using resistance types transducers. In these transducers, the resistance of the transducer changes as a function of some physical quantity. Commonly used resistance transducers are thermisistors, photoconductor cells, and strain gauges.

14.4.1 Temperature Indicators Using Thermistor

The Thermistor is a relative passive type of temperature resistance transducer. They are basically semiconductors.

In many respects, a thermistor resembles a conventional resistor. It is usually a two-terminal device. It has resistance as its fundamental property. It is generally installed and operated in the manner of an ordinary resistor. But its great difference is that it has a negative temperature coefficient (NTC) or positive temperature coefficient (PTC) type. Most thermistors exhibit an NTC characteristic. An NTC type is one in which its resistance decreases with increase in temperature. The temperature coefficient is expressed in ohms/°C.

Since it is a thermally sensitive resistor, it has a high temperature coefficient of resistance and is therefore well suited for temperature measurement and control.

If in the bridge circuit of Fig. 14.25 the transducer used is a thermistor, the circuit can thus be used as a temperature indicator. The output meter is then calibrated in °C or °F. The bridge is balanced initially at a desired reference condition. As the temperature varies, the resistance of the thermistor also changes, unbalancing the bridge, which in turn produces a meter deflection at the output. By selecting the appropriate gain for the differential instrumentation amplifier, the meter can be calibrated to read a desired temperature. In this circuit, the meter movement (deflection) depends on the amount of unbalance in the bridge, which is caused by a change in the value of thermistor resistance ΔR. The change ΔR for the thermistor can be determined as follows.

ΔR = temperature coefficient of resistance
[final temperature – reference temperature]

If the meter in this circuit is replaced by a relay, and if the output of the differential instrumentation amplifier drives the relay that controls the current in the heat-generating circuit, a temperature controller can be formed. A properly designed circuit should energise a relay when the temperature of the thermistor drops below a desired value, causing the heater unit to turn on.

Example 14.3 *In the circuit of Fig. 14.25, R_1 = 2.2 k, R_F = 10 k, $R_A = R_B = R_C$ = 120 k, E = + 5 V, and Opamp supply voltage = ± 15 V. The transducer is a thermistor with the following specifications.*

R_T = 120 K at a reference temperature of 25°C

temperature coefficient of resistance – 1 k/°C or 1%/°C

Determine the output voltage at 0°C and 100°C.

Solution At 25°C, $R_A = R_B = R_C$ = 120 k

Therefore, the bridge is balanced and $V_a = V_b$. Therefore V_o = 0.

At 0°C the change ΔR in the resistance of the thermistor is

$$\Delta R = \frac{(-1\text{ k})}{°\text{C}} \times (0°\text{C} - 25°\text{C}) = 25\text{ k}$$

Therefore, the output voltage is given by

$$V_o = \frac{-(\Delta R)E}{2(2R + \Delta R)} \times \frac{R_F}{R_1}$$

$$V_o = \frac{-(25\text{ k}) \times (5)}{2(240\text{ k} + 25\text{ k})} \times \frac{10\text{ k}}{2.2\text{ k}} = -1.07\text{ V}$$

Similarly at 100°C

$$\Delta R = \frac{(-1\text{ k})}{°\text{C}} \times (100°\text{C} - 25°\text{C}) = -75\text{ k}\Omega$$

and $$V_o = \frac{-(-75\text{ k}) \times (5)}{2(240\text{ k} + 74\text{ k})} \times \frac{10\text{ k}}{2.2\text{ k}} = 5.17\text{ V}$$

Hence, when V_o = – 1.07 V, the meter dial can be marked as 0°C and when V_o = 5.17 V, the meter dial can be marked as 100°C. But at 25°C. V_o = 0, hence a centre zero meter is required. Assuming that the resistance temperature characteristics of the thermistor is linear, the meter may be interpolated linearly from 25°C – 0°C and 25°C –100°C.

14.4.2 Light Intensity Meter

If in the bridge circuit of Fig. 14.25, the transducer is a photocell, this circuit can be used as a light intensity meter. Initially the bridge is balanced for the dark condition, therefore when exposed to light the bridge becomes unbalanced, to a degree depending upon the intensity of light. This unbalance causes the meter to deflect. To measure the change in light intensity, the meter is calibrated in terms of Lux or Lumen.

The light intensity meter can also be designed using a single input inverting or non-inverting opamp, but the light intensity meter using an instrumentation amplifier is more accurate and stable, because the common mode (noise) voltages are effectively cancelled by the differential mode.

14.4.3 Analog Weight Scale

Figure 14.25 can be converted into a simple analog weight scale by connecting strain gauges in the bridge circuit. These strain gauges are connected in all the four arms of the bridge,as shown in Fig. 14.26. The strain gauge elements are mounted on a base of the specially made weight platform, on which an external force or weight is placed. One pair of strain gauge elements in opposite arms elongates, (i.e. R_{T1} and R_{T3} both increases in resistance) while the other pair compresses (R_{T2} and R_{T4} both decreases in resistance), and vice-versa.

The bridge is balanced when no external force or weight is applied, i.e. $R_{T1} = RT_2 = RT_3 = RT_4 = R$, and the output voltage of the weight scale is zero.

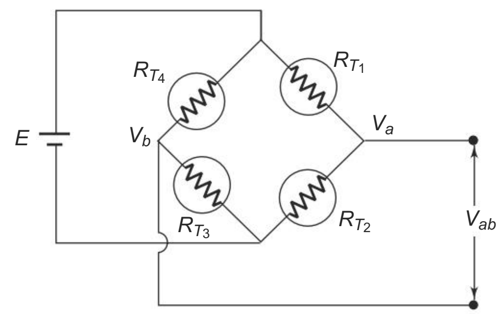


Fig. 14.26 Strain gauge bridge circuit for analog weight scale

Suppose a weight is placed on the scale platform and R_{T1} and R_{T3} increases in resistance. Then R_{T2} and R_{T4} decrease in resistance by the same value ΔR and the bridge is unbalanced, thereby giving an unbalanced output voltage. This unbalanced voltage V_{ab} is given by

$$V_{ab} = +E\left(\frac{\Delta R}{R}\right).$$

where E – excitation voltage of the bridge.

$R = R_{T1} = R_{T2} = R_{T3} = R_{T4}$ = unstrained gauge resistance

ΔR – change in gauge resistance.

The differential instrumentation amplifier then amplifies the voltage V_{ab}, giving a deflection on the meter movement. As the gain of the amplifier is $(+R_F/R_1)$, the output voltage V_o is given by

$$V_o = E \times \left(\frac{\Delta R}{R}\right) \times \left(\frac{R_F}{R_1}\right).$$

The gain of the amplifier is selected depending on the sensitivity of the strain gauge and on the full scale deflection requirements of the meter. The meter can be then calibrated in grams or kilograms.

For better accuracy and resolution, a micro based digital weight scale may be constructed. However, such a scale is much more complex and expensive then the analog scale.

Example 14.4 *The circuit of Fig. 14.25, is converted into an analog weight scale with the help of four strain gauge elements, each having an unstrain resistance of 50 Ω.*

Assume that the gain of the differential instrumentation amplifier is 100, E = + 10 V and the opamp supply voltages = ± 15 V.

When a certain weight is placed on the scale platform, the output voltage is 1.5 V. Determine the change in resistance of each gauge element. (Assuming that the output voltage is initially zero.)

Solution Using the formula

$$V_o = E\left(\frac{\Delta R}{R}\right) \times \left(\frac{R_F}{R_1}\right)$$

$$1.5 = 10\left(\frac{\Delta R}{100}\right) \times 100$$

Therefore $\quad \Delta R = \dfrac{1.5}{10} = 0.15\ \Omega$

This means that R_{T1} and R_{T3} decrease by 0.15 Ω and R_{T2} and R_{T4} increase by 0.15 Ω when a certain weight is placed on the scale platform.

CHOPPED AND MODULATED DC AMPLIFIER 14.5

A simple ac amplifier may be used to amplify a dc input through the use of additional circuit component known as chopper. In this circuit, the dc signal is first converted into an ac signal, amplified by a standard amplifier, and finally converted back to a dc signal. The chopper can be electronic or mechanical.

Figure 14.27 shows a chopper type dc amplifier. V_i is the input dc voltage, this voltage is alternately connected to terminals A and B. When the switch is in position A, the direction of current flow of the current is in one direction. When the switch is connected to position B, the current flows in the opposite direction.

This means that an ac voltage will be induced in the secondary winding of the input transformer.

For an ideal transformer, this voltage is of perfect square wave shape. The peak value of the induced voltage is proportional to the dc input. The ac signal is amplified by a standard ac amplifier. An amplified square wave appears at the primary winding of the output transformer. The ac signal is converted back to dc. The secondary winding of the output transformer is centrally tapped, with an output switch ganged (mechanically coupled) to the input switch. The input voltage across the primary winding of the output transformer is shown in Fig. 14.28 (a) and the output of the secondary winding is shown in Fig. 14.28(b).

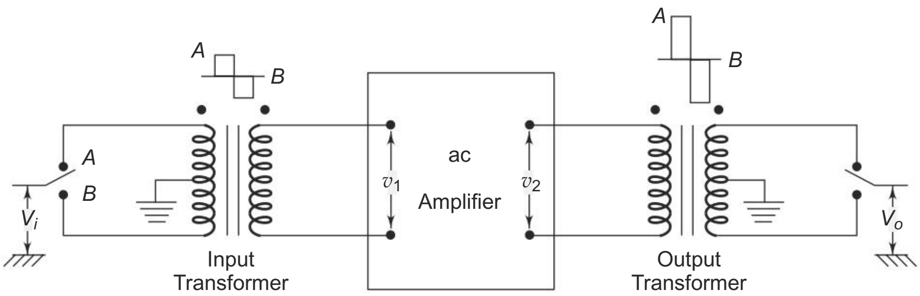


Fig. 14.27 Chopper type dc amplifier

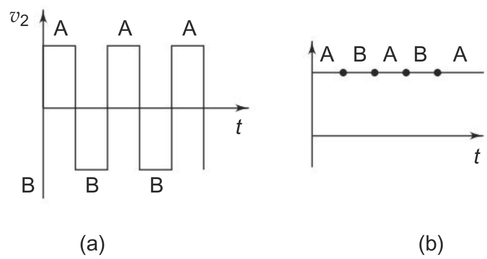


Fig. 14.28 (a) Input voltage across the primary windings of output transformer
(b) Output of the secondary winding of output transformer

The amplifier of Fig. 14.27 is referred to as a chopper amplifier, since the dc input voltage is literally chopped to produce ac signal. The chopping action may be accomplished by either mechanical or electronic means. A vibrating reed is used when mechanical methods are employed.

MODULATORS 14.6

An ac amplifier must see ac in order to perform useful work. A fair number of chopping and modulation techniques are available to change direct current, or to change very slowly varying signals into square waves (or pulses). Associated circuitry, inserted at the amplification system common output, may then be used to rectifier or demodulate the carrier wave and thereby produce a common dc product.

Small dc increments are converted to pulses, suitable for amplification in high gain ac stages, in an electro-mechanical chopper shown in Fig. 14.29. The device is basically identical to the well known dc (vibrator) chopper.

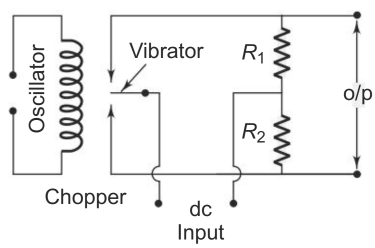


Fig. 14.29 Electromechanical chopper

The chopper is usually driven by a medium frequency oscillator, with drive frequencies for the coil between 50–400 Hz or more. High quality choppers handle voltages as high as 100 V, and voltages in the microvolt range. Most contacts are composed of similar metals and carry a maximum current of 2 mA. The approximate life is about 5000 hours.

14.6.1 Self Oscillator Converter (Transistorised)

A transistorised dc to ac converter suitable for low level dc signal is illustrated in Fig. 14.30.

The unique property of this assembly is that it starts to oscillate vigorously at very low levels of dc collector voltages at only a few micro-watts of collector power. The circuit shown in Fig. 14.30, is connected in a tickler feedback, common emitter configuration. Referring to Fig. 14.31, the centre tapped transformer provides the output for the ac amplifier. The dc gain can be varied, within limits, by the 100 k potentiometer connected in the collector circuit. The AF oscillator is completely inoperative, unless an input signal is applied to it. The polarity must be reversed for a PNP transistor. Rectification of the ac output may be done with the simple full wave rectifier circuit. This circuit connected to the amplifier output secondary transformer is shown in Fig. 14.31.

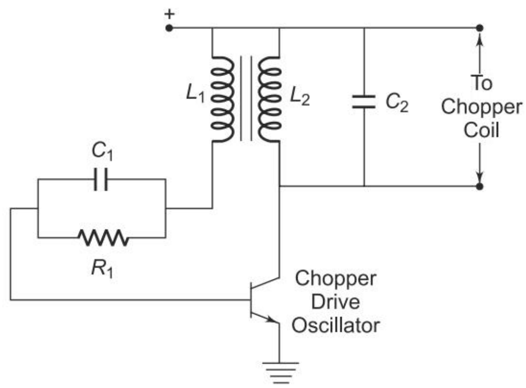


Fig. 14.30 Transistorised chopper circuit

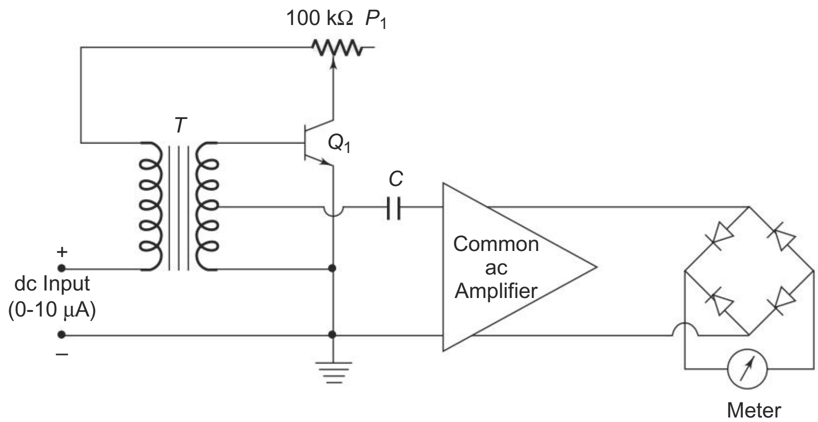


Fig. 14.31 dc to ac converters for an ac amplifier

14.6.2 Magnetic Modulator

The most reliable substitutes for electromechanical choppers are magnetic modulators. These devices reach a null stability as low as 10 μV, but their response time is often less than 2 Hz and the ambient temperature range is restricted. The 3-leg saturable reactor, shown in Fig. 14.32 falls within the category of magnetic modulator devices. A magnetic modulator is based on the principle of saturation. If dc product is applied to windings L_1 and L_2, the core will be saturated according to the B–H loop characteristics.

As shown in Fig. 14.32, L_2 is the ac excitation winding and $L_1 - L_3$ are the dc winding.

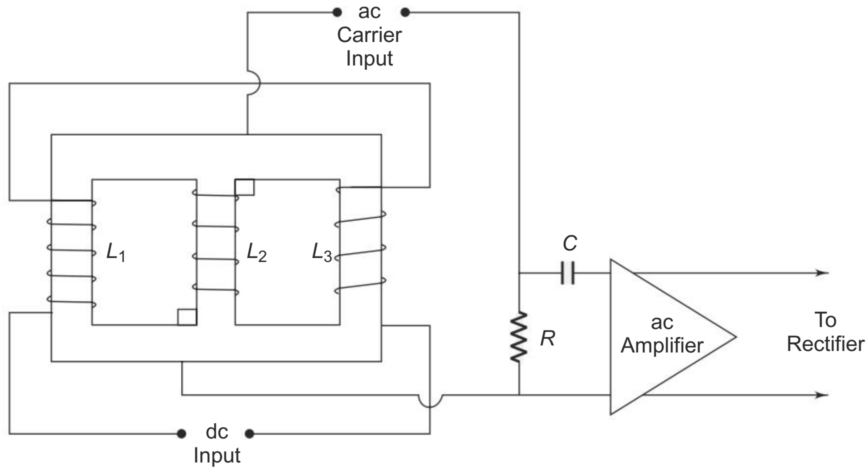


Fig. 14.32 Magnetic modulator

A given degree of magnetic saturation causes a change of inductance in L_2. If no dc voltage is applied to the reactor, the impedance of L_2 will set the value of a given ac current flowing through it. Now if dc voltage is applied, it causes an increase in the core saturation changing the L_2 impedance characteristics which in turn causing a change in ac current seen by the follow-up ac amplifier. The dc excitation windings are connected in Buck type configurations. This configuration prevents the ac current to enter the dc windings by transformer action.

14.6.3 Diode Bridge Modulator

Figure 14.33 shows a silicon diode ring modulator with an associated ac coupled dc amplifier. The amplifier itself has a gain of 65 db and a flat response within ± 1 db, from 8 Hz to 80 kHz. Precautions are taken to safeguard the amplifier against sudden surge voltages having an active magnitude above the supply voltage of 9 V dc by using a 10 V zener diode as a guard element.

The function of diode bride modulator can be best understood from Fig. 14.33 (a). The device can be regarded as a polarity sensitive switch in which the ac excitation current cycles turn the input dc ON or OFF.

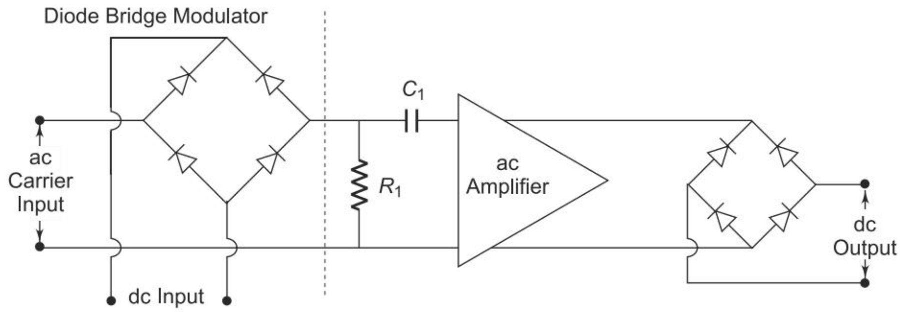


Fig. 14.33 (a) Diode bridge type modulator

The signal is developed across the resistor R_1 and coupled to the amplifier via the capacitor C_1. Each pair of diodes conducts on alternate half cycles of the ac excitation. During one half cycle, the effect is to open the path between the dc signal input and the follow up ac amplifier. During the interval of the other cycle, the conductance path is closed. A transformer may be inserted for electrical isolation and/or voltage boosting as shown in Fig. 14.33 (b). The modulator circuit shown in Fig. 14.33 (b) uses a conventional diode bridge which modulates low level dc signal (across R_2), amplify the modulated signal, and then demodulates it to obtain the dc signal at high level. The centre taps of the transformer are critical for successful operation, and the silicon diodes used here require matched forward characteristics and a reverse current smaller than 10^{-8} A. The output waveform is basically a square wave filtered by the output transformer. The amplitude of the output signal available across R_3 is proportional to the magnitude of the dc input signal. The phase of the output signal with respect to the carrier signal is proportional to the sign of the dc signal.

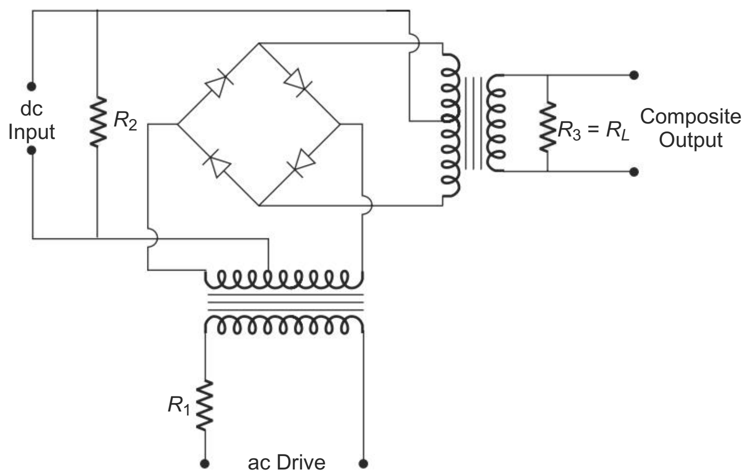


Fig. 14.33 (b) Diode bridge modulator with transformer coupling

14.6.4 Transistor Choppers

A single transistor used as a dynamic switch to convert low level dc signal to an ac waveform is shown in Fig. 14.34.

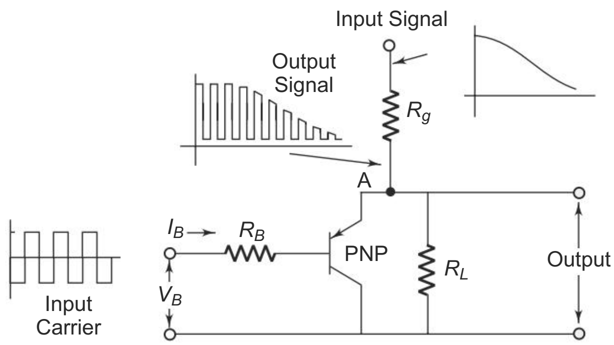


Fig. 14.34 Transistor chopper

If the voltage applied to the transistor base is negative (in the case of a PNP transistor), both emitter and collector diodes are shorted, and the output voltage approximates V_1 as shown in Fig. 14.35. However, if a positive base voltage is applied to the chopping transistor (PNP), the diodes are open-circuit. The resistance from point *A* in Fig. 14.34 will then be in the Meg ohms range, and signal V_B applied to the input of this circuit, is not attenuated.

Devices of this type use low drive power and have a good frequency response, but there is a leakage current. Since the junctions of a transistor are in physical contact with each other, there is a residual feed through of signals from the drive circuit to the output circuit, and such a minor signal may be interpreted by the following amplifier as a true input signal requiring amplification. If no voltage is applied at the composite system input, the voltage at point A (Fig. 14.34) during the high impedance portion of the cycle is approximately $-I_1 \times (R_g R_L)/R_g + R_{\Delta,}$ where R_L is load resistance.

Some typical coordinates of point *B* (Fig. 14.35) are

$$V_1 = 1 \text{ mV dc } I_1 < 1\ \mu\text{A}.$$

Some of the temperature dependent drift of voltage V_1 may be reduced by the use of two transistors whose connection scheme is shown in Fig. 14.36. The coordinate values of the matched silicon transistor may thereby made to remain between ± 0.2 mV dc and = ± 0.2 µA over a very wide temperature range. If R_g or R_L is kept at a small value, undesired drift due to current I_1 can be minimised.

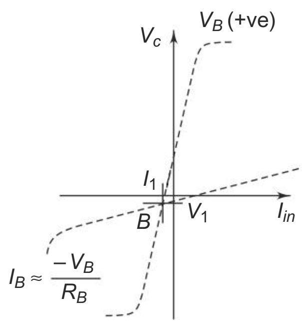


Fig. 14.35 Operating characteristics of a transistor chopper

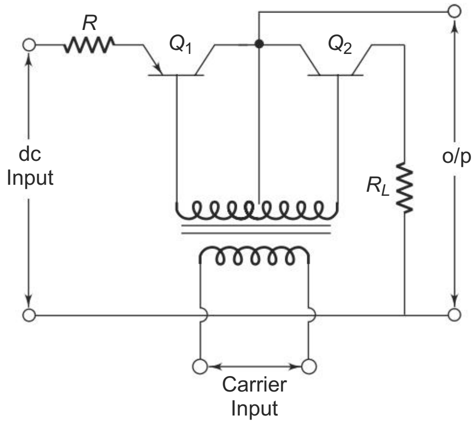


Fig. 14.36 Balanced transistor chopper

In any chopper circuit, keeping the active chopper elements at constant temperature reduces drift. If two elements are used to cancel drift, it is important to keep the two at the same temperature, and a good way to prevent temperature drift is to make both elements a part of the same integrated circuit. This is done by the use of an integrated double emitter bipolar transistor.

14.6.5 Capacitance Modulator

A specially processed junction diode that assumes the properties of a variable capacitor (about 5 – 250 pf) when its dc reverse voltage is varied is called a Varactor diode shown in Fig. 14.37. This unique behaviour can be used in special modulator circuits designed to turn a low voltage dc signal input into an ac output of proportional magnitude. A representative circuit is shown in Fig. 14.38.

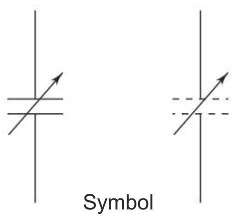


Fig. 14.37 Varactor

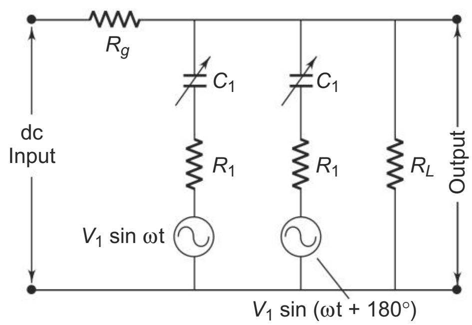


Fig. 14.38 Capacitance modulator

The modulator is essentially a 'pumped' parametric amplifier which can provide power gain. Assuming that the voltage variable capacitor (diodes) have the form $C_1 = C_o + Kf(v)$,

where C_o = leakage capacitance,

K = constant, and

$f(v)$ = variable voltage function

The magnitude of the output voltage at R_L can be shown to be proportional to the product of the magnitude of the carrier signal and the dc input signal supplied ahead of R_g. The system's pump sources (carrier insertion) are provided by V_1 and V_2. These modulating units are usually composed in bridge configuration and the drift voltage of the modulator is limited by the precise balance of the two varicaps (varactor diode).

14.6.6 Synchronous Modulator and Demodulator

Low dc current may be transformed into a high voltage dc by simple chopper action. Although an inductive type transformation process is required, the output dc may be obtained without rectifying devices.

Figure 14.39 shows a circuit function of a synchronous vibrator. Contacts *a* and *b* connect the input dc to the primary side of the transformer in an alternating fashion, thus producing an ac product. By adding another set of points to the normally composite vibrator reed, the output of the transformer can be rectified whenever a and b close the circuit in unison with the primary side contacts *a′* and *b′*. The output product will be unfiltered direct current. This principle of coinciding phase switching may be used in dc amplifier chopping circuits.

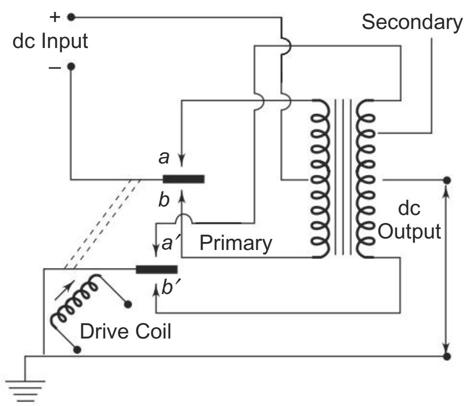


Fig. 14.39 Synchronous vibrator

A functional diagram of a synchronous chopper is shown in Fig. 14.40. Using this chopper it is possible to obtain an amplified output signal of the original dc input from the ac amplifier. In this circuit no rectification is necessary to produce the dc signal. The proper polarity of the signal is always maintained in the proper phase.

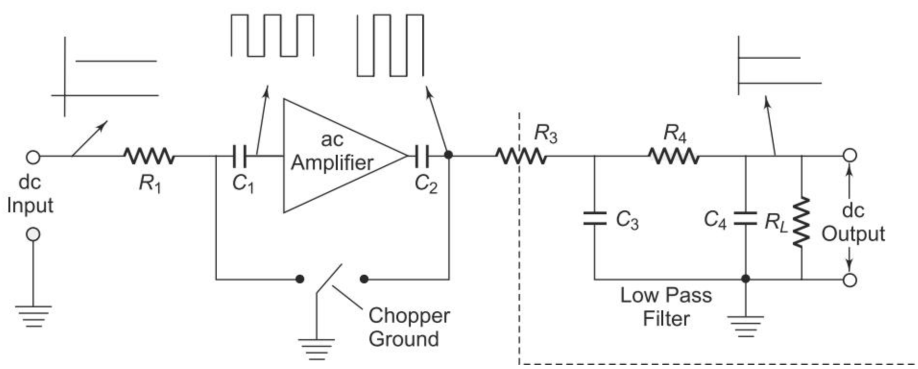


Fig. 14.40 Functional circuit of a chopper dc amplifier

14.6.7 Solid State Modulator/Demodulator Circuit

A solid state modulator/demodulator circuit is shown in Fig. 14.41. The transformer T is driven by an ac source and couples each secondary connected to the two transistor pairs, $Q_1 - Q_2$ and $Q_3 - Q_4$.

During the first half cycle, point X is positive with respect to point Y. This forward biases transistors $Q_1 - Q_2$, which are driven in full conduction. The resistance of the transistors falls. The current flow through the path is shown in Fig. 14.41 (arrow A). The other transistor $Q_3 - Q_4$ is reverse biased and its resistance increases. Therefore the output voltage increases.

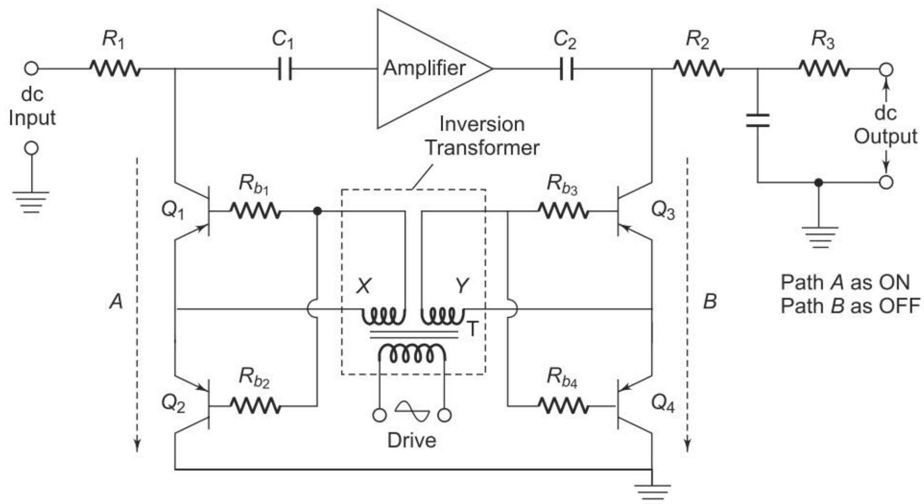


Fig. 14.41 Solid state modulator/demodulator

During the second half cycle, the polarity reverses, making Y positive with respect to X; now transistor $Q_3 - Q_4$ conducts fully, reducing its resistance, and output. Transistor $Q_1 - Q_2$ are reverse biased. Its resistance increases, hence the output developed across it increases. Thus, it is seen that $Q_1 - Q_2$ act as a modulator which feeds an ac signal to the ac amplifier and $Q_3 - Q_4$ act as a demodulator, which demodulates the amplified ac signal to obtain dc again. The dc output obtained is an amplified value of the dc input.

14.6.8 Photo Optical Modulator

The principle of an optical dc chopper is illustrated in Fig. 14.42. Devices of this type have been used widely in infrared signal detectors, whose output is a slowly varying dc product. The properties of the parallel connected photo resistor are such that the device produces a decrease in its internal resistance when struck by light. The photo resistor may have a dark resistance of several Meg-ohms, but has a dynamic resistance of 6 Ω or less when light falls on it. (In order to make use of this effect in a chopping system, a motor driven chopping disc supplied with a slit is rotated in front of the photo resistor. The latter's chopping frequency is determined by the number of the slots in the disc, and by the speed at which it is being rotated. A 1 slot disc, being rotated at 3600 rpm, would thus produce a chopping frequency of 60 Hz. Two slots with rpm constant give 120 Hz; and with 10 slots the frequency would be 600 Hz.) The bulb may be driven by a simple relaxation oscillator in place of the motor.

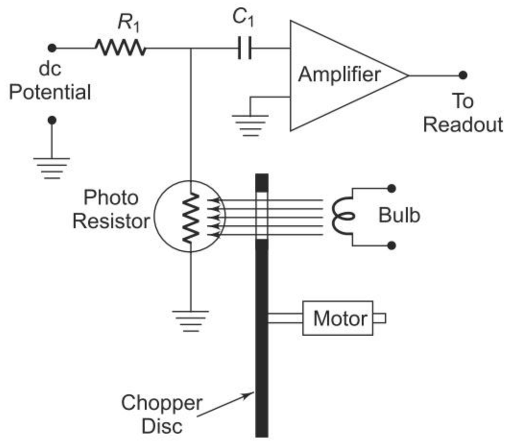


Fig. 14.42 Photo optical dc chopper

Review Questions

1. What is a signal conditioner? What are the basic elements of a signal conditioner?
2. What is an opamp?
3. What are the important features of an opamp?
4. How is an opamp used as a/an
   (i) inverting and non-inverting amplifier.
   (ii) summing amplifier.
   (iii) integrator.
   (iv) differentiator
   (v) comparator.
5. What is an instrumentation amplifier?
6. What are the applications of an instrumentation amplifier? What are Choppers? How is the drift in a high gain dc amplifier eliminated?
7. State the different types of chopper.
8. List the differences between electronic and mechanical choppers.
9. On what principle does the mechanical chopper operate?
10. Explain the working of a capacitance modulator.
11. Explain the working of a synchronous modulator/demodulator.
12. Explain how a transistor can be used as a chopper.
13. Explain the working of a solid state modulator/demodulator.
14. Explain the working principle of an optical chopper.

Further Reading

1. G.B. Clayton, *Operational Amplifiers*, Butterworth & Co., 1979.
2. Ramakant Gayakwad, *Op Amps & Linear Integrated C & T Technology*, Prentice-Hall of India, 1987.
3. Coughlin and Driscoll, *Op Amps & Linear Integrated Circuits*, 3rd Edition, Prentice-Hall of India Pvt. Ltd., 1987.
4. Ernest. O. Doebelin, *Measurement Systems Applications & Design*, 4th Edition, McGraw-Hill, 1990.
5. L. George Lawrence, *D.C. Instrumentation Amplifiers*, W. Foulsham & Co. Ltd, 1965.

The McGraw·Hill Companies

Chapter

Filters 15

INTRODUCTION 15.1

A network designed to attenuate certain frequencies but pass others without attenuation is called a filter. A filter circuit thus possesses at least one pass band, which is a band of frequencies in which the output is approximately equal to the input (attenuation is zero) and an attenuation band in which output is zero (attenuation is infinite). The frequencies that separate the various pass and attenuation bands are called the cutoff frequencies.

An important characteristics of all filter networks is that they are constructed from purely reactive elements.

In measurement systems, the transducer often does not measure the physical parameter (measurand) precisely. The information present is not in standard form, hence the need for further filtering and analysis. The main factors that must be considered for accuracy are signal to noise (S/N) ratio, response time and bandwidth over which the measurements are desired. Among these the S/N ratio is the most important parameter that needs to be considered, and the use of a signal filter becomes a necessity when low level or high resolution measurements are required. The signal originating from a transducer is fed to the signal conditioner. The output signal obtained from a transducer must be reproduced faithfully. In order to obtain a faithful reproduction of the signal, it becomes necessary to eliminate any spurious or unwanted signals which may get introduced into the system, either at the transduction stage or at the signal conditioning stage.

The filter are thus designed to pass signals of desired frequencies and reject signals of unwanted frequencies (harmonics and noise).

FUNDAMENTAL THEOREM OF FILTERS 15.2

Assuming the filter is correctly terminated in its characteristic impedance Z_o the following is applicable.

"Over the range of frequencies for which the characteristic impedance Z_o of a filter is purely resistive (real), the attenuation (α) will be zero."

"Over the range of frequencies for which the characteristic impedance Z_o is purely reactive (imaginary), the attenuation (α) will be greater than zero."

Mathematical Proof of the Theorem The fundamental theorem on filters can be proved by considering a simple filter circuit in the form of a symmetrical T network. Let the T network consist of series and shunt elements, as shown in Fig. 15.1(a).

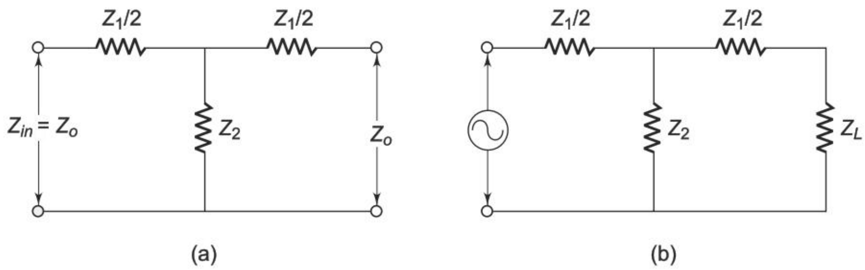


Fig. 15.1 (a) T-network (b) Electrical circuit

The characteristic impedance Z_o of the network in Fig. 15.1 is given by

$$Z_o = \sqrt{\frac{Z_1^2}{4} + Z_1 Z_2}$$

Similarly, the current ratio or the propagation constant

$$e^{\gamma} = \frac{Z_1}{2Z_2} + \frac{Z_o}{Z_2} = \frac{I_s}{I_R}$$

Being a filter network, both Z_1 and Z_2 are reactive elements. Let $Z_1 = jX_1$ and $Z_2 = jX_2$.

There will be two cases for impedance values of Z_o, depending on the value and sign of X_1 and X_2.

Hence $$Z_o = \sqrt{j^2\frac{X_1^2}{4} + j^2 X_1 X_2} = \sqrt{-\left(\frac{X_1^2}{4} + X_1 X_2\right)} \tag{15.1}$$

Similarly $$e^{\gamma} = \frac{jX_1}{2jX_2} + \frac{Z_o}{jX_2}$$

$$= 1 + \frac{X_1}{2X_2} - j\frac{\left(\sqrt{-\left(\frac{X_1^2}{4} + X_1 X_2\right)}\right)}{X_2}$$

$$e^{\gamma} = 1 + \frac{X_1}{2X_2} - j\frac{Z_o}{X_2} \tag{15.2}$$

Z_o will be real if $\frac{X_1^2}{4} + X_1 X_2$ is negative (say equal to $-A$)

Z_o will be imaginary if $\frac{X_1^2}{4} + X_1 X_2$ is positive.

Case 1 When Z_o is real

Let $$\left[\frac{X_1^2}{4} + X_1X_2\right] = -A$$

Then from Eq. (15.1) we get

$$Z_o = \sqrt{-(-A)} = \sqrt{A}$$

Also from Eq. (15.2) we have

$$e^{\gamma} = 1 + \frac{X_1}{2X_2} - j\frac{\sqrt{A}}{X_2} = \frac{I_s}{I_R} \tag{15.3}$$

Since $$Z_o = \sqrt{A}$$

but $$e^{\gamma} = e^{\alpha + j\beta} = e^{\alpha} + e^{j\beta}$$

where α = attenuation constant

β = phase shift constant

Considering attenuation only

$$|e^{\gamma}| = e^{\alpha}$$

$\therefore$ From Eq. (15.3)

$\therefore$ $$e^{\alpha} = \left|\frac{I_s}{I_R}\right| = \sqrt{\left(1 + \frac{X_1}{2X_2}\right)^2 + \left(\frac{\sqrt{A}}{X_2}\right)^2}$$

$$= \sqrt{1 + \frac{2X_1}{2X_2} + \frac{X_1^2}{4X_2^2} + \frac{A}{X_2^2}}$$

But as $$A = -\left(\frac{X_1^2}{4} + X_1X_2\right)$$

$\therefore$ $$e^{\alpha} = \sqrt{1 + \frac{X_1}{X_2} + \frac{X_1^2}{4X_2^2} - \frac{\left(\frac{X_1^2}{4} + X_1X_2\right)}{X_2^2}}$$

$\therefore$ $$e^{\alpha} = \sqrt{1 + \frac{X_1}{X_2} + \frac{X_1^2}{4X_2^2} - \frac{X_1^2}{4X_2^2} - \frac{X_1}{X_2}}$$

$$e^{\alpha} = \sqrt{1}$$

$\therefore$ $$e^{\alpha} = 1$$

$\therefore$ $$\alpha = \log_e 1 = 0 \tag{15.4}$$

From Eq. (15.4) we see that attenuation is zero, if Z_o is real.

Case 2 When Z_o is imaginary [purely reactive]

For Z_o to be imaginary $X_1^2/4 + X_1\,X_2$ should be positive.

$\therefore$ let $$\frac{X_1^2}{4} + X_1 X_2 = B$$

$\therefore$ from Eq. (15.1) we get

$$Z_o = \sqrt{-\left(\frac{X_1^2}{4} + X_1 X_2\right)} = \sqrt{-(B)}$$

$$Z_o = \sqrt{(-1)(B)} \text{ as } j^2 = -1, j = \sqrt{(-1)}$$

$\therefore$ $$Z_o = j\sqrt{B}$$

Similarly the value of e^{γ} from Eq. (15.2) will be

$$e^{\gamma} = 1 + \frac{X_1}{2X_2} - \frac{j\left(j\sqrt{B}\right)}{X_2} = \frac{I_s}{I_R}$$

as $$Z_o = j\sqrt{B}$$

$$e^{\gamma} = 1 + \frac{X_1}{2X_2} + \frac{\sqrt{B}}{X_2} = \frac{I_s}{I_R}$$

$\therefore$ $$e^{\gamma} = \left(1 + \frac{X_1}{2X_2}\right) + \frac{1}{X_2}\left(\sqrt{\frac{X_1^2}{4} + X_1 X_2}\right) = \frac{I_s}{I_R}$$

Taking the modulus

$$\left|\frac{I_s}{I_R}\right| = \left(1 + \frac{X_1}{2X_2}\right)^2 + \left(\frac{1}{X_2}\left(\sqrt{\frac{X_1^2}{4} + X_1 X_2}\right)\right)^2 = e^{\alpha}$$

Since $$B = \frac{X_1^2}{4} + X_1 X_2$$

$$\left|\frac{I_s}{I_R}\right| = \left(1 + \frac{X_1}{2X_2}\right)^2 + \frac{1}{X_2^2}\left(\frac{X_1^2}{4} + X_1 X_2\right)$$

$$\left|\frac{I_s}{I_R}\right| = \left(1 + \frac{X_1}{2X_2}\right)^2 + \left(\frac{X_1^2}{4X_2^2} + \frac{X_1 X_2}{X_2^2}\right) = e^{\alpha}$$

$\therefore$ $$e^{\alpha} = \left(1 + \frac{X_1}{2X_2}\right)^2 + \left(\frac{X_1^2}{4X_2^2} + \frac{X_1}{X_2}\right) \tag{15.5}$$

Hence α is greater than one and real. Therefore if Z_o is imaginary, α cannot be zero, and the filter will have an attenuation band.

PASSIVE FILTERS 15.3

15.3.1 Introduction

Passive filters are mainly networks using inductors, capacitors and resistors. The classical theory employed was based on the image parameter theory which in turn was based on the filter's characteristics and performance. Its component values were calculated by considering a source having a specified source resistance and feeding it into a constant load impedance called the termination impedance, resulting in constant – K type filter or prototype filter. This filter is a T or π section in which the series arm Z_1 and shunt Z_2 are connected by the relationship $Z_1 \cdot Z_2 = R_o^2$ where R_o is a real constant called the design impedance.

The constant – K filter section cannot have an idealised attenuation versus frequency characteristic. Hence the response characteristics in the attenuation band can be modified to an almost ideal respone curve by suitable changes in the shunt or series arm without affecting the design impedance of the filter circuit. Such derived filter circuits of the prototype filter section are called *m*-derived filters, where *m* is the design parameter, which is a constant. Constant-K filters were often modified with *m*-derived end sections both at the input and output side, to obtain better impedance matching between the source and filter network on the input side and between the filter and load on the output side.

One of the main difficulties with these passive filters was the necessity of always using the specified source and termination impedance. The cascading of these filters was also not straight forward, in spite of designing the filters so that their source and termination impedance were equal. Additional isolation amplifiers often had to be used between cascaded sections. This was done to prevent severe distortion of the filter characteristics due to non-ideal matching of impedance between the filter end sections and the source, as well as the terminating load. Another disadvantage was the necessity of using bulky and often nonlinear inductances for low and very low frequencies. Due to the relatively low values of inductive reactance at low frequency, in addition to the non-linearity at high current levels (due to saturation of cores), the circuit was designed to keep the signal levels low.

The availability of opamps in integrated form has changed the saturation significantly, giving rise to active filters.

Today, a majority of low frequency filters are of this type, particularly for frequencies below 100 kHz. The special advantage of the active circuit for use in low frequency filters is the fact that inductors can be totally eliminated. In addition, active capacitance multiplication enables the use of capacitors of low practical values even for cutoff frequencies down to a fraction of 1 Hz. However due to the limited gain-BW product of the ICs and their effect on the filter characteristics, and due to the advantages of the inductor in high frequency range, passive filters are preferred for frequencies above a few 100 kHz.

15.3.2 Types of Filters

Filters may be of any physical form—electrical, mechanical, pneumatic, hydraulic, acoustical, etc. The most commonly used filters are of the electrical type. Filters may be classified as follows.

1. Low pass
2. High pass
3. Band pass
4. Band stop
5. All pass

Consider a basic configuration of an electrical filter shown in Fig. 15.1 (b). The source is sinusoidal, of variable frequency. The filter circuit may be so designed that some frequencies are passed from the input to the output of the filter with very little attenuation (pass band) while others are greatly attenuated (attenuation band).

The responses of various filters are shown in Fig. 15.2. These are ideal responses that cannot be achieved in actual practice.

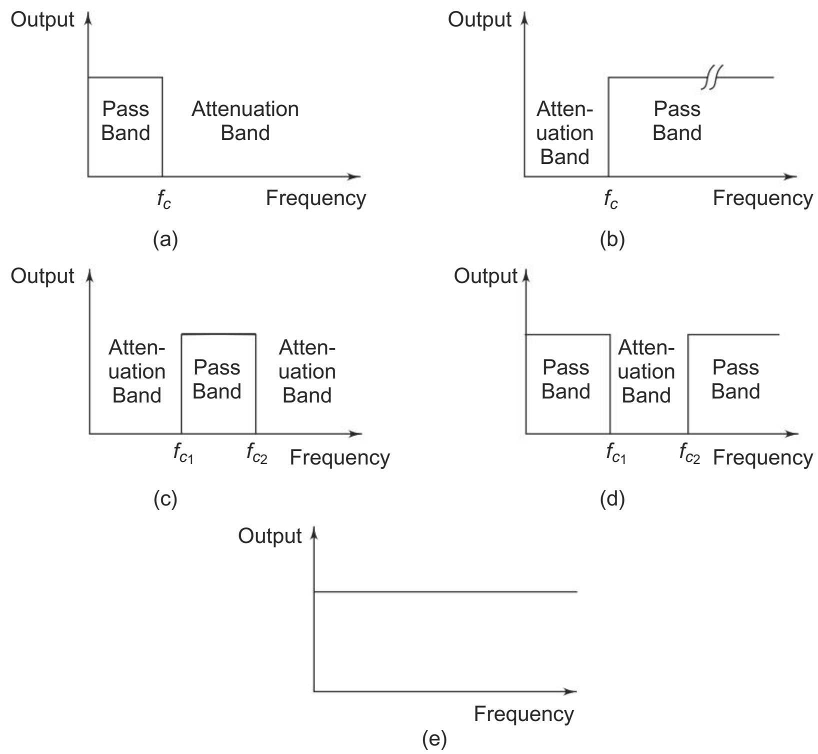


Fig. 15.2 Ideal response of various filters (a) Low pass (b) High pass (c) Band pass (d) Band stop (e) All pass

The ideal response of a low pass filter is shown in Fig. 15.2 (a). The voltage gain is given by output/input.

Hence $$\text{gain } A = \left|\frac{V_o}{V_{in}}\right|$$

This gain is constant over a frequency range starting from zero to a cut off frequency f_c. The output of any signal having a frequency greater than f_c will be attenuated, i.e. there will be no output voltage for frequencies greater than the cutoff frequency f_c. Hence output will be available faithfully from 0 to f_c with constant gain, and is zero from f_c onward.

The characteristics of a High Pass filter are shown in Fig. 15.2 (b). The HP filter has a zero gain starting from zero to a frequency f_c, the cut off frequency. Above this cutoff frequency, the gain is constant and equal to A. Hence signal of any frequency beyond f_c will be faithfully reproduced with a constant gain, and frequencies from 0 to f_c will be attenuated.

The band pass filter characteristics are shown in Fig. 15.2 (c). It faithfully reproduces signals falling between f_{c1} and f_{c2}, while signals between 0 and f_{c1}, and frequencies greater than $f_{c2}$are attenuated. There is an output corresponding to signals having frequencies between f_{c1} and f_{c2}, but no output for signals having frequencies below f_{c1} and above f_{c2}. Hence this filter passes a band of frequencies.

Figure 15.2(d) shows the characteristics of a band stop filter. This filter attenuates a particular band of frequencies from f_{c1} to f_{c2}, while passing all frequencies between 0 and f_{c1}, and f_{c2} onwards. This filter is also called a notch filter.

Figure 15.2(e) shows the characteristics of an all pass filter. In this filter all frequencies are passed without attenuation. The important feature of this filter is that it provides predictable phase shift for frequencies of different input signals. These filters are mostly used in communications.

The filters discussed above have ideal characteristics and a sharp cutoff. In actual practice the characteristics shown in Fig. 15.2 are not practicable in a practical filter circuit. Attenuation outside the pass band, that is in the attenuation band, is finite. The attenuation can be made sufficiently large by adding several filters in tandem. Also, it is not possible to have a pure inductor (inductor without resistance). Hence the attenuation in the pass band will not be zero. Only those characteristics of filters which can be practically realised are discussed below.

15.3.3 Basic Low Pass Filter

An electric low pass filter is shown in Fig. 15.3 (a). It is an RC network. At low frequencies, the capacitive reactance is very high and the capacitor circuit can be considered as open circuit. Under this condition, the output equals the input ($V_o = V_{in}$) or voltage gain is equal to unity. At very high frequencies, the capacitive reactance is very low and the output voltage V_o is small as compared with the input voltage V_{in}. Hence the gain fall and drops off gradually as the frequency is increased, as shown in Fig. 15.3 (b).

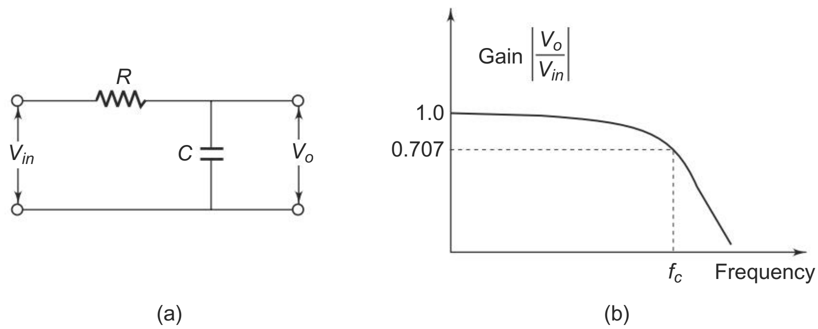


Fig. 15.3 Basic low pass filter and its frequency response (a) Low pass filter (b) Frequency response

The transfer function is

$$\frac{V_o}{V_{in}}(s) = \frac{1/sC}{R + 1/sC} = \frac{1}{1 + sRC} = \frac{1}{1 + s\tau} \tag{15.6}$$

The sinusoidal transfer function of a low pass filter is

$$\frac{V_o}{V_{in}}(j\omega) = \frac{1}{1 + (j\omega)RC} = \frac{1}{1 + (j\omega)\tau}$$

$$\text{Gain} = \quad A = \left|\frac{V_o}{V_{in}}(j\omega)\right| = \frac{1}{\sqrt{1 + (\omega RC)^2}} = \frac{1}{\sqrt{1 + (\omega C)^2}} \tag{15.7}$$

The gain drops to 0.707 at the cut off frequency ω_c

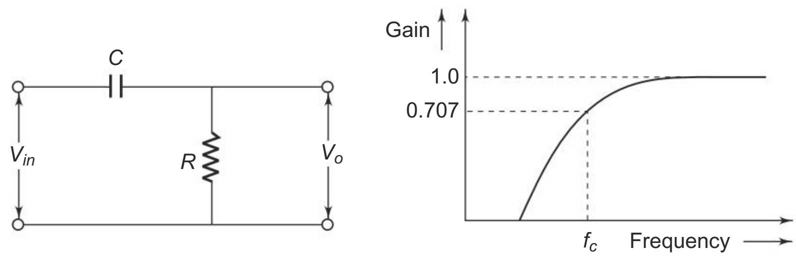
$$0.707 = \frac{1}{1 + (\omega_c RC)}$$

or cut off frequency

$$\omega_c = \frac{1}{RC} => f_c = \frac{1}{2\pi RC} \tag{15.8}$$

15.3.4 Basic High Pass Filter

An electrical high pass filter is shown in Fig. 15.4(a). When the frequency is low, the capacitive reactance is high, hence minimum output is available and the gain is small. When the frequency is high, the capacitive reactance is small, the output equals the input and the gain approaches unity. Hence this circuit passes high frequencies while rejecting low frequencies. The response of a high pass filter is as shown in Fig. 15.4(b).



(a) (b)

Fig. 15.4 (a) H.P. filter circuit (b) Frequency response

The transfer function of a high pass filter is given by

$$\frac{V_o}{V_{in}}(s) = \frac{R}{R+1/sC} = \frac{sRC}{1+sRC} = \frac{s\tau}{1+s\tau} \tag{15.9}$$

Sinusoidal transfer function of a high pass filter is

$$\frac{V_o}{V_{in}}(j\omega) = \frac{j\omega RC}{1+j\omega RC} = \frac{j\omega\tau}{1+j\omega\tau} = \frac{1}{1+1/j\omega\tau}$$

$$\text{Gain } A = \left|\frac{V_o}{V_{in}}(j\omega)\right| = \frac{\omega\tau}{\sqrt{1+(\omega\tau)^2}} \tag{15.10}$$

ACTIVE FILTERS 15.4

15.4.1 Introduction

An electric filter is often a frequency selective circuit that passes a specificed band of frequencies and blocks or attenuates signals of frequencies outside this band. Filters may be classified in a number of ways as follows.

1. Analog or digital filters
2. Passive or active filters
3. Audio (AF) or radio (RF) filters

Analog filters are designed to process analog signals using analog techniques, while digital filters process analog signals using digital techniques. Depending on the type of elements used in their construction, filters may be classified as active or passive (discussed earlier).

The elements used in passive filters are R, C and L. Active filters, on the other hand, employ transistors or opamps in addition to resistors and capacitors. The type of element used dictates the operating frequency range of the filter.

For example, RC filters are commonly used for audio or low frequency operation, whereas LC filters or crystal filters are employed at RF or high frequency. Because of their high Q value (figure of merit), the crystals provide stable operation at high frequency. Inductors are not used in the audio range because they are large, costly and may dissipate a lot of power. An active filter offers the following advantages over a passive filter.

1. Gain and Frequency Adjustment Flexibility Since the opamp is capable of providing a gain, the input signal is not attenuated, as in a passive filter. Also, an active filter is easier to tune.

2. No Loading Problem Because of the high input resistance and low output resistance of the opamp, the active filter does not cause loading of the source or load.

3. Cost Active filters are typically more economical than passive filters. This is because of the variety of cheaper opamps available, and the absence of inductors.

Although active filters are most extensively used in the field of communications and signal processing, they are employed in one form or another in almost all sophisticated electronic systems, radio, TV, telephones, radar, space satellites and biomedical equipment.

15.4.2 Classification of Active Filters

The most commonly used active filters are as follows.

1. Low pass
2. High pass
3. Band pass
4. Band stop
5. All pass

Each of these filters uses an opamp as the active element and R, C as the passive element.

Figure 15.5 shows the frequency response characteristics of the types of active filters. The ideal response is shown by the dashed lines, while solid lines indicate the practical filter response.

A low pass filter has a constant gain from 0 Hz to a high cutoff frequency f_H. Therefore, the band width is also f_H.

At f_H, the gain is down by 3 dbs, for $f > f_H$, it decreases with increase in input frequency. Frequencies between 0 Hz and f_H are known as pass band frequencies, while the range of frequencies beyond f_H, that are attenuated, are called the stop band frequencies.

Figure 15.5(a) shows the frequency response of a low pass filter. As indicated by the dashed line, an ideal filter has a zero attenuation in its pass band and infinite attenuation in the attenuation band. However, ideal filter response is not practical because linear network cannot produce discontinuities. But, it is possible to obtain a practical response that approximates the ideal response by using special design techniques as well as precision components and high speed opamps.

Butterworth, Chebyschev, Bessel and Elliptic filters are the most commonly used practical filters for approximating the ideal response. In many low pass filter applications, it is necessary for the closed loop gain to be as close to unity as possible within the pass band.

The Butterworth filter is best suited for this application. The key characteristic of the Butterworth filter is that it has a flat pass band as well as a flat stop band. For this reason it is sometimes called a flat-flat filter.

The Chebyschev filter has a ripple pass band and a flat stop band, while the Elliptic filter has a ripple pass band and a ripple stop band. Generally, the Elliptic filter gives the best stop band response among the three.

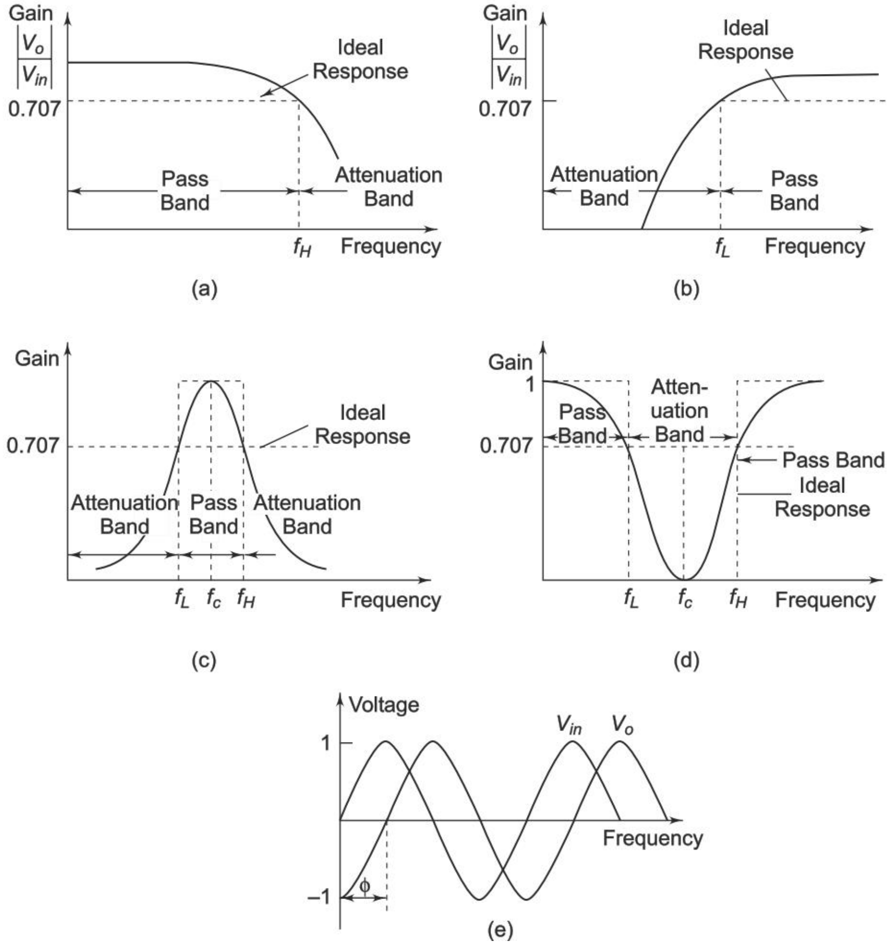


Fig. 15.5 Frequency response of major active filters (a) Low pass filter (b) High pass filter (c) Band pass filter (d) Band stop filter (e) Phase relation of input and output waveform in an all pass filter

Figure 15.5 (b) shows a high pass filter with a stop band $0 < f < f_H$ and a pass band $f > f_L$, where f_L is the low cutoff frequency and f is the operating frequency.

A band pass filter has a pass band between the two cutoff frequencies f_H and f_L, where $f_H > f_L$, and two stop bands at $0 < f < f_L$ and $f > f_H$. The bandwidth of the band pass filter therefore equals ($f_H - f_L$). The frequency response of the band pass filter is as shown in Fig. 15.5(c).

The band reject filter is exactly opposite to the band pass. It has a band stop between two cutoff frequencies f_H and f_L, and two pass bands, $0 < f < f_L$ and $f > f_H$. The band reject is also called a band stop or band elimination filter. The frequency response of a band stop filter is shown in Fig. 15.5(d). In this, f_c is called the centre frequency, since it is approximately at the centre of the pass or stop band.

Figure 15.5(e) shows the phase shift between the input and output voltages of an all pass filter. This filter passes all frequencies equally well, i.e. the output and input voltages are equal in amplitude for all frequencies, with the phase shift between the two, a function of frequency. The highest frequency up to which the input and output amplitudes remain equal is dependent on the unity gain bandwidth of the opamp. At this frequency, however, the phase shift between the input and output is maximum. For applications where the phase shift is important, the Bessel filter which has a minimal phase shift is used, even though its cutoff characteristics are not very sharp.

As shown in Figs 15.5 (a) to (d), the actual response curve of the filters in the stop band either steadily increase or decrease with increase in frequency. The rate at which the gain of the filter changes in the stop band is determined by the order of the filter. For example, in the case of Butterworth filters, for the first order low pass filter, the gain rolls off at rate of – 20 db/decade in the stop band ($f > f_H$). On the other hand, for the second order low pass filter the roll off rate is – 40 dB/decade and so on.

Figure 15.6 (a) shows the ideal response (solid line) and the practical (dashed lines) frequency response for three types of Butterworth low pass filters. As the roll of becomes steeper, they approach the ideal filter characteristics more closely.

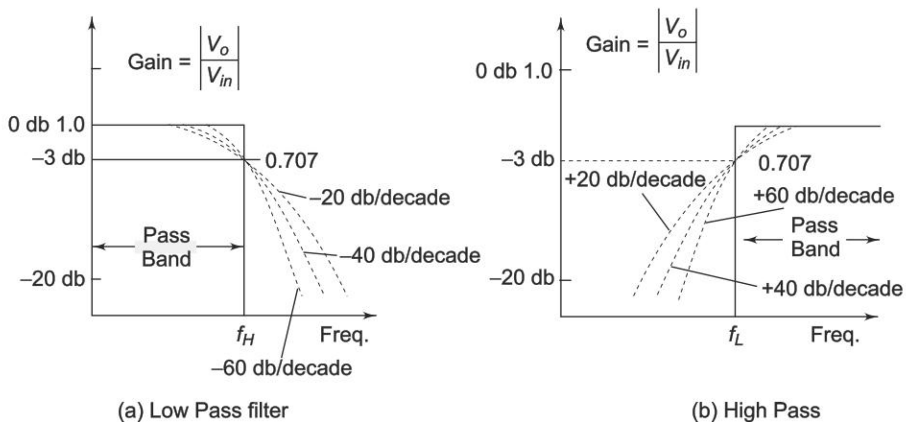


Fig. 15.6 Frequency response for three types of low pass and high pass butterworth filters (a) Low pass filter (b) High pass filter

By contrast, for the first order high pass filter, the gain increases at the rate of 20 db/decade in the stop band, that is, until $f = f_L$. The increase is 40 db/decade for the second order high pass filter and so on. Figure 15.6 (b) shows the frequency response for three types of high pass Butterworth filters.

BUTTERWORTH FILTER 15.5

In many low pass filter applications, it is necessary for the closed loop gain to be as close to unity as possible within the pass band. The Butterworth filter is best suited for this type of application. This filter is also called a maximally flat or flat-flat filter. Figure 15.6 shows the ideal (solid line) and practical (dashed line) frequency response for three types of Butterworth filters. As the roll offs become steeper, they approach the ideal filter response more closely.

15.5.1 Basic Low Pass Filter

The circuit of Fig. 15.7 is commonly used for low pass active filters. The filtering is done by the use of an RC network. The opamp is used as a unity gain amplifier. The resistor R_F is equal to R_1.

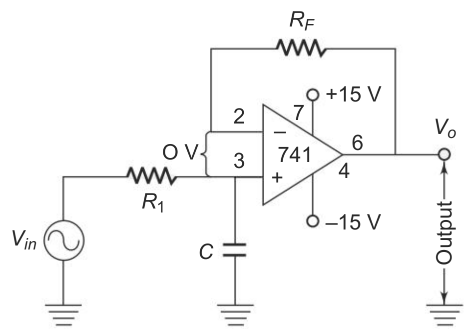


Fig. 15.7 Low pass filter for a roll off – 20 db/decade circuit

(At dc, the capacitive reactance is infinite and the dc resistance path to ground for both input terminals must be equal.) The difference voltage between inverting and non-inverting inputs is essentially O V. Hence, the voltage across the capacitor C equals the output voltage. Since this circuit is a voltage follower, V_{in} divides between R and C. The capacitor voltage V_o is given by

$$V_o = \frac{1/j\omega C}{R + 1/j\omega C} V_{in} \tag{15.11}$$

where ω is the frequency of V_{in} in radians per second ($\omega = 2\pi f$) and j is the imaginary term. To obtain the closed loop voltage gain A_{CL} we have,

$$A_{CL} = \frac{V_o}{V_{in}} = \frac{1}{1 + 1/j\omega CR} \tag{15.12}$$

Consider the Eq. (15.12). At very low frequencies, as ω approaches 0, $|A_{CL}| = 1$, and at very high frequencies, as ω approaches infinity, $|A_{CL}| = 0$. Hence this filter is a low pass filter.

Figure 15.8 shows a frequency response of ω versus $|A_{CL}|$. For frequencies greater than the cutoff frequency ω_c, $|A_{CL}|$ decreases at a rate of 20 db/decade. This is the same as saying that the voltage gain is divided by 10 when the frequency of ω is increased by 10.

The cutoff frequency is defined as that frequency of V_{in} where $|A_{CL}|$ is reduced to 0.707 times its low frequency value. The cutoff frequency is calculated from

$$\omega_c = \frac{1}{RC} = 2\pi f_c$$

$$\text{Therefore } f_c = \frac{1}{2\pi RC} \tag{15.13}$$

where f_c = is the cutoff frequency in Hz

R = resistance in Ω

C = capacitance in Farad.

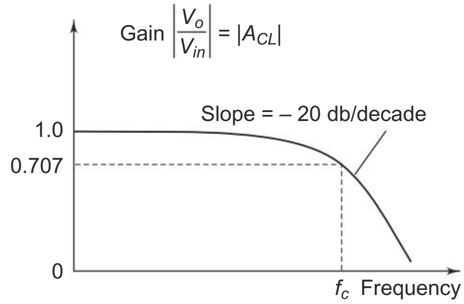


Fig. 15.8 Low pass active filter frequency response for a roll off of – 20db/decade

The Eq. (15.13) can be rearranged to solve R, ignore to give

$$R = \frac{1}{2\pi f_c C}$$

A first order low pass Butterworth filter can be obtained from the basic low pass filter using an RC filter network.

Figure 15.9 shows a first order low pass Butterworth filter that uses an RC network for filtering. The opamp is used in the non-inverting configuration, which does not load the RC network. R_1 and R_F determine the gain of the filter (in this case unity).

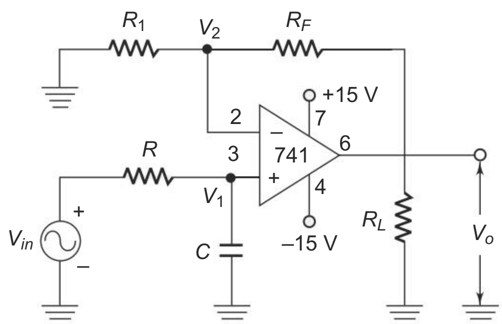


Fig. 15.9 First order low pass butterworth filter circuit

Using the voltage divider rule, the voltage across the capacitor, i.e. at the non-inverting input is

$$V_1 = \frac{-jX_c}{R - jX_c} \times V_{in}, \text{ but } -jX_c = \frac{1}{j2\pi f_C}$$

Simplifying, we get

$$V_1 = \frac{V_{in}}{j2\pi f CR + 1}$$

As output voltage $V_o = \left(1 + \dfrac{R_F}{R_1}\right) \times V_1$

Therefore $V_o = \left(1 + \dfrac{R_F}{R_1}\right) \times \left(\dfrac{V_{in}}{1 + j2\pi f CR}\right)$

Therefore $\dfrac{V_o}{V_{in}} = \dfrac{A_F}{1 + j2\pi f RC}$ But $f_H = \dfrac{1}{2\pi RC}$

$$\frac{V_o}{V_{in}} = \frac{A_F}{1 + j\left(f / f_H\right)} \tag{15.14}$$

where V_o/V_{in} = Gain of the filter as a function of frequency

$AF = 1 + R_F/R_1$ = pass band gain of the filter

f = frequency of the input signal

$f_H = 1/2\,\pi RC$ = high cutoff frequency

The gain magnitude and phase angle can be obtained by applying modulus to Eq. (15.14).

i.e.
$$\left|\frac{V_o}{V_{in}}\right| = \frac{A_F}{\sqrt{1 + \left(f / f_H\right)^2}} \tag{15.15}$$

and
$$\phi = -\tan^{-1}\left(\frac{f}{f_H}\right)$$

where ϕ is the phase angle in degrees.

The operation of the low pass filter can be verified from the gain magnitude Eq. (15.15).

1. At very low frequency, that is $f < f_H$,
$$\frac{V_o}{V_{in}} \cong A_F$$
2. At $f = f_H$, $\dfrac{V_o}{V_{in}} = \dfrac{A_F}{\sqrt{2}} = 0.707\, A_F$
3. At $f > f_H$, $\dfrac{V_o}{V_{in}} < A_F$

Hence the low pass filter has a constant gain, A_F, from 0 Hz to the high cutoff frequency f_H. At f_H, the gain is 0.707 A_F and after f_H, the gain decreases at a constant rate with increase in frequency; when the frequency is increased 10 times (one decade), the voltage gain is divided by 10. In other words, the gain decreases by 20 db (20 log 10) each time the frequency is increased by 10. Hence the rate at which the gain rolls off after f_H is 20 db/decade or 6 db/octave, where octave signifies a two fold increase in frequency. The frequency $f = f_H$ is called the cutoff frequency.

The procedure of converting a cutoff frequency to a new cutoff frequency is called frequency scaling.

To obtain a new cutoff frequency, R or C (but not both) is multiplied by the ratio of the original cutoff frequency to the new cutoff frequency.

In filter design, the values required for R and C are often not standard, and a variable capacitor C is not commonly used. Hence, we choose a standard value of the capacitor and then calculate the value of the resistor required for a desired cutoff frequency. This is because for a non-standing value of a resistor, a potentiometer can be used.

Example 15.1 *Design a low pass filter having a cutoff frequency of 2 kHz with a pass band gain of 2.*

Solution Following are the design steps.

1. $f_H = 2$ kHz
2. Let $C = 0.01$ µF
3. Then R is calculated as

$$R = \frac{1}{2\pi f C} = \frac{1}{2 \times 3.14 \times 2\,\text{kHz} \times 0.01\,\mu\text{F}} = \frac{1}{4 \times 3.14 \times 10\,\mu\text{F}}$$

$$= \frac{100}{4 \times 3.14} \times 10^3 = 7.95\ \text{k}\Omega \text{ (practical value 8.2 k)}$$

4. Since gain is given by $A_F = 1 + \frac{R_F}{R_1}$, therefore $\frac{R_F}{R_1} = A_F - 1$

$$\frac{R_F}{R_1} = 2 - 1 = 1 \text{ therefore } R_F = R_1$$

In this case $R_1 = R_F = 10$ kΩ is selected.

Example 15.2 *Referring to Fig. 15.9, calculate the value of resistance for a cutoff frequency of 20 k radians/s and C = 0.01 µF.*

Solution Given $\omega_c = 20$ k radians/s

$C = 0.01$ µF

$R =$ to be determined

As $$\omega_c = \frac{1}{RC},\ R = \frac{1}{\omega_c C} = \frac{1}{20\,\text{k} \times 0.01\,\mu\text{F}} = 5\ \text{k}$$

Hence the resistance value required is 5 kΩ.

15.5.2 Second Order Low Pass Butterworth Filter

A stop band response having a 40 db/decade roll off is obtained with a second order filter.

A second order filter can be obtained by coupling two active filters having a – 20 db/decade roll off to produce a roll off of – 40 db/decade, but this would not be economical because it requires two opamps.

A second order filter can be obtained by the use of a single opamp first order low pass filter by simply using an additional RC network, as shown in Fig. 15.10.

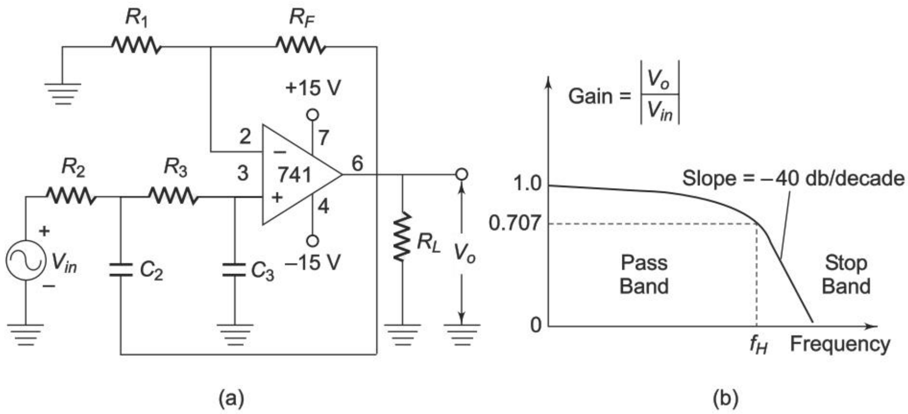


Fig. 15.10 (a) Second order low pass butterworth filter circuit
(b) Second order low pass butterworth filter frequency response

The gain of the second order filter is set by R_1 and R_F, while the high cutoff frequency f_H is determined by R_2, C_2, R_3 and C_3 as given below.

$$f_H = \frac{1}{2\pi\sqrt{R_2 R_3 C_2 C_3}} \tag{15.16}$$

The voltage gain magnitude equation for a second order low pass Butterworth response is given by

$$\left|\frac{V_o}{V_{in}}\right| = \frac{A_F}{\sqrt{1+\left(f/f_H\right)^4}} \tag{15.17}$$

where $A_F = 1 + R_F/R_1$ = pass band gain of the filter
f = frequency of the input signal, in Hz
f_H = High cutoff frequencies, in Hz

The normalised Butterworth polynomials are given in Table 15.1.

Table 15.1

Order n	*Factors of Polynomials B_n (s)*
1	$s + 1$
2	$s^2 + \sqrt{2s} + 1$
3	$(s + 1)(s^2 + s + 1)$
4	$(s^2 + 0.765 s + 1)(s^2 + 1.848 s + 1)$

where $s = j\omega$ and coefficient of $s = 2k$, where k is the damping factor.

15.5.3 Mathematical Proof for a Second Order Low Pass Filter

Consider a prototype second order low pass filter shown in Fig. 15.11.

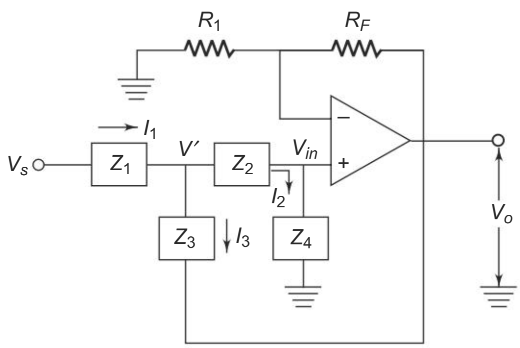


Fig. 15.11 General prototype second order filter circuit

Referring to the circuit in Fig. 15.11, we see that

$$I_1 = I_2 + I_3 \tag{15.18}$$

$$I_1 = \frac{V_s - V'}{Z_1} \tag{15.19}$$

$$I_2 = \frac{V' - V_{in}}{Z_2} = \frac{V'}{Z_2 + Z_4} \tag{15.20}$$

$$I_3 = \frac{V' - V_o}{Z_3} \tag{15.21}$$

Substituting Eqs (15.19), (15.20) and (15.21) in Eq. (15.18)

we have
$$\frac{V_s - V'}{Z_1} = \frac{V'}{Z_2 + Z_4} + \frac{V' - V_o}{Z_3} \tag{15.22}$$

Also
$$V_{in} = \left(\frac{Z_4}{Z_2 + Z_4}\right) \times V',$$

therefore
$$V' = \frac{Z_2 + Z_4}{Z_4} \times V_{in} \tag{15.23}$$

But the open loop voltage gain $A_{vo} = V_o / V_{in}$, therefore

$$V_{in} = \frac{V_o}{A_{V_o}} \tag{15.24}$$

Substituting for V_{in} in Eq. (15.23)

Therefore
$$V' = \left(\frac{Z_2 + Z_4}{Z_4}\right) \times \frac{V_o}{A_{V_o}} \tag{15.25}$$

Substituting for V' in Eq. (15.22) we have

$$\frac{V_s - \dfrac{Z_2+Z_4}{Z_4}\times\dfrac{V_o}{A_{Vo}}}{Z_1} = \frac{\dfrac{Z_2+Z_4}{Z_4}\times\dfrac{V_o}{A_{Vo}}}{Z_2+Z_4} + \frac{\dfrac{Z_2+Z_4}{Z_4}\times\dfrac{V_o}{A_{Vo}} - V_o}{Z_3}$$

$$\frac{V_s}{Z_1} - \frac{Z_2+Z_4}{Z_1 Z_4}\times\frac{V_o}{A_{Vo}} = \frac{Z_2+Z_4}{Z_4(Z_2+Z_4)}\times\frac{V_o}{A_{Vo}} + \frac{Z_2+Z_4}{Z_3 Z_4}\times\frac{V_o}{A_{Vo}} - \frac{V_o}{Z_3}$$

$$\frac{V_s}{V_1} - \frac{Z_2+Z_4}{Z_1 Z_4}\times\frac{V_o}{A_{Vo}} = \frac{1}{Z_4}\times\frac{V_o}{A_{Vo}} + \frac{Z_2+Z_4}{Z_3 Z_4}\times\frac{V_o}{A_{Vo}} - \frac{V_o}{Z_3}$$

$$\frac{V_s}{Z_1} = \frac{V_o}{A_{Vo}}\left[\frac{1}{Z_4} + \frac{Z_2+Z_4}{Z_1 Z_4} + \frac{Z_2+Z_4}{Z_3 Z_4} - \frac{A_{Vo}}{Z_3}\right]$$

$$\frac{V_s}{Z_1} = \frac{V_o}{A_{Vo}}\left[\frac{Z_1 Z_3 + Z_3(Z_2+Z_4) + Z_1(Z_2+Z_4) - A_{Vo}(Z_1 Z_4)}{Z_1 Z_3 Z_4}\right]$$

$$\frac{V_s}{Z_1} = \frac{V_o}{A_{Vo}}\left[\frac{Z_1 Z_3 + Z_2 Z_3 + Z_3 Z_4 + Z_1 Z_2 + Z_1 Z_4\,(1-A_{Vo})}{Z_1 Z_3 Z_4}\right]$$

$$\frac{V_s}{Z_1} = \frac{V_o}{A_{Vo}}\left[\frac{Z_3(Z_1+Z_2+Z_4) + Z_1 Z_2 + Z_1 Z_4(1-A_{Vo})}{Z_1 Z_3 Z_4}\right]$$

$$\therefore \qquad V_s = \frac{V_o}{A_{Vo}}\left[\frac{Z_3(Z_1+Z_2+Z_4) + Z_1 Z_2 + Z_1 Z_4(1-A_{Vo})}{Z_3 Z_4}\right]$$

$$\therefore \qquad \frac{V_o}{V_s} = \frac{Z_3 Z_4 A_{Vo}}{(Z_1+Z_2+Z_4)Z_3 + (1-A_{Vo})Z_1 Z_4 + Z_1 Z_2} \tag{15.26}$$

But this is equal to the transfer function(s).

$$Av(s) = \frac{V_o}{V_s} = \frac{Z_3 Z_4 A_{Vo}}{(Z_1+Z_2+Z_4)Z_3 + (1-A_{Vo})Z_1 Z_4 + Z_1 Z_2} \tag{15.27}$$

(This is the general equation.)

Therefore $Z_1 = R_1, Z_2 = R_2, Z_3 = 1/sC_1, Z_4 = 1/sC_2$. Now for a low pass filter, Z_1 and Z_2 are resistors and Z_3 and Z_4 are capacitors, where s is the transfer function. Therefore

$$A\,v(s) = \frac{A_{V_o}\times 1/sC_1 \times 1/sC_2}{(R_1+R_2+1/sC_2)\,1/sC_1 + (1-A_{V_o})\,R_1/sC_2 + R_1 R_2}$$

Example 15.3 *Design a second order low pass filter at a high cutoff frequency of 2 kHz.*

Solution To design the second order low pass filter the following steps are followed:

1. $f_H = 2$ kHz.
2. Let $C_2 = C_3 = 0.0033\ \mu$ F

Then $f_H = \dfrac{1}{2\pi RC}$ where $R = R_2 = R_3$ and $C = C_2 = C_3$

Therefore $$R = \frac{1}{2\pi f_H C} = \frac{1}{2 \times 3.14 \times 2\ \text{kHz} \times 0.0033\ \mu F} = 24\ \text{k}\Omega$$

Hence $R = R_2 = R_3 = 22$ k is selected.

Since $R_F/R_1 = 0.586$, therefore $R_F = 0.586 \times R_1$.

Let $R_1 = 10$ k then $R_F = 0.586 \times 10$ k $= 5.86$ k $= 5.6$ k

Hence the required components are

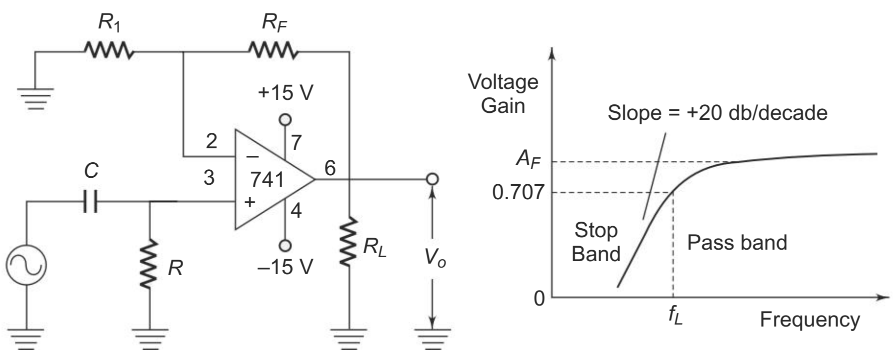
$R_2 = R_3 = 22$ k
$C_2 = C_3 = 0.0033$ μF
$R_F = 5.6$ k
$R_1 = 10$ k

15.5.4 First Order High Pass Butterworth Filter

High pass filters are often formed by interchanging frequency determining resistors and capacitors in a low pass filter.

A first order high pass filter is formed from a first order low pass type by interchanging components R and C.

Figure 15.12 shows a first order high pass Butterworth filter with a low cutoff frequency f_L, the frequency at which the magnitude of the gain is 0.707 times its pass band value. All frequencies higher than f_L are in the pass band with the highest frequency determined by the closed loop band width of the opamp.



(a) (b)

Fig. 15.12 (a) First order high pass butterworth filter circuit
(b) First order high pass butterworth filter frequency response

For the first order high pass filter, the output voltage is

$$V_o = \left(1 + \frac{R_F}{R_1}\right) \times \left(\frac{j2\pi f RC}{1 + j2\pi f RC}\right) \times V_{in}$$

or
$$\frac{V_o}{V_{in}} = A_F \times \left(\frac{j(f/f_L)}{1 + j(f/f_L)}\right) \tag{15.28}$$

where $A_F = 1 + R_F/R_1$ – pass band gain of the filter
f = frequency of the input signal
$f_L = 1/(2\,\pi\,R\,C)$ – low cutoff frequency

Hence the magnitude of the voltage gain is

$$\left|\frac{V_o}{V_{in}}\right| = \frac{A_F\,(f/f_L)}{\sqrt{1 + (f/f_L)^2}} \tag{15.29}$$

The frequency scaling procedures are the same as for the low pass filter because high pass filters are formed from low pass filters by simply interchanging Rs and Cs.

15.5.5 Second Order High Pass Butterworth Filter

As in the case of a first order filter, a second order high pass filter can be formed from a second order low pass filter by simply interchanging the frequency determining resistance and capacitor. Figure 15.13 (a) shows a second order high pass filter.

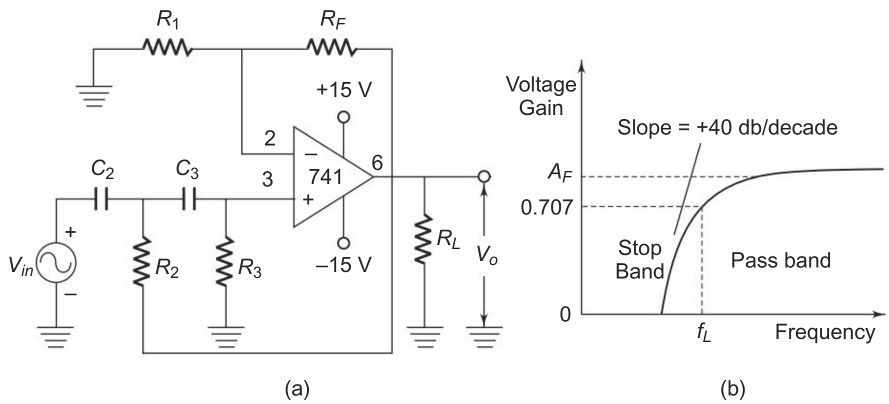


Fig. 15.13 (a) Second order high pass butterworth filter circuit
(b) Second order high pass butterworth filter frequency response

The voltage gain magnitude equation of the second order high pass filter is given by

$$\left|\frac{V_o}{V_{in}}\right| = \frac{A_F}{\sqrt{1 + (f_L/f)^4}} \tag{15.30}$$

where $A_F = 1.586$ = pass band gain for the second order Butterworth response
f = frequency of the input signal in Hertz.
f_L = low cutoff frequency

This voltage gain equation can be mathematically determined by using Eq. (15.27), and replacing Z_1 and Z_2 by capacitors and Z_3 and Z_4 by resistors.

Example 15.4 *Determine the low cutoff frequency f_L of a second order high pass Butterworth filter having the following components.*

$R_2 = R_3 = R = 47\ k\Omega$

$C_2 = C_3 = C = 0.0022\ \mu F$

Solution The low cutoff frequency for a high pass filter is given by

$$f_L = \frac{1}{2\pi RC} = \frac{1}{2 \times 3.14 \times 47\,\text{k} \times 0.0022\,\mu\text{F}} = 1.54\ \text{kHz}$$

15.5.6 Higher Order Filter

From the preceding discussion on filters we can conclude that in the stop band the gain of the filter changes at the rate of 20 db/decade for first order filters and 40 db/decade for second order filters. This means that as the order of the filter is increased, the actual stop band response of the filter approaches its ideal stop band characteristics.

Higher order filters, such as third, fourth, fifth and so on are formed simply by using first and second order filters. For example, a third order low pass filter is formed by cascading or connecting first and second order low pass filters in series. A fourth order low pass filter is composed of two cascaded second order low pass filter sections. There is no limit to the order of the filter that can be formed; as the order of the filter increases, so does its size. Also, the accuracy declines, in that the difference between the actual stop band response and the ideal stop band response increases with an increase in the order of the filter.

Figure 15.14, shows third and fourth order low pass Butterworth filters. In a third order filter the voltage gain of the first order section is one and that of the second order section is two.

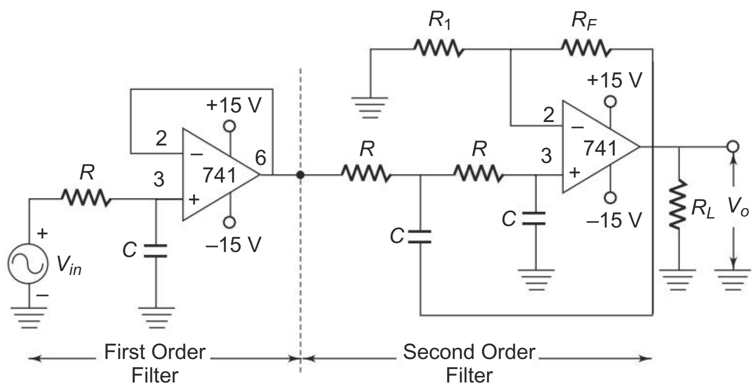


Fig. 15.14 (a) Third order low pass butterworth filter circuit

In a fourth order filter the gain of the first section is 1.152 while that of the second section is 2.235. These gain values are necessary to guarantee Butterworth response and must remain the same regardless of the filter cutoff frequency. Also, the overall filter gain is equal to the product of the individual voltage gains of the filter sections. Hence, the overall gain of a third order filter is 2.0 and that of the fourth order is $(1.152) \times (2.235) = 2.57$.

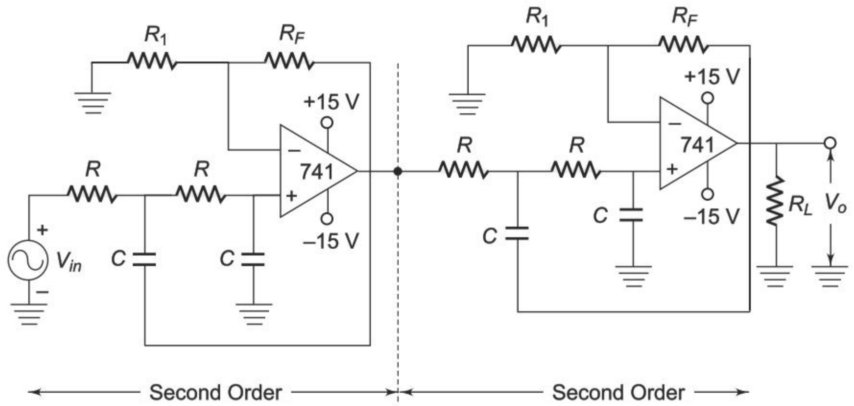


Fig. 15.14 (b) Fourth order low pass filter (Butterworth)

Since the frequency determining resistors and capacitors are equal, the high cutoff frequencies of the third and fourth order low pass filters in Figs 15.14 (a) and (b) must also be equal, and are given by

$$f_H = \frac{1}{2\pi RC}$$

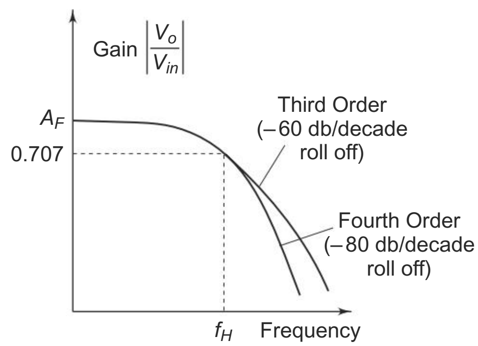


Fig. 15.14 (c) Third order and fourth order frequency response

As with first and second order filters, third and fourth order high pass filters are formed by simply interchanging the position of frequency determining resistors and capacitors. The overall gain of higher order filters is fixed because all the frequency determining resistors and capacitors are equal.

BAND PASS FILTER 15.6

A band pass filter is a circuit designed to pass signals only in a certain band of frequencies while rejecting all signals outside this band. There are basically two types of band pass filters,

(i) Wide band pass

(ii) Narrow band pass

A band pass filter is defined as a wide band pass if its figure of merit or quality factor $Q < 10$. While there is no firm dividing line between the two, if $Q > 10$, the filter is a narrow band pass filter. Hence Q is a measure of selectivity meaning the higher the value of Q, the more selective is the filter, or the narrower is the band width. The relationship between Q, 3 db band width and the centre frequency f_c is given by

$$Q = \frac{f_c}{BW} = \frac{f_c}{f_H - f_L} \tag{15.31}$$

For the wide band pass filter, the centre frequency can be defined as

$$f_c = \sqrt{f_H f_L} \tag{15.32}$$

where f_H = high cutoff frequency

f_L = low cutoff frequency of the wide bandpass

Figure 15.15, shows the frequency response of a band pass filter. This type of filter has a maximum output voltage V_{max} at one frequency called the resonant frequency, f_c. If the frequency is varied away from resonance, the output voltage decreases. There is one frequency above f_c and one below f_c at which the voltage is $0.707 \times V_{max}$ (3 db point). These frequencies are the high and low cutoff frequencies. The band of frequencies between f_H and f_L is the band width. Therefore the band width is given by BW = $f_H - f_L$.

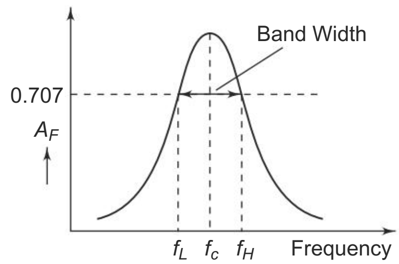


Fig. 15.15 Frequency response of a band pass filter

A narrow band filter is one that has a band width of less than 1/10th the resonant frequency (band width < 0.1 f_c). A wide band filter has a band width greater than 1/10th the resonant frequency (band width > 0.1 f_c).

The ratio of resonant frequency to band width is known as the quality factor Q.

15.6.1 Wide Band Pass Filter

Band pass can be realised by a number of possible circuits. A wide band pass filter can be formed by simply cascading high pass and low pass section and is generally the choice for simple to design.

To obtain a ± 20 db/decade band pass filter, a first order high pass filter and a first low pass sections are cascaded, for a ± 40 db/decade band pass filter, second order high pass filter and second order low pass filter are cascaded and so on for higher orders.

In other words, the order of the band pass filter depends upon the order of the High pass and Low pass sections.

Fig. 15.16(a), shows a circuit of a ± 20 db/decade wide band pass filter, which is composed of a first order high pass and a first order low pass filter. Figure 15.16(b) shows the frequency response.

Figure 15.16(c) shows a circuit of a ± 40 db/decade wide band pass filter.

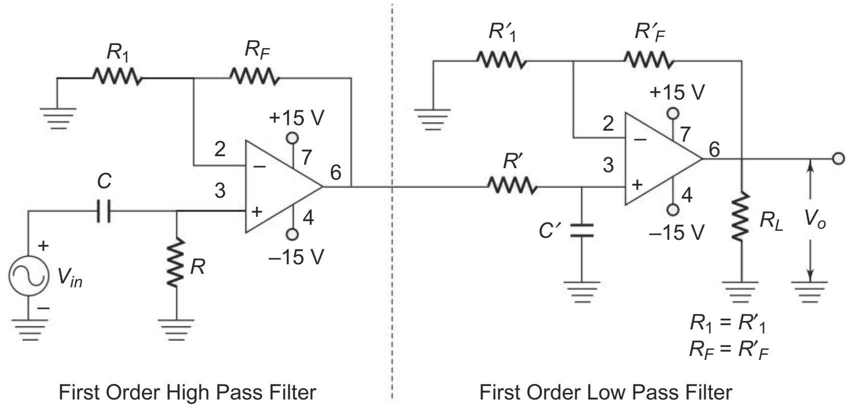


Fig. 15.16 (a) ±20 db/decade band pass filter

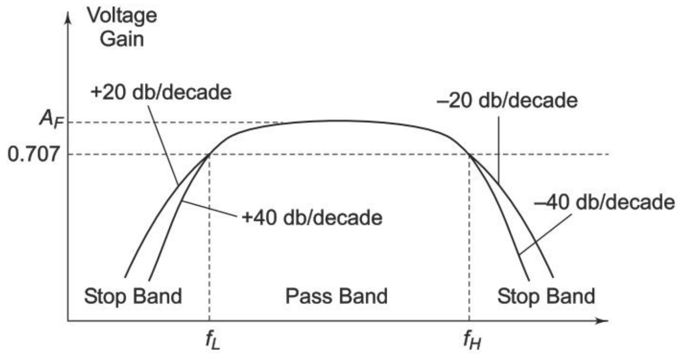


Fig. 15.16 (b) Frequency response of a band pass filter

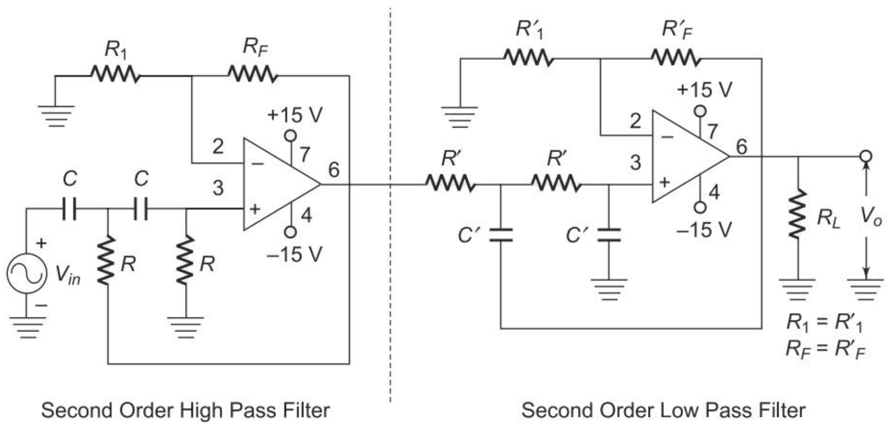


Fig. 15.16 (c) Second order band pass filter

Example 15.5

(i) *Design a wide band pass filter with f_L = 100 Hz, f_H = 1 kHz and a pass band gain of 4.*

(ii) *Calculate the value of Q for this filter.*

Solution

(i) For a low pass filter f_H = 1 kHz and $f_H = \dfrac{1}{2\pi RC}$

Let $C = 0.01\ \mu\text{F}$

then $$R = \frac{1}{2\pi f_H C} = \frac{1}{2 \times 3.14 \times 1\,\text{kHz} \times 0.01\,\mu\text{F}} = 15.9\ \text{k}\Omega$$

As in the case of high pass filters, f_L = 100 Hz and $f_L = \dfrac{1}{2\pi RC}$

Let $C = 0.01\ \mu\text{F}$

$$R = \frac{1}{2\pi f_L C} = \frac{1}{2 \times 3.14 \times 100 \times 0.01\,\mu\text{F}} = 159\ \text{k}\Omega$$

Since the band pass gain is 4, the gain of the high pass and low pass filter sections are set each equal to 2. Therefore $R_1 = R_F$ = 10 k for both sections.

(ii) $$f_c = \sqrt{f_L f_H} = \sqrt{100 \times 1\text{kHz}} = 316.2$$

Therefore $$Q = \frac{f_c}{f_H - f_L} = \frac{316.2}{1\,\text{kHz} - 100} = \frac{316.2}{900} = 0.35$$

Hence Q is less than 10, as expected for a wide band pass filter.

15.6.2 Narrow Band Pass Filter

The narrow band pass filter using multiple feedback is shown in Fig. 15.17(a). As shown in this circuit, the filter uses only one opamp. This filter is unique in the following respects.

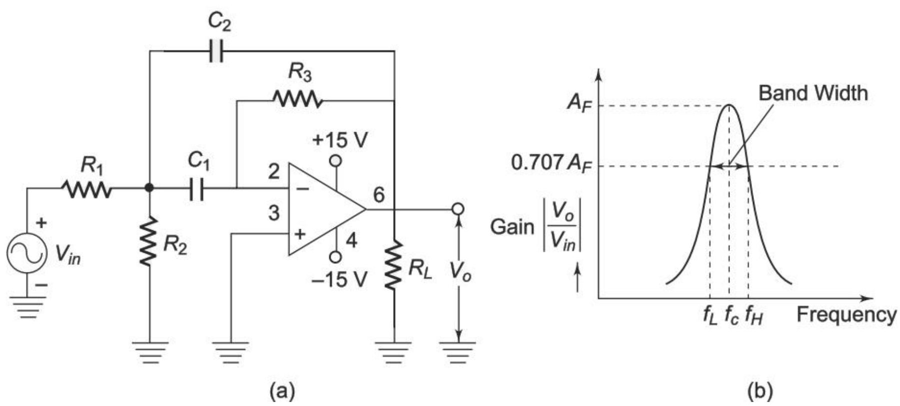


Fig. 15.17 (a) Narrow band pass filter using multiple feedback
(b) Frequency response of a narrow band pass filter

1. It has two feedback paths, which is why it is called a multiple feedback filter.
2. The opamp is used in the inverting mode.

Generally a narrow band pass filter is designed for specific values of centre frequency f_c and Q, or f_c and band width.

Figure 15.17(b) shows the frequency response of a narrow band pass filter.

15.6.3 Mathematical Derivation for Narrow Band Pass Filter

If we wish to realise a resonant band pass filter having the narrow band characteristic of a tuned circuit, we have to use a multiple feedback network in conjunction with an opamp as shown in Fig. 15.17 (a).

The transfer function of a tuned circuit is given by

$$Av(s) = \frac{A_{Vo} s(\omega_o/Q)}{s^2 + (\omega_o/Q)s + \omega_o^2} \tag{15.33}$$

Now let us consider a CR network comprising C_1 C_2 and R_3 as shown in Fig. 15.18.

Referring to figure, the z-parameters or impedance parameters are

$$V_1 = Z_{11} I_1 + Z_{12} I_2 \tag{15.34}$$

$$V_2 = Z_{21} I_1 + Z_{22} I_2 \tag{15.35}$$

where Z_{11} = input impedance with output open for ac
Z_{12} = reverse transfer impedance with input open for ac
Z_{21} = forward transfer impedance with output open for ac
Z_{22} = output impedance with input open for ac

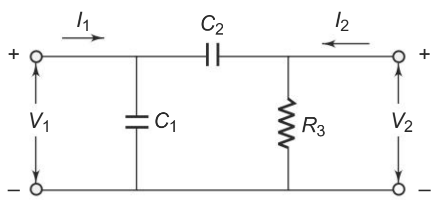


Fig. 15.18 Equivalent CR network

The equivalent circuit for a narrow band pass filter is shown in Fig. 15.19.

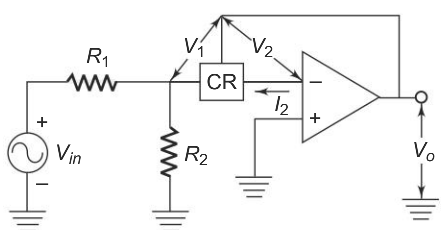


Fig. 15.19 Equivalent circuit for narrow band pass filter

Referring to the figure,

$I_2 \cong 0$, Eqs (15.34) and (15.35) reduces to

$V_1 = Z_{11} I_1$ and $V_2 = Z_{21} I_1$

Also due to virtual ground

$V_o + V_2 = 0$ therefore $V_o = -V_2$

But $V_2 = Z_{21} I_1$, therefore $V_o = -Z_{21} I_1$ (15.36)

The current flowing through R_2 is given by

$$I_2 = \frac{V_o + V_1}{R_2}$$

Now consider the mesh $V_{in}\, R_1\, R_2$

$$V_{in} = R_1\left(I_1 + \frac{V_o + V_1}{R_2}\right) + V_o + V_1$$

But $V_1 = Z_{11} I_1$ and $I_1 = -V_o/Z_{21}$.

$$\therefore \quad V_{in} = -V_o\left[\frac{1}{Z_{21}}\left(R_1 + \frac{R_1}{R_2}Z_{11} + Z_{11}\right) + V_o\left(\frac{R_1}{R_2} + 1\right)\right]$$

$$V_{in} = -V_o\left[\frac{1}{Z_{21}}\left(R_1 + \frac{R_1}{R_2}Z_{11} + Z_{11}\right) - \left(\frac{R_1}{R_2} + 1\right)\right]$$

$$V_{in} = -V_o\left[\frac{R_1 R_2 + R_1 Z_{11} + R_2 Z_{11} - R_1 Z_{21} - R_2 Z_{21}}{Z_{21} R_2}\right]$$

Simplifying it we have

$$V_{in} = -\frac{V_o}{Z_{21}}\left[(Z_{11} - Z_{21}) \times \left(\frac{R_1}{R_2} + 1\right) + R_1\right]$$

Therefore

$$\frac{V_o}{V_{in}} = \frac{-Z_{21}}{(Z_{11} - Z_{21}) \times \left(\frac{R_1}{R_2} + 1\right) + R_1}$$

$$\frac{V_o}{V_{in}} = \frac{-Z_{21} R_2}{(Z_{11} - Z_{21}) \times (R_1 + R_2) + R_1 R_2} \tag{15.37}$$

We now calculate the values of Z_{11} and Z_{21}. Referring to the figure for a CR network we have

$$Z_{11} = \frac{1/sC_1\,(R_3 + 1/sC_2)}{1/sC_1 + 1/sC_2 + R_2}$$

which reduces to

$$Z_{11} = \frac{R_3 C_2 + 1/s}{(C_1 + C_2) + sR_3 C_1 C_2} \tag{15.38}$$

Similarly

$$Z_{21} = \frac{R_3\, C_2}{(C_1 + C_2) + s\, R_3\, C_1\, C_2} \tag{15.39}$$

Therefore

$$Z_{11} - Z_{21} = \frac{1/s}{(C_1 + C_2) + s\, R_3\, C_1\, C_2} \tag{15.40}$$

Substituting Eqs (15.38), (15.39) and (15.47) in Eq. (15.37), we get

$$\frac{V_o}{V_{in}} = \frac{-\dfrac{R_2\, R_3\, C_2}{(C_1 + C_2) + s\, R_3\, C_1\, C_2}}{\dfrac{1}{s}\left[\dfrac{1}{(C_1 + C_2) + s\, R_3\, C_1\, C_2}\right](R_1 + R_2) + R_1\, R_2}$$

Therefore, gain

$$Av\,(s) = \frac{V_o}{V_{in}} = \frac{-R_2\, R_3\, C_2}{(1/s)(R_1 + R_2) + R_1\, R_2\left[(C_1 + C_2) + s\, R_3\, C_1\, C_2\right]}$$

Multiplying the numerator and the denominator by $s/R_1\, R_2\, R_3\, C_1\, C_2$ we get

$$Av\,(s) = \frac{-s/R_1\, C_1}{s^2 + \dfrac{s}{R_3}\left(\dfrac{C_1 + C_2}{C_1\, C_2}\right) + \dfrac{1}{R_3\, C_1\, C_2}\left(\dfrac{R_1 + R_2}{R_1\, R_2}\right)}$$

Let $\quad C_1 = C_2 = C$

$$\therefore \quad Av\,(s) = \frac{-s/R_1\, C_1}{s^2 + \dfrac{2s}{R_3\, C} + \dfrac{1}{R_3\, C^2}\left(\dfrac{R_1 + R_2}{R_1\, R_2}\right)} \tag{15.41}$$

Now comparing Eq. (15.41) with (15.33) we see that

$$\frac{A_{Vo}\, \omega_o}{Q} = -\frac{1}{R_1\, C_1} \tag{15.42}$$

$$\frac{\omega_o}{Q} = \frac{2}{R_3\, C} \tag{15.43}$$

$$\omega_o^2 = \frac{1}{R_3\, C^2}\left(\frac{R_1 + R_2}{R_1\, R_2}\right) \tag{15.44}$$

But $\omega_o = 2\pi f_c$ where f_c = cutoff frequency.
Therefore Eqs (15.42), (15.43) and (15.44) now become

$$R_1 = \frac{Q}{2\pi\, f_c A_{V_o}\, C} \tag{15.45}$$

$$R_3 = \frac{Q}{\pi\, f_c\, C} \tag{15.46}$$

and with slight manipulations with Eq. (15.44) we get

$$R_2 = \frac{Q}{2\pi f_c C\left(2Q^2 - A_{V_o}\right)} \tag{15.47}$$

But the gain A_{Vo} at f_c is given by

$$A_{Vo} = \frac{R_3}{2R_1} \tag{15.48}$$

which is derived from a slight manipulation of Eqs (15.45) and (15.46)

Another condition that the gain must satisfy is

$$A_{Vo} < 2Q^2 \tag{15.49}$$

Else we get $R_2 = 0$ which is not possible.

Example 15.6

(i) Design a narrow band pass filter as shown in Fig. 15.17, with a centre frequency f_c = 1 kHz, $Q = 5$, $A_{Vo} = 8$

(ii) Change the centre frequency to 1.5 kHz., keeping A_{Vo} and bandwidth constant.

Solution

(i) Let $C_1 = C_2 = C_3 = 0.01$ μF

But

$$R_1 = \frac{Q}{2\pi f_c C A_{V_o}} = \frac{5}{2\times 3.14\times 1\text{kHz}\times 0.01\mu\text{F}\times 8}$$

$$= 9.9\text{ k} = 10\text{ k}$$

$$R_2 = \frac{Q}{2\pi f_c C\left(2Q^2 - A_{V_o}\right)}$$

$$= \frac{5}{2\times 3.14\times 1\text{kHz}\times 0.01\mu\text{F}\left[2(25)-8\right]}$$

$$= \frac{5}{2\times 3.14\times 1\text{kHz}\times 0.01\mu\text{F}\times 42}$$

$$= 1.89\text{ k2} = 2\text{ k}\Omega\text{ (approx.)}$$

$$R_3 = \frac{Q}{\pi f_c C} = \frac{5}{3.14\times 1\text{kHz}\times 0.01\,\mu\text{F}} = 159\text{ k}\Omega$$

$$= 150\text{ k}\Omega\text{ (approximately)}$$

(ii) To change the centre frequency f_c to a new centre frequency f'_c without changing the gain or band width, simply change R_2 to R'_2, so that

$$R'_2 = R_2\left(\frac{f_c}{f'_c}\right)^2$$

To change f_c = 1 kHz to f'_c = 1.5 kHz.

$$\therefore \qquad R'_k = 2\,\text{k}\left(\frac{1\,\text{kHz}}{1.5\,\text{kHz}}\right)^2 = 2\ \text{k} \times 0.444 = 888.88\ (\text{approx. } 820\ \Omega)$$

BAND REJECT (STOP) FILTER 15.7

In this filter, frequencies are attenuated in the stop band and passed outside it, as shown in Fig. 15.20(b).

As with band pass filters, band reject filters can also be classified as (i) wide and (ii) narrow band.

The narrow band reject filter is also called the notch filter. Because of its higher Q which is greater than 10, the bandwidth of the narrow band reject filter is much smaller than that of the wide band reject filter. The band reject filter is also called a band stop or band elimination filter because it eliminates a certain band of frequencies.

15.7.1 Wide Band Reject Filter

Figure 15.20 (a) shows wide band reject filter using a low pass filter, a high pass filter and a summing amplifier. For a proper band reject response, the low cutoff frequency f_L of the high pass filter must be larger than the high cutoff frequency f_H of the low pass filter. Also, the pass band gain of both high pass and low pass sections must be equal.

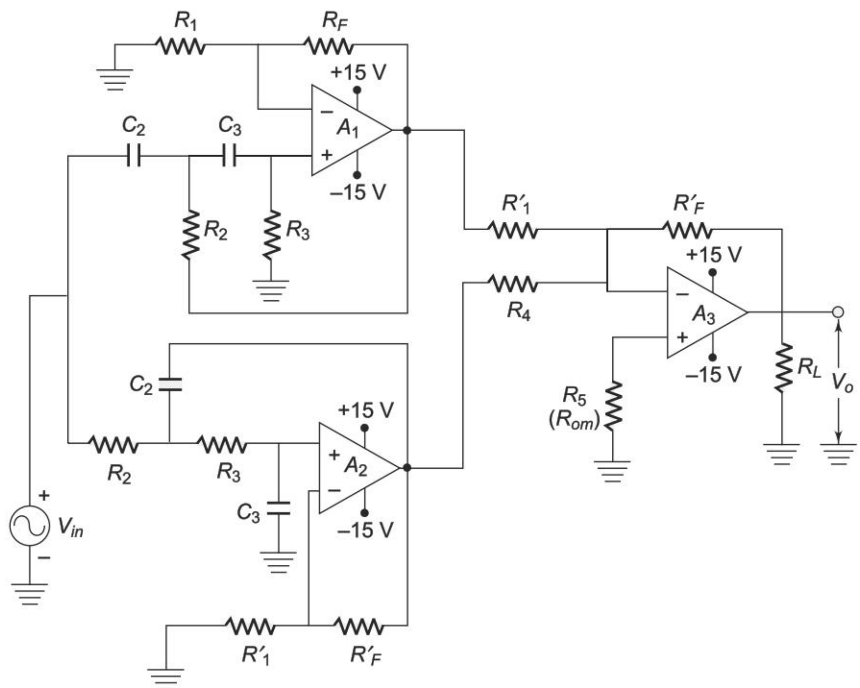


Fig. 15.20 (a) Wide band reject filter using a low pass, high pass and summing amplifier

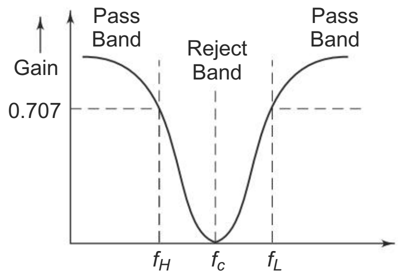


Fig. 15.20 (b) Frequency response of a wide band reject filter

Example 15.7 *Design a wide band reject filter having f_L = 100 Hz and f_H = 1 kHz.*

Solution As in the wide band pass filter (example 15.5), f_L = 100 Hz and f_H = 1 kHz.

In this example, these band frequencies are interchanged, that is f_L = 1 kHz and f_H = 100 Hz.

Therefore, the values of R and C for low pass filters are interchanged for values of R and C for high pass filters.

Hence R = 159 k, C = 0.1 μF for low pass filter, and
R = 15.9 k, C = 0.01 μF for high pass filter.

Again, a gain of 2 is used for each section and

$$R_1 = R_F = R'_1 = R'_F = 10\ \text{k}\Omega.$$

Further, the gain of the summing amplifier is set to 1,

therefore $R_2 = R_3 = R_4 = 10\ \text{k}\Omega$

The value of $R_{om} = R_2 \parallel R_3 \parallel R_4 = 10\ \text{k} \parallel 10\ \text{k} \parallel 10\ \text{k} = 3.3\ \text{k}\Omega$.

15.7.2 Narrow Band Reject Filter

The narrow band reject filter, often called the notch filter, is commonly used for the attenuation of a single frequency. For example, it may be necessary to attenuate 60 Hz or 400 Hz noise or hum signals in a circuit. The most commonly used notch filter is the Twin T network, shown in Fig. 15.21(a), which is a passive filter composed of two T shaped networks.

One T network is made up of two resistors and a capacitor, while the other is made of two capacitors and a resistor. The frequency at which maximum attenuation occurs is called the notch-out frequency, given by

$$f_N = \frac{1}{2\pi RC} \tag{15.50}$$

One disadvantage of the passive twin T network is that it has a relatively low figure of merit, Q. As discussed earlier, the higher the value of Q, the more selective is the filter. Therefore, to increase the Q of the twin T network significantly, it should be used with a voltage follower, as shown in Fig. 15.21(b). Figure 15.21(c) shows the frequency response of a notch filter.

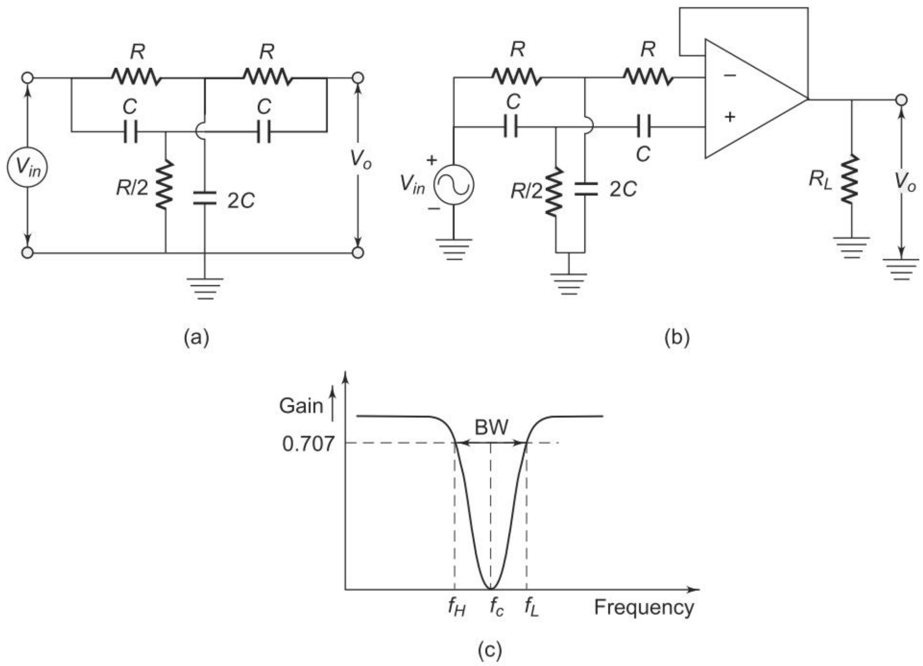


Fig. 15.21 (a) Twin-T notch filter (b) Active notch filter (c) Frequency response of a notch filter

The Notch filters are used in communications, biomedical instruments, etc. where the elimination of certain frequencies is necessary.

Example 15.8 *Design a 50 Hz active notch filter.*

Solution Let $C = 0.047\ \mu F$, and using Eq. (15.50).

We get

$$R = \frac{1}{2\pi f_N C} = \frac{1}{2 \times 3.14 \times 50 \times 0.047\ \mu F} = 67.73\ k = 68\ k\Omega$$

For $R/2$, two 68 kΩ resistors connected in parallel are used and for the $2C$ component, two 0.047 μF capacitors connected in parallel are used.

ALL PASS FILTER 15.8

The all pass filter is one that passes all frequency components of the input signal without attenuation. Any ordinary wire can be used to perform this characteristic but the most important factor in an all pass filter is that it provides predictable phase shifts for different frequencies of the input signal.

These filters are widely used in communications. For example, when signals are transmitted over transmission lines, such as telephone wires, from one point to another, they undergo a change in phase. All pass filters are used to compensate for these phase changes. They are also called delay equalizers or phase correctors.

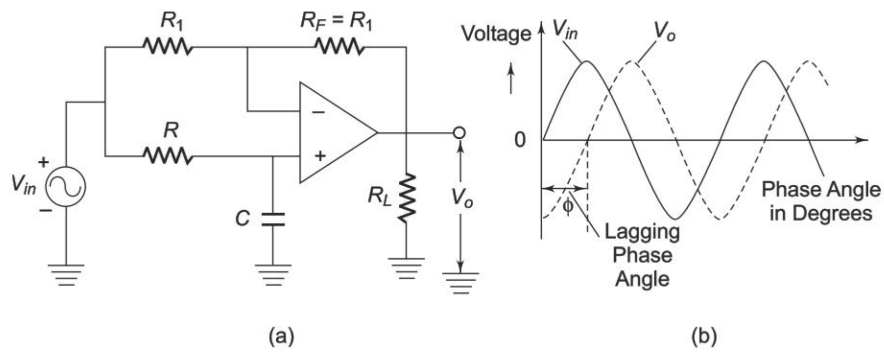


Fig. 15.22 (a) All pass filter circuit (b) Input-output waveform relation of an all pass filter (lagging)

Figure 15.22 shows an all pass filter with the output lagging the input. The output voltage V_o of the filter can be obtained by using the superposition theorem, as follows. From Fig. 15.22(a), we see that

$$V_o = -V_{in} + \frac{-jX_c}{R - jX_c} \times V_{in} \times 2 \tag{15.51}$$

where $\quad j = \sqrt{-1}$ and $X_c = \dfrac{1}{2\pi fC}$

$$V_o = -V_{in} + \frac{X_c \times V_{in} \times 2}{j(R - jX_c)}$$

$$V_o = -V_{in} + \frac{X_c \times V_{in} \times 2}{jR + X_c}$$

$$V_o = -V_{in} + \frac{\dfrac{1}{2\pi fC} \times V_{in} \times 2}{jR + \dfrac{1}{2\pi fC}}$$

$$V_o = -V_{in} + \frac{2V_{in}}{j2\pi fCR + 1}$$

$$V_o = -V_{in}\left(-1 + \frac{2}{j2\pi fCR + 1}\right)$$

Therefore

$$V_o = V_{in}\left(\frac{1 - j2\pi fCR}{1 + j2\pi fCR}\right) \tag{15.52}$$

The above equation indicates that the amplitude of V_o/V_{in} is unity, that is $|V_o| = |V_{in}|$ throughout the useful frequency range and the phase shift between V_o and V_{in} is a function of input frequency f. The phase angle ϕ is given by

$$\phi = 2 \tan^{-1} (2 \pi f C R) \tag{15.53}$$

where ϕ = phase in degrees
f = frequency in Hz
R = resistance in Ω
C = capacitance in F

Referring to Fig. 15.22(a), if the positions of R and C are interchanged, the phase shift between input and output becomes positive. That is, output V_o leads input V_{in}.

Example 15.9 *For the all pass filter of Fig. 15.22(a), determine the phase angle ϕ if the input frequency is 2.5 kHz.*

Solution Let $C = 0.01$ µF and $R = 15$ k.

From Eq. (15.60).

$\phi = 2 \tan^{-1} (2 \pi f C R)$

$\phi = 2 \tan^{-1} (2 \times 3.14 \times 2.5 \text{ k} \times 15 \text{ k} \times 0.01 \text{ µF}) = -134°$

This means that the output voltage V_o has the same frequency and amplitude as the input voltage but lags it by –134°, as shown in Fig. 15.22(b).

UNIVERSAL ACTIVE FILTERS 15.9

With the advance of integrated circuit technology, integrated circuits with improved capabilities are appearing in ever increasing numbers. Innovative design methods and fabrication procedures have not only helped to produce a large variety of new integrated circuits but also improved the old ones.

Often the use of specialised ICs produces a simpler and more accurate circuit, such as Datel's universal filter FLT-U2 which has simultaneous low pass, high pass and band pass output responses. Notch and all pass functions are also available by combining these output responses in the uncommitted opamp. Because of its versatility, this filter is called a universal filter. The universal filter is sometimes also called a state variable filter.

The various filter networks discussed in this chapter are used in many circuits, but in critical applications specially designed filter ICs are preferred. Besides being more accurate, specially designed IC filters are simpler, easier to use and more flexible. Detal's FLT-U2 is a typical example of such a specialised IC filter.

Datel's FLT-U2 is a universal filter which uses the state variable active filter principle to implement second order low pass, high pass and band pass outputs functions. These output functions are simultaneously available at the outputs of the three committed opamps (Pin nos 3, 13 and 5), as shown in the FLT-U2 block diagram of Fig. 15.23(a).

A fourth uncommitted opamp can be used as a gain stage buffer amplifier, or to raise the order of the low pass, high pass or band pass functions. The uncommitted opamp can also be used to realise the notch and all pass functions.

Frequency tuning is accomplished by using two external resistors and Q tuning by using a third resistor.

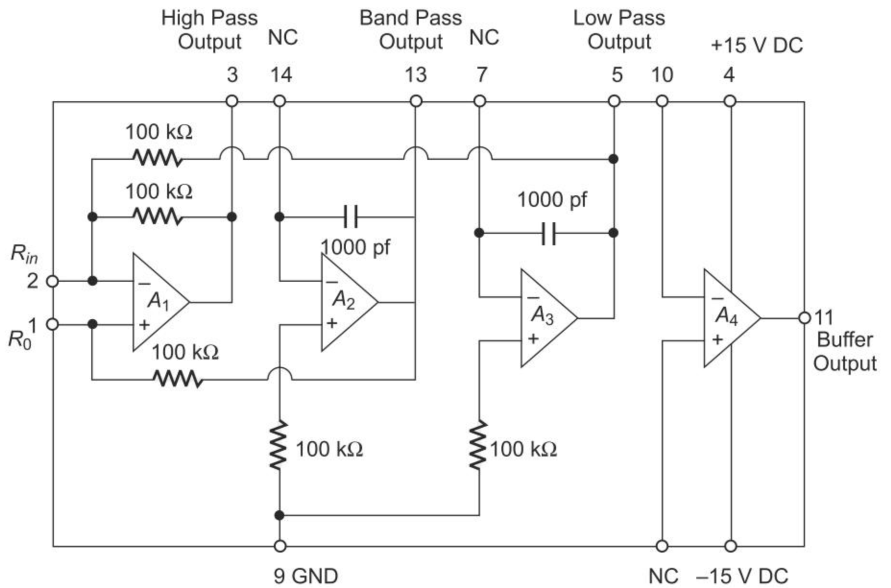


Fig.15.23 (a) FLT-U2 block diagram

In addition, any of the filter types, such as Butterworth, Chebyschev or Bessel, may be designed by the proper selection of external components by using the FLT-U2.

Figure 15.23(b) shows an FLT-U2 as a second order low pass, high pass and band pass filter.

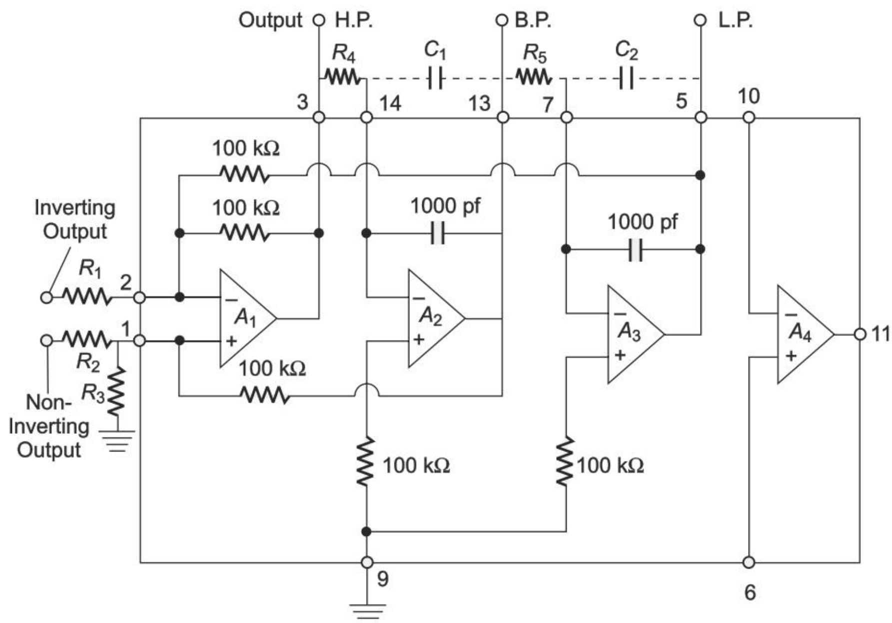


Fig. 15.23 (b) FLT-U2 as a second order low pass, high pass and band pass filter

DESIGNING PROCEDURES FOR FLT-U2 15.10

The following procedure should be carried out for designing an FLT-U2.

1. For a desired function-low pass, high pass or band pass use Table 15.2 to select an appropriate filter configuration for inverted or non-inverted output.

Table 15.2 Filter Configuration

Configuration	*Low Pass*	*High Pass*	*Band Pass*
Inverting input	Inverted	Inverted	Non-inverted
Non-inverting input	Non-inverted	Non-inverted	Inverted

2. For an inverting configuration, follow this step; otherwise go to step 3. Using the desired value of Q, calculate R_1 and R_3 from Table 15.3. Note, since it is an inverting configuration R_2 is open.

Table 15.3 Inverting Configuration

Type	R_1	R_3
Low Pass	100 kΩ	$\dfrac{100\text{k}}{(3.80)(Q)-1}$
High Pass	10 kΩ	$\dfrac{100\text{ k}}{(6.64)(Q)-1}$
Band Pass	Q (31.6 kΩ)	$\dfrac{100\text{k}}{(3.48)(Q)}$

3. For an non-inverting configuration follow this step: otherwise go to step 4. Calculate R_2 and R_3 value from Table 15.4, using the desired value of Q. Note, since it is an non-inverting configuration R_1 is open.

Table 15.4 Non-inverting Configurations

Types	R_2	R_3
Low Pass	$\dfrac{316\text{k}}{Q}$	$\dfrac{100\text{k}}{3.16Q-1}$
High Pass	$\dfrac{31.6\text{k}}{Q}$	$\dfrac{100\text{k}}{0.316Q-1}$
Band Pass	100 k	$\dfrac{100\text{k}}{3.48-1}$

4. Using the desired value of resonant frequency f_1, which is the centre frequency for the band pass and cutoff frequency for the low pass or high pass, calculate the R_4 and R_5 values from the equation

$$R_4 = R_5 = \frac{(5.03) \times 10^7}{f_1}\ \Omega \tag{15.54}$$

For a band pass filter the centre frequency can be varied without affecting the bandwidth by varying R_5, with R_4 fixed.

5. If centre or cutoff frequency, $f_1 < 50$ Hz, the 1000 pf internal capacitors should be shunted with external capacitors equal to the values of internal capacitors across pins 5 – 7 and 13 – 14. The values of R_4 and R_5 can then be computed from the equation

$$R_4 = R_5 = \frac{(5.03) \times 10^{10}}{f_1 C}\ \Omega \tag{15.55}$$

where C is the total capacitance in pf, that is, the sum of external capacitors across pins 5 and 7 or across 13 and 14 in pf and the internal 1000 pf capacitor.

6. The procedure described above is based on unity gain output for the desired function.
However, if additional gain is required, the fourth uncommitted opamp should be used as an inverting or non-inverting gain stage following the selected output, as shown in Fig. 15.24.

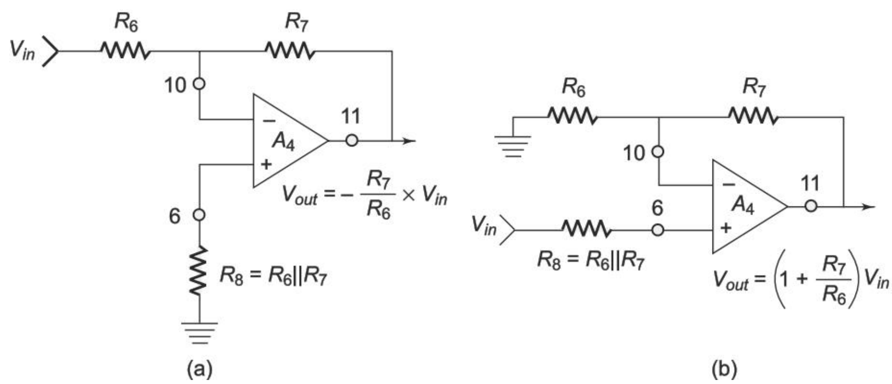


Fig. 15.24 (a) Uncommitted Opamp gain with inverting configuration
(b) Uncommitted Opamp gain with non-inverting configuration

Similarly the order of the filter function can be raised by adding a capacitor to the gain stage, as shown in Fig. 15.25.

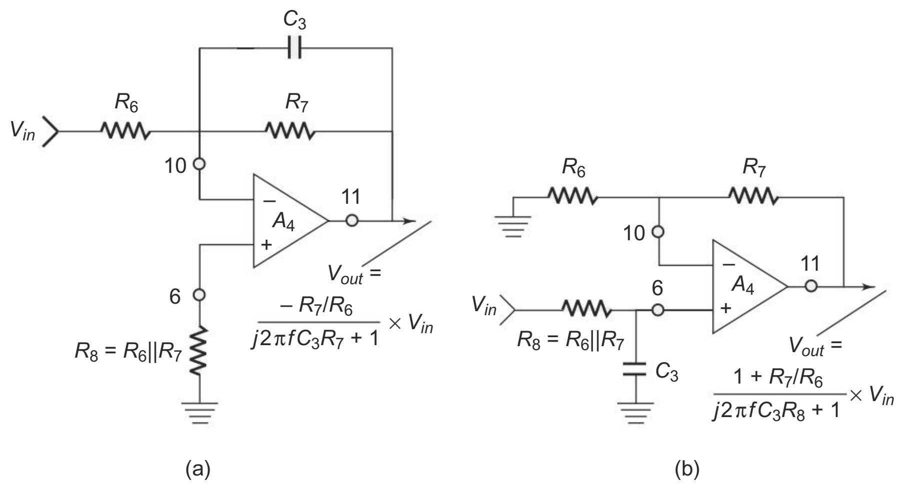


Fig. 15.25 (a) Using the uncommitted opamp to raise the order of low pass function in the inverting configuration (b) Using the uncommitted opamp to raise the order of low pass function in the non-inverting configuration

To construct a notch filter, the simplest method is to use the FLT-U2 as an inverting band pass filter and then to sum the output of the band pass filter with the input signal by means of the uncommitted opamp.

Higher order filters can be made by using cascaded FLT-U2 stages. The FLT-U2 universal filter is used in audio tone signalling, data acquisition and feedback control systems.

Example 15.10 *The FLT-U2 is to be used as a second order inverting Butterworth low pass filter with a dc gain of 6, cutoff frequency of 1.5 kHz and Q = 10. Determine the values of the external components.*

Solution The values of the external components are determined by the following design steps.

1. According to Table 15.1, the inverting configurations would normally be used to give an inverting low pass output. However, to obtain a gain of 6, an inverting uncommitted opamp has to be used, hence the non-inverting filter configuration must be used.
2. From Table 15.3, using $Q = 10$

$$R_2 = \frac{316\,\text{k}}{10} = 31.6\ \text{k}\Omega$$

$$R_3 = \frac{100\,\text{k}}{(3.16)(10) - 1} = 3.27\ \text{k} \cong 3.3\ \text{k}$$

$$R_1 = \text{open}$$

3. Substituting f_1 = 1.5 kHz in the following equations,

$$R_4 = R_5 = \frac{(5.03)\times 10^7}{f_1} = \frac{(5.03)\times 10^7}{1.5\,\text{kHz}} = 33.53\ \text{k}\Omega$$

Use $R_4 = R_5 = 33$ kΩ.

4. The last step is to use the uncommitted opamp as an inverting amplifier with a gain of 6.

Let R_6 = 1.8 k, then $R_7 = 6 \times 1.8$ k = 10.8 k (use 10 k)

$R_8 = R_6 \parallel R_7$ = 1.8 k || 10.8 k = 1.542 k

Example 15.11 *Using FLT-U2, design a notch filter with a 4 kHz notch-out frequency and Q = 8.*

Solution The FLT-U2 can be used as a notch filter by summing the inverted output of the band pass filter designed with the input signal by means of the uncommitted opamp.

For a band pass (according to Table 15.3)

R_2 = 100 k

$$R_3 = \frac{100\ \text{k}}{(3.48)\,8 - 1} = 3.72\ \text{k (use 3.9 k)}$$

R_1 = open

Since f_1 = 4 kHz,

$$R_4 = R_5 = \frac{(5.03)\times 10^7}{4\,\text{kHz}} = 12.5\ \text{k} = 12\ \text{k}\Omega$$

Let $R_6 = R_7 = R_8$ = 10 k

Then $R_9 = R_6 \parallel R_7 \parallel R_8$ = 3.3 k

Figure 15.26 shows the complete circuit diagram of the notch filter using FLT-U2.

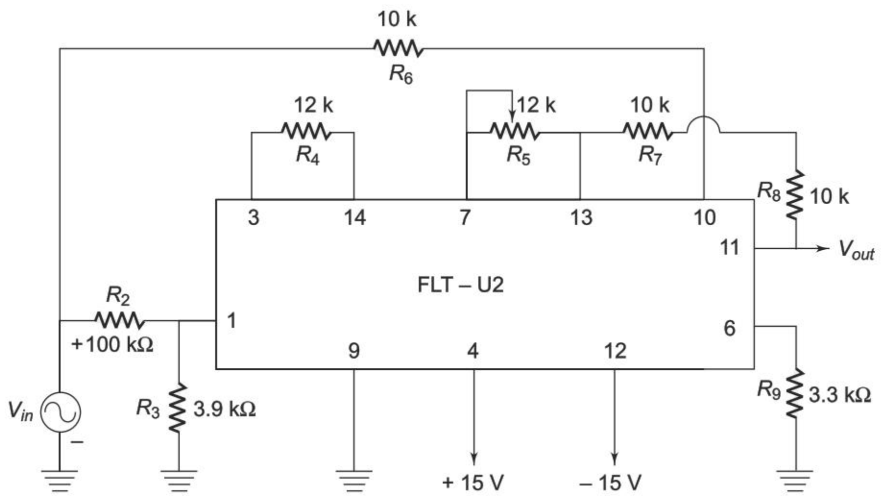


Fig. 15.26 Notch filter of Example 15.11

TYPES OF ACTIVE FILTERS 15.11

There are basically four useful types of active filters.

1. Butterworth
2. Chebyschev
3. Bessel
4. Elliptic

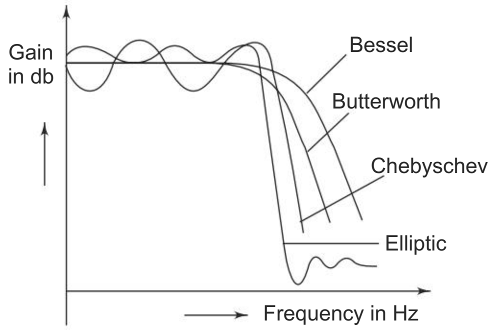


Fig. 15.27 Frequency response of various filters

15.11.1 Butterworth Filter

The Butterworth filter has an essentially flat amplitude versus frequency response up to the cutoff frequency. The sharpness of the cutoff can be seen in Fig. 15.27, where it is compared with the Chebyschev and the Bessel filter.

It is to be noted that all three filters reach a roll-off slope of – 40 db/decade at frequencies much greater than cut off. Although Butterworth filters achieve the sharpest attenuation, their phase shift as a function of frequency is non-linear.

Butterworth filters are also known as maximally flat type filters. This class of filters approximates the ideal filter well in the pass band. It has a monotonic decrease in gain with frequency in the cutoff region and a maximally flat response below cut-off, as shown in Fig. 15.28.

The Butterworth filter has characteristics somewhere between Chebyschev and Bessel filter. It has a moderate roll-off of the skirt and a slightly non-linear phase responses.

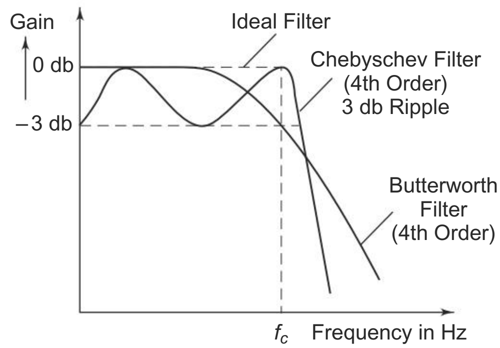


Fig. 15.28 Frequency response comparison of 4th order Butterworth, Chebyschev and ideal response

15.11.2 Chebyschev Filter

The Chebyschev filter, also called the equal ripple filter, gives a sharper cutoff than the Butterworth filter in the pass band. A comparison of these two filters is shown in Fig. 15.28.

Both Butterworth and Chebyschev filters exhibit large phase shifts near the cutoff frequency.

If very sharp roll-off is desired, a Chebyschev filter is used. However, this is obtained at the expense of a gain ripple in the lower frequency pass band.

A disadvantage of the Chebyschev filter is the appearance of gain maxima and minima below the cutoff frequency. This gain ripple, expressed in dbs, is an adjustable parameter in filter design.

The faster the roll-off, the greater the peak-to-peak ripple in the pass band. The phase response is highly non-linear in the skirt region. Such unequal delays of data frequency in the pass band causes severe pulse distortion and thus increased errors at modem demodulators. This can overcome somewhat by increasing the band width of the filter so that the phase region is extended.

15.11.3 Bessel Filter

For applications where the phase is important, the Bessel filter, which is a minimal phase shift filter is used, even though its cut off characteristics is not very sharp.

The Bessel filter provides ideal phase characteristics with an approximately linear phase response up to nearly cutoff frequency. The Bessel filter has a very linear phase response but a fairly gentle skirt slope, as illustrated in Fig. 15.29. It is well suited for pulse applications.

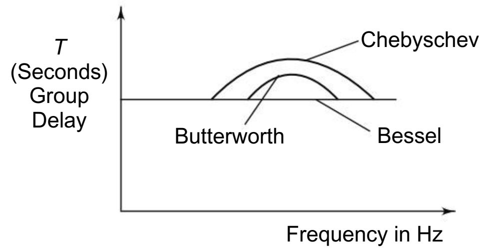


Fig. 15.29 Group delay (filter characteristics)

15.11.4 Elliptic Filter

The elliptic filter has the sharpest roll-off of all filters in the transition region, but has ripples in both the pass band and stop band regions, as shown in Fig. 15.27.

The elliptic filter can be designed to have very high attenuation for certain frequencies in the stop band, which lessens the attenuation for other frequen-cies in the stop band.

Table 15.5 Summary of Filter Response Characteristics

Bessel	*Butterworth*	*Chebyschev*	*Elliptic*
Maximally flat time delay in the dc vicinity.	Maximally flat amplitude response in the dc vicinity.	Equiripple amplitude in the pass band.	Equiripple amplitude in both pass band and stop band.
Monotonically decreasing in amplitude with frequency.		Fastest roll-off for the same order of filter when compared to Bessel or Butterworth filters.	Sharper transition band and higher attenuation in the lower portion of the stop band.
Best phase response, poorest amplitude response.	Compromise between amplitude and phase when connected to Bessel and Chebyschev filters.	Best amplitude response (more attenuation in the stop band as ripple is increased). Poorest phase response.	High frequency attenuation, lower than that of the other filters.
Attenuation Continues to Increase as Frequency Increases			Sensitive to Component Values

DIGITAL FILTERS 15.12

It has been seen that control of the band width of a circuit or signal is of paramount importance in optimising the measurement of a signal. Formerly band width control was implemented using frequency, selective active or passive networks of resistors, capacitors and inductors. However, the results obtained were somewhat less than satisfactory, because of undesirable attenuation and phase shift of signal frequencies, resulting in distortion of signal information, or less than optimal enhancement of the signal to noise (S/N) ratio. Digital filtering overcomes many of these problems.

Digital filtering simply involves band width control using software. A particularly powerful and informative route for the implementation of a digital filter is through the use of a Fourier transformation. The frequency composition can be determined from the amplitude-time waveform by carrying out a Fourier transformation. The action of an analog filter on an amplitude-time waveform can be accurately described by multiplying the frequency spectrum of the waveform by the frequency response curve for the filter.

This same process can be carried out on the computer after calculation of the Fourier transform signal waveform. However, essentially any desired frequency response curve can be set up, including many that would be simply impossible to design with hardware. Filters with no phase shift, filters with square cutoffs, high pass filters, differentiating filters and unique discrete frequency filters are all easy to implement.

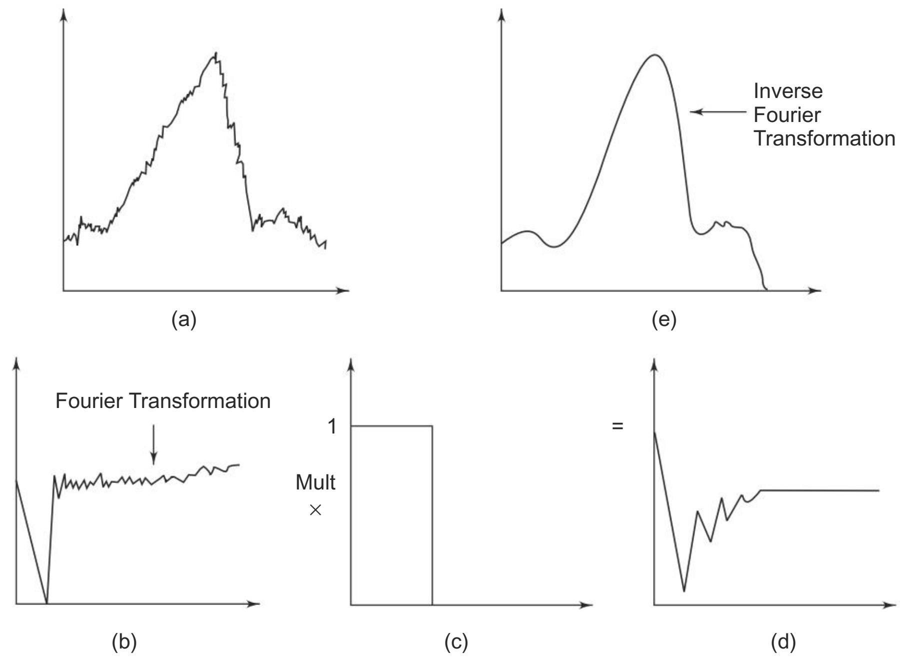


Fig. 15.30 Digital filtering (a) Noisy signal (b) Real output (c) Digital low pass filter (d) Modified real output (e) Filtered signal

A simple example is presented in Fig. 15.30. A noisy signal is shown in Fig. 15.30(a). The real output of the Fourier transform is shown in Fig. 15.30(b).

The important point is that the distribution of signal and noise information in this plot is the same as it was in the amplitude spectra with the signal information concentrated near the origin and the noise information spread throughout. A digital low pass filter is shown in Fig. 15.30(c). This filter has characteristics of no phase shift and an abrupt cutoff, which are impossible to achieve with analog filters.

Filtering is implemented by simply multiplying the real output by this filter response function, resulting in the modified real output shown in Fig. 15.30(d). The filtered signal can be regenerated by an inverse Fourier transform. The result is shown in Fig. 15.30(e). Note that the noise level has been significantly reduced. Analogous reduction of the noise level with analog techniques would have been difficult to achieve without distorting the signal information. Distortion can result from a digital filter if signal information is attenuated by the filter. Simple precautions can render such errors unlikely.

Although this example was quite simple, the basic power and flexibility of this approach to filtering should be obvious. Many unique and useful filters can be designed and implemented with this approach. It is also important to note that digital filtering as discussed in this section is exactly analogous to the moving average smoothing operations which are frequently utilized to enhance the S/N of signals.

To understand the concept of digital filters, the reader should be well versed in complex variables, Fourier series and Fourier transformation.

The mathematical proofs are beyond the scope of this book; only mathematical formulas/theorems are given.

DISCRETE FUNCTIONS 15.13

The properties of linear digital filters defined as transformations of discrete functions are described in the following.

Discrete functions are defined only for discrete values of their variables, that is, they are sequences of numbers. These numbers can be obtained as quantised samples of continuous functions (representing an analog signal or image) or they can be values of a discrete variable, such as the readings indicated by, or the output data of a digital counter. The sequences thus obtained are then processed to obtain other sequences.

The notation which is useful for one dimensional (1-D) sequences is the following

$$\{x(n)\}\ N_1 \le n \le N_2 \tag{15.56}$$

where $N_1 = -\infty$ and $N_2 = \infty$ for infinite sequences.

For two dimensional (2-D) sequences, the notation is

$$\{x(n_1, n_2)\}\ N_1 \le n_1 \le N_2,\ M_1 \le n_2 \le M_2 \tag{15.57}$$

where $N_1 = M_1 = -\infty$ and $N_2 = M_2 = \infty$ for double infinite sequences.

We can observe that with the terms, the sequence number and numerical transformation of sequences, the quantization aspect is conceptually included due to the implicit finite precision needed for any numerical representation. However, in the development of the theory of discrete linear systems we consider the more general case of sequences of numbers with infinite precision. Thus we use functions whose variable is defined on a discrete set of values, but whose amplitude can assume values in a continuous way within a specified range.

In many applications, sequences of the type (1-D) and (2-D) are obtained by taking samples of continuous or analog signals and images. Therefore, we consider, in greater precision the way an analog signal is related to its sample, leading to the sampling theorem.

THE 1-D SAMPLING THEOREM 15.14

One of the most important applications of digital filtering is the processing of sequences of samples derived from continuous or analog signals. This is made possible due to the results and implications of the sampling theorem.

This theorem can be stated as follows:

"A continuous analogue function $x(t)$ which has a limited Fourier spectrum, that is a spectrum $x(j\omega)$ such that $x(j\omega) = 0$ for $\omega > \omega_m$, is uniquely described from a knowledge of its values at uniformly spaced time instants T units apart, where $T = 2\pi/\omega_s$ and $\omega_s \geq 2\omega_m$."

THE 2-D SAMPLING THEOREM 15.15

This theorem can be stated as follows.

A function of two variables, $x(x_1, x_2)$ whose 2-D Fourier transform is equal to zero for $\omega_1 > \omega_{1m}$ and $\omega_2 > \omega_{2m}$, is uniquely determined by the values taken at uniformly spaced points in the x_1 and x_2 plane, if the spacing X_1 and X_2 satisfy the conditions $X_1 \leq \pi/\omega_{1m}$ and $X_2 \leq \pi/\omega_{2m}$.

THE 1-D Z-TRANSFORM 15.16

Given a sequence $\{x(n)\}$ with $-\infty \leq n \leq \infty$, its Z-transform is defined as

$$X(Z) = \sum_{n=-\infty}^{\infty} x(n)\, Z^{-n} \tag{15.58}$$

where Z is a complex variable.

The Z-transform can be inverted and $x(n)$ can be obtained as

$$x(n) = \frac{1}{2\pi j} \oint_C X(Z)\, Z^{n-1}\, dZ \tag{15.59}$$

where C is the counter clockwise closed contour in the region of convergence of $X(Z)$ and encircling the origin of the Z-plane.

FUNDAMENTAL PROPERTIES OF 1-D DIGITAL SYSTEMS 15.17

A digital system can be defined as an operator which transforms an input sequence $\{x(n)\}$ to an output sequence $\{y(n)\}$

$$\{y(n)\} = \mathrm{T}[\{x(n)\}] \tag{15.60}$$

A system of this type is said to be linear, indicated by the symbol L, if and only if the principle of superposition holds. This is equivalent to saying that if $\{y_1(n)\}$ and $\{y_2(n)\}$ are the outputs of the system with inputs $\{x_1(n)\}$ and $\{x_2(n)\}$, then

$$\begin{aligned} \mathrm{L}[(a\, x_1(n) + b\, x_2(n)\}] &= a\mathrm{L}[(x_1(n)\}] + b\mathrm{L}[\{x_2(n)\}] \\ &= a\,\{y_1(n)\} + b\,\{y_2(n)\} \end{aligned} \tag{15.61}$$

where a and b are arbitrary constant.

Since linearity is valid (for absolutely convergent systems) for the sum of an infinite number of terms, the linear system can be uniquely determined by its response to a unit sample sequence defined as the sequence.

$$\delta(n) = 1,\ n = 0 \text{ and } \delta(n) = 0,\ n \neq 0 \tag{15.62}$$

We can obtain another input-output relation as follows. By observing that the generic sample $x(n)$ of the input signal can be written in the form

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\,\delta(n-k) \tag{15.63}$$

The input-output relation can be written as

$$y(n) = \mathrm{L}\left[\left\{\sum_{k=-\infty}^{\infty} x(k)\delta(n-k)\right\}\right] \tag{15.64}$$

and from linearity properties

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)\,\mathrm{L}[\{\delta(n-k)\}] = \sum_{k=-\infty}^{\infty} x(k)\,h(k, n) \tag{15.65}$$

where $\{h(k, n)\}$ is the response to the unit sample sequence, which in general is a function of k and n.

A more useful relation is obtainable for a subclass of linear systems, the shift-invariant linear system. These systems are characterised by the property that if $\{y(n)\}$ is the response to $\{x(n)\}$ then $\{y(n-k)\}$ is the response to $\{x(n-k)\}$.

In this case the response of the system to the impulse $\{(n-k)\}$ is $\{h(n-k)\}$ and the input-output relation can be written as:

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)\,h(n-k) \tag{15.66}$$

This relation, which is commonly written in the symbolic form

$$\{y(n)\} = \{x(n) * h(n)\} \tag{15.67}$$

is called the convolution sum of $\{x(n)\}$ and $\{h(n)\}$. The sequence $\{h(n)\}$ is called the impulse response of the system.

In Z-transform domain, the convolution equation reduces to the form

$$Y(Z) = H(Z)\,X(Z) \tag{15.68}$$

where
$$H(Z) = \sum_{k=-\infty}^{\infty} h(k)\, Z^{-k} \tag{15.69}$$

is the Z-transform of the unit-impulse response. $H(Z)$ is called the Z-transform function of the system and characterizes completely the linear shift invariant system.

FUNDAMENTAL PROPERTY OF 2-D DIGITAL SYSTEMS 15.18

The same definitions as previously can be given in the 2-D case. A 2-D system can be characterized by an operator transforming an input 2-D sequence $\{x(n_1, n_2)\}$ to an output 2-D sequence $\{y(n_1, n_2)\}$. The system is linear if and only if the principle of superposition holds if

$$L\,[\{ax_1\,(n_1, n_2) + bx_2(n_1, n_2)\}] = aL\,\{x_1\,(n_1, n_2)\} + bL\,\{x_2\,(n_1, n_2)\} \tag{15.70}$$

If $\{h(n_1 - k, n_2 - i)\}$ is the output of the system to the unit pulse $\{\delta(n_1 - \mathrm{k}, n_2 - i)\}$, the convolution sum can be obtained by writing the output signal in the form

$$y(n_1, n_2) = \mathrm{L}\left[\sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} x(k,i)\,\delta\,(n_1 - k, n_2 - i)\right]$$

$$= \sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} x(k, i)\ \mathrm{L}[\delta(n_1 - k, n_2 - i)]$$

$$= \sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} x(k, i)\, h(n_1 - k, n_2 - i) \tag{15.71}$$

In the 2-D Z-domain the convolution operator reduces to a multiplication

$$Y(Z_1, Z_2) = H(Z_1, Z_2)\, X(Z_1, Z_2)$$

where the function

$$H(Z_1, Z_2) = \sum_{k=-\infty}^{\infty}\sum_{i=-\infty}^{\infty} h(k, i)\, Z_1^{-k}\, Z_2^{-i} \tag{15.72}$$

is called the Z-transfer function of the system. Obviously this function is the 2-D Z-transform of the impulse response $h(k, i)$.

FREQUENCY DOMAIN REPRESENTATION 15.19

A very important representation of discrete time systems can be obtained in terms of sinusoidal or complex-exponential signals.

This is possible because the linear shift invariant systems have the interesting property that the response to a sinusoidal or complex exponential input is also sinusoidal or complex-exponential, having the same frequency and amplitude, and a phase which is characteristics of the system under study.

Let us consider a discrete sequence

$$\{x(n)\} = \{e^{jn\omega}\} \tag{15.73}$$

defined for $n = -\infty$ to ∞, that is, a sampled complex exponential with radian frequency ω. The output of the system having impulse response $\{h(k)\}$ can be written as

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)\, e^{j(n-k)\omega} = e^{jn\omega} \sum_{k=-\infty}^{\infty} h(k)\, e^{-jk\omega} \tag{15.74}$$

Let $$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h(k)\, e^{-jk\omega} \tag{15.75}$$

we can write

$$y(n) = e^{jn\omega} H(e^{j\omega}) \tag{15.76}$$

The function $H(e^{j\omega})$, which is called the frequency response of the system, describes the change in phase and amplitude of the input exponential, provided the series for $H(e^{j\omega})$ converges. $H(e^{j\omega})$ is in general a complex number, and can therefore be represented in terms of its real and imaginary parts.

$$H(e^{j\omega}) = HR(e^{j\omega}) + i\, HI(e^{j\omega}) \tag{15.77}$$

or in terms of its magnitude and phase

$$H(e^{j\omega}) = |H(e^{j\omega})|\, e^{j \arg H(e)^{j\omega}} \tag{15.78}$$

In this representation the phase can be substituted by the group delay, defined as the negative of the first derivative of the phase with respect to ω.

15.19.1 1-D Linear Systems Described by the Difference Equation

A very important class of linear shift invariant systems is the one described by the following equation

$$y(n) = \sum_{k=0}^{N-1} a(k)\, x(n-k) = \sum_{k=0}^{M-1} b(k)\, y(n-k) \tag{15.79}$$

where $x(n)$ are the samples of the input sequence, $y(n)$ are the samples of the output sequence, and $a(k)$ and $b(k)$ are coefficients which define the system.

In general, of course, a digital filter is not uniquely specified by the difference Eq. (15.79), that is, to any solution of Eq. (15.79) we can add a component which satisfies the homogeneous difference equation (the difference equation with the LHS equal to 0) so that the overall sum satisfies Eq. (15.79).

Therefore, as with difference equations in continuous time, it is necessary to specify the initial conditions of the system. These initial conditions must be such that the system is linear and recursive.

For recursivity of the system, it is necessary that any output samples be computable from a knowledge of previously computed samples or from initial conditions.

If we take the Z-transform of both sides of Eq. (15.79), we obtain,

$$\sum_{n=-\infty}^{\infty} \left[\sum_{k=0}^{N-1} a(k)\, x(n-k) \right] Z^{-n} = \sum_{n=-\infty}^{\infty} \sum_{k=0}^{M-1} \left[b(k)\, y(n-k) \right] z^{-n}$$

$$\sum_{k=0}^{N-1} a(k) \sum_{n=-\infty}^{\infty} x(n-k) Z^{-n} = \sum_{k=0}^{M-1} b(k) \sum_{n=-\infty}^{\infty} y(n-k) Z^{-n}$$

$$X(Z) \sum_{k=0}^{N-1} a(k) Z^{-k} = Y(Z) \sum_{k=0}^{M-1} b(k) Z^{-k}$$

from which it is possible to obtain an input-output relation in the Z-domain in the form

$$Y(Z) = H(Z)\, X(Z) \tag{15.80}$$

where
$$H(Z) = \frac{\sum_{k=0}^{N=1} a(k) Z^{-k}}{\sum_{k=0}^{M-1} b(k) Z^{-k}} \tag{15.81}$$

Two cases can now be considered, leading to two fundamental classes of filters. If the coefficients satisfy the following condition,

$$b(0) = 1 \quad \text{and} \quad b(k) = 0, (k) \neq 0 \tag{15.82}$$

the difference Eq. (15.79) reduces to

$$y(n) = \sum_{k=0}^{N-1} a(k)\, x(n-k) \tag{15.83}$$

and hence the transfer function reduces to

$$H(Z) = \sum_{k=0}^{N-1} a(k)\, Z^{-k} \tag{15.84}$$

In this case the output samples of the filter depend only on the input samples, without any feedback of the part output samples on the current output sample.

These filters have finite responses of length N to the unit sample sequence $f(n)$. For this reason, they are known as Finite Impulse Response (FIR) filters, as opposed to Infinite Impulse Response (IIR) filters which are described by Eq. (15.79), where at least one (in addition to $b(o)$ of $b(k)$) coefficient with $k \neq 0$ is different from zero.

Transfer functions of the type (15.81) and (15.84) are uniquely determined with a constant factor by the roots of the polynomials denominator and numerator, together with the causality condition in the IIR case.

The roots of numerator polynomials are generally called zeros of the filter, while the roots of the denominator polynomials are known as the poles of the filters. Thus FIR filters have only finite zeros, while IIR filters have both zeros and poles.

15.19.2 2-D Linear Systems Described by Linear difference Equations

The 2-D difference equation which defines a linear shift invariant filter has the form

$$\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} a(n_1, n_2)\, x(m_1 - n_1, m_2 - n_2)$$
$$= \sum_{n_1=0}^{M_1-1}\sum_{n_2=0}^{M_2-1} b(n_1, n_2)\, y(m_1 - n_1, m_2 - n_2) \qquad (15.85)$$

where $\{x(n_1, n_2)\}$ is the input matrix, $\{y(n_1, n_2)\}$ is the output matrix and $\{a(n_1, n_2)\}$, $\{b(n_1, n_2)\}$ are the coefficient matrices which define the transfer function of the filter.

In the (Z_1, Z_2) plane the filter (15.85) is described by the transfer function

$$H(Z_1, Z_2) = \frac{\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} a(n_1, n_2) Z_1^{-n_1} Z_2^{-n_2}}{\sum_{n_1=0}^{M_1-1}\sum_{n_2=0}^{M_2-1} b(n_1, n_2) Z_1^{-n_1} Z_2^{-n_2}} \qquad (15.86)$$

If $b(0, 0) = 1$ and $b(n_1, n_2) = 0$ for n_1 and $n_2 \neq 0$, Eq. (15.86) reduces to

$$H(Z_1, Z_2) = \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} a(n_1, n_2)\; Z_1^{-n_1}\; Z_2^{-n_2} \qquad (15.87)$$

With a corresponding difference equation in the form

$$y(m_1, m_2) = \sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1} a(n_1, n_2)\, x(m_1 - n_1, m_2 - n_2) \qquad (15.88)$$

As in the 1-D case, Eqs (15.87) and (15.88) define an FIR system, while Eqs (15.85) and (15.86) represent an IIR system.

In the 2-D case, it is necessary to select the choice of suitable initial conditions and recursivity problems, which are of greater importance due to increase in dimensionality. In the 2-D case, the problem is more involved, because the possibility of different recursive relations with zero initial conditions is increased. 2-D linear filters are further divided into quadrant recursive filters and half plane recursive filters.

15.19.3 Recursive and Non-Recursive Realisation

In recursive realisation, the present output value depends on the input (past and present values), as well as the previous value of the output. It can usually be recognized by the presence of both a and b terms in the input and output relationship [Eq. (15.79)].

The recursive realisation is illustrated in Fig. 15.31(a). Since the impulse response of recursive filters extends to infinity, this realisation is particularly suited for implementation of IIR filters. In non-recursive realisation, the present value of the output depends only on the present and past values of the input.

Equivalently, the coefficients b in Eq. (15.79) are zero for this case. A typical realisation is shown in Fig. 15.31(b). Since the impulse response yielded by this type of realisation if finite, the resulting filters are always stable, in contrast to the possible instability of filters resulting from recursive realisation.

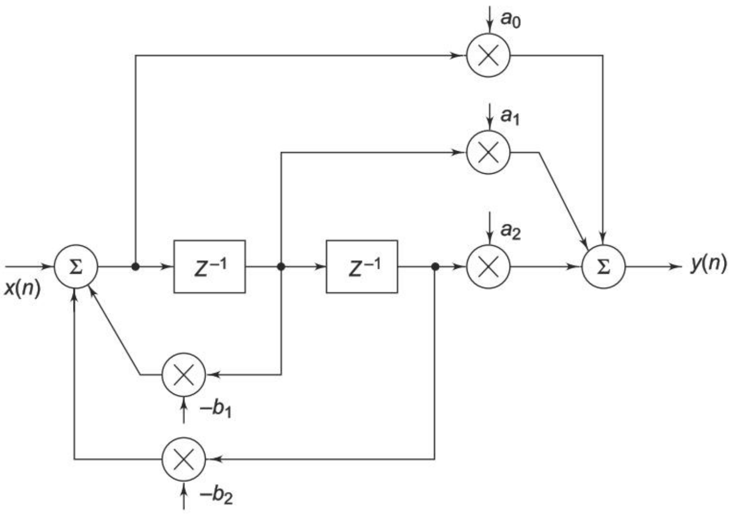


Fig. 15.31 (a) Recursive realisation

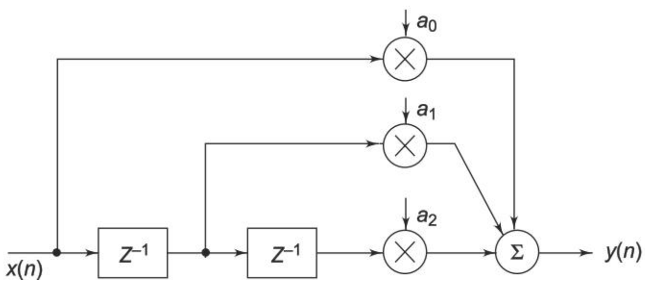


Fig. 15.31 (b) Non-recursive realisation

15.19.4 Fast Fourier Transform

The Fast Fourier Transform is a computational tool which facilitates signal analysis, such as power spectrum analysis and filter simulation using digital computers.

In digital analysis of signals, the continuous waveform to be analysed is first sampled and then converted to digits that represent the amplitude of the samples. The digital analyser operates on this number sequence and produces another number sequence as digital processing. If the resulting sequence represents the samples of a time sequence, it can be converted to analog form by the use of D/A converter, to form a continuous waveform in the time domain. In the case of an FFT of the number sequence, the result of the digital process is a number sequence, which represents in the frequency domain the characteristics of the continuous waveform that formed the basis for the input number sequence.

The Fourier transform of a number sequence is referred to as the Discrete Fourier Transforms (DFT) and is analogous to the Fourier series and the Fourier intergral transform of continuous and transient time signals.

The DFT defines the spectrum of a time series. The convolution of a two time series is equivalent to the multiplication of the DFTs of the two time.

The FFT is a highly efficient procedure for computing the DFT of a time series. The calculation of the DFT of a time series have $N = 2^n$ samples involves N^2 arithmetic operations; the same can be done by a FFT with only $2Nn = 2N \log_2 N$ arithmetic operations.

In a computer that takes about half an hour to do the DFT calculation in the conventional way for $N = 8192$ samples, the calculation time required using FFT is only about five seconds.

FIR 1-D DIGITAL FILTER DESIGN (THE WINDOW METHOD) 15.20

We describe the Fourier series technique as the first design method of 1-D FIR digital filters [Eq. (15.83)]. In this method, given a frequency function $H(e^{j\omega})$ with period 2π, an FIR digital filter with a transfer function corresponding to $H(e^{j\omega})$ can be obtained by simply computing the coefficients of the Fourier series of $H(e^{j\omega})$ and using these coefficients as the impulse response of the filter.

The main problem is that in general the Fourier series so obtained has an infinite number of coefficients, resulting in a filter which cannot be imple-mented. It is therefore necessary, in this case to truncate the Fourier series so obtained in some way. This truncation unfortunately leads to Gibb's pheno-menon, which causes the appearance of an overshoot in the approximated frequency response near a discontinuity.

To illustrate this point, let us consider the design of a low pass filter whose ideal response is as shown in Fig. 15.32(a). The magnitude of this filter is defined equal to unity in the interval $(0 - \omega_c)$ and equal to zero in the interval $\omega_c - \pi$. The impulse response $\{h(n)\}$ can be obtained by computing the coefficients of the Fourier series for $H(e^{j\omega})$, which are given by

$$h(n) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega = \frac{\sin(n\omega_c)}{\pi n} \tag{15.89}$$

This response has the form shown in Fig. 15.32 (b) for $N = 101$ (where N = number of impulse response).

If now we consider the frequency transfer function of the filter resulting from a truncated Fourier series whose coefficients are as shown in Fig. 15.32(b), i.e., Eq. (15.89), we obtain the result shown in Fig. 15.32(c).

Figure 15.32(c) shows the amount of the error due to Gibb's phenomenon and the fact that a finite transition band substitutes for the discontinuity in the ideal frequency response.

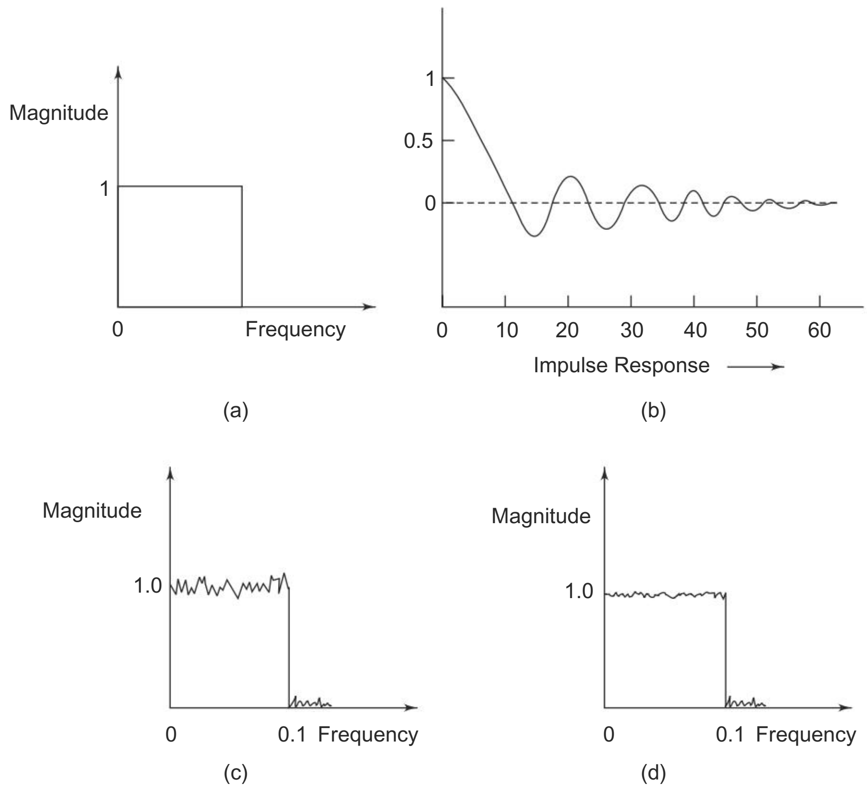


Fig. 15.32 (a) Frequency response of an ideal filter
(b) Impulse response of the ideal filter (low pass) for N = 101
(c) Freq. response of a rectangular low pass filter (N = 101)
(d) Freq. response of a rectangular window low pass filter (N = 201)

The error introduced by the transition of the Fourier series, which in the case of an ideal low pass filter or band pass filter is 9% of the value of discontinuity, is independent of the length of the impulse response.

When the impulse response is lengthened, the maximum error is the same but the oscillations are confined to a smaller frequency range, as can be seen in Fig. 15.32(d), where the transfer function of the low pass filter with $N = 201$ is shown.

The window method can also be directly applied to the design of 2-D FIR filters.

A 2-D filter with a frequency response $H\left(e^{j\omega_1}\, e^{j\omega_2}\right)$, periodic in ω_1 and ω_2 can be synthesised by using a impulse response, the sequence $\{h(n_1, n_2)\}$ obtained by computing the coefficients of a 2-D Fourier series of the periodic frequency response.

DESIGN METHODS FOR IIR DIGITAL FILTERS 15.21

15.21.1 Transformation of Passive Ladder Filter

Fettveis first suggested the use of one port reflection coefficients for the transfer of analog filter properties from the analog domain to the digital filter domain, and thus produced the wave digital filter structure.

The reason why reflection coefficients should be used is that a direct signal flow graph representation of the ladder in terms of its *V-I* relationship leads to structures in the digital filter domain using the bilinear transformation which are unrealizable, in that they contain loops without delay.

Basically a one port impedance Z, as shown in Fig. 15.33, can be described by its reflection coefficient

$$B = A\frac{Z - R}{Z + R} \tag{15.90}$$

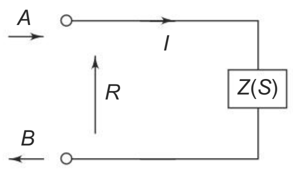


Fig. 15.33 One port of a typical ladder component

where R is an arbitrary reference resistance.

The numerator of Eq. (15.90), if delay free loops are to be avoided, must contain an overall factor of Z^{-1}, that is it must vanish for $Z^{-1} = 0$. Hence we must have

$$Z - R = 0 \tag{15.91}$$

Thus all impedances of a ladder can be treated as one port and suitably transformed. Equation (15.91) needs to be applied with different values of R to all ladder impedances.

The interconnection of these one ports however requires an impedance level adjustment which in the Fettveis nomenclature corresponds to an adapter. For example, if we take the series situation of Fig. 15.34 with $R_1 \ldots R_N$ as shown, the interconnections means that

$$i_1 = i_2, \ldots i_N \tag{15.92}$$

and
$$V_1 + V_2 + \ldots + V_N = 0 \tag{15.93}$$

in addition
$$A_r = V_r + R_r I_r \tag{15.94}$$
$$B_r = V_r - R_r I_r$$

($r = 1, \ldots N$) and hence

$$B_r = A_r - \alpha_r A_o \tag{15.95}$$

where

$$\alpha_r = 2R_r/R$$
$$A_o = A_1 + A_2 + \ldots + A_N$$
$$R = R_1 + R_2 + \ldots + R_N \tag{15.96}$$
$$\alpha_1 + \alpha_2 + \ldots + \alpha_N = 2 \tag{15.97}$$

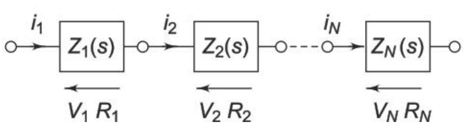


Fig. 15.34 Series connection of impedances

The digital structure realizing these conditions is known as a series adapter, and is represented symbolically as in Fig. 15.35.

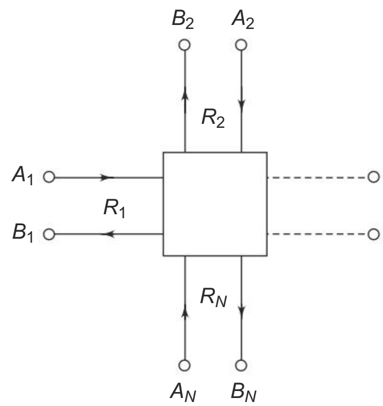


Fig. 15.35 Series adapter

Similar structures can be derived for parallel connection of one port. A two port series-parallel adapter is shown in Fig. 15.36. It facilitates the interconnection of two ports of different normalization resistances.

$$\alpha = (R_2 - R_1)/(R_2 + R_1)$$
$$B_1 = A_2 + \alpha(A_1 - A_2)$$
$$B_2 = A_1 + \alpha(A_1 - A_2)$$

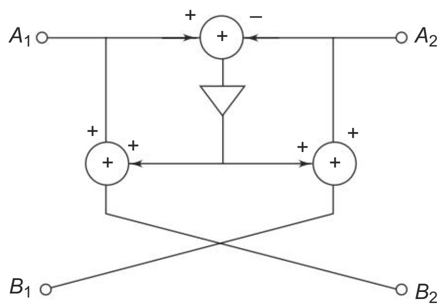


Fig. 15.36 A series-parallel adapter

A simple example of a series tuned circuit equivalent is shown in Fig. 15.37.

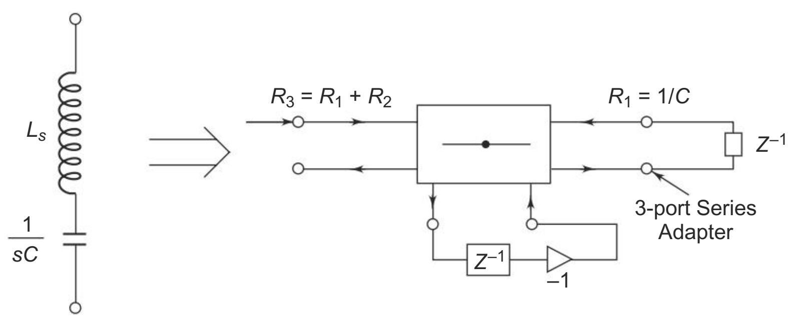


Fig. 15.37 Series tuned circuit and its digital equivalent

A parallel tuned circuit and its digital equivalent are shown in Fig. 15.38.

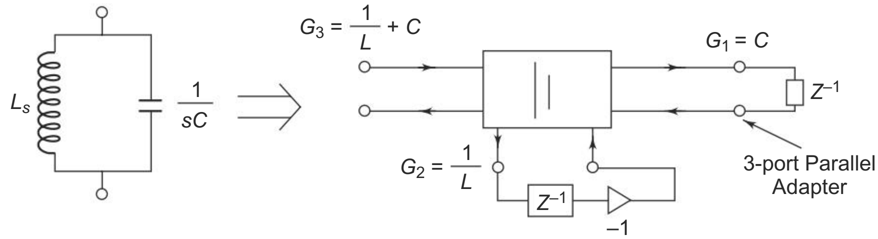


Fig. 15.38 Parallel tuned circuit and its digital equivalent

Digital structure realisation theory has also been applied to derive the filter of Fig. 15.39, where the original passive filter is also shown.

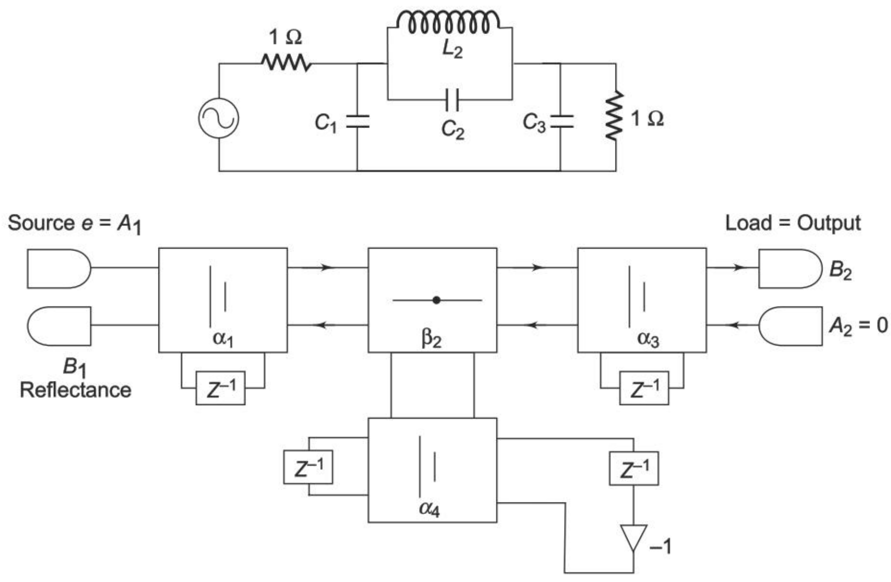


Fig. 15.39 Example of a wave digital filter obtained by the fettveis one port approach

Note that the digital realization structure requires adders, multipliers and delays. The number of delays, of course, is consistent with the number of storage elements of the original passive filters. A more direct approach to wave digital filters is to treat the ladder elements of the passive ladder as two ports. In that case we need floating impedance and shunt admittances to realize a ladder. We shall not go into the details, since they are beyond the scope of this book.

1-D IIR FILTER DESIGN 15.22

In this section some practical considerations relating to the methods for the design of IIR digital filters are presented. In particular, the bilinear transform is considered along with its application to the digitization of rational transfer functions of classical analog filters defined on the s-plane or in the framework of the wave digital filter approximation.

15.22.1 Review of Classical Analog Filters

Before going into the details of digitization methods, let us briefly review the main properties of the common families of analog low pass filters.

The first family of filters considered is the Butterworth filter, which consists of filter which are normally flat, that is, filters having the $2N - 1$ derivative of the squared magnitude. For an Nth order filter with zero at the origin, the squared magnitude response has the form

$$|H(j\,\omega)|^2 = \frac{1}{1+\left(j\,\omega/\omega_c\right)^{2N}} \tag{15.98}$$

where ω_c is the cutoff frequency at which the magnitude response has an amplitude of –3db. These filters with monotonic pass band and stop band behaviour have equally spaced poles on the s-plane, on a circle of radians.

Chebyschev filters have an equiripple behaviour, which minimises the maximum error in one of the two bands, the pass band or the stop band. On the other hand in which the minimal condition is not applied, the behaviour is monotonic. The squared magnitude response is written in the form

$$|H(j\,\omega)|^2 = \frac{1}{1+\varepsilon^2 V_N^2\,(\omega/\omega_c)} \tag{15.99}$$

where V_N is a Chebyschev polynomial of order N.

A third family of filters is that of elliptic filters. An elliptic filter has equiripple behaviour in both stop and pass bands. It is an optimum filter, in the sense that for a given order and given ripple specifications, it has a narrow transition band width. The squared magnitude response is given by

$$|H(j\,\omega)|^2 = \frac{1}{1+\varepsilon^2 S_n^2\left(\omega,k_1\right)} \tag{15.100}$$

where ε is a parameter related to the pass band filter ripple equal to $1 \pm 1/(1+\varepsilon^2)$, where $k_1 = \varepsilon/\sqrt{A^2-1}$ is a parameter related to the stop band $1/A^2$, and S_n is the Jacobi elliptic function.

15.22.2 Design of Digital IIR Filters by Means of the Bilinear Transform

The poles of the filter discussed above can be easily obtained, and expressed in the form

$$U_m = \frac{1-x_m^2-y_m^2}{\left(1-x_m^2\right)+y_m^2} \text{ and } V_m = \frac{2y_m}{\left(1-x_m^2\right)+y_m^2} \tag{15.101}$$

where
$$x_m = a\tan\frac{\omega_c}{2}\cos\frac{m\pi}{N} \tag{15.102}$$

$$y_m = b \tan \frac{\omega_c}{2} \sin \frac{m\pi}{n}$$

m = 0, ..., 2N – 1, where N is odd and

$$x_m = a \tan \frac{\omega_c}{2} \cos \frac{2m+1}{2N} \pi \tag{15.103}$$

$$y_m = b \tan \frac{\omega_c}{2} \sin \frac{2m+1}{2N} \pi$$

m = 0, 1, ..., 2N – 1, when N is even, a and b are equal to the 1 in the Butterworth case.

Thus when the parameters of the design, N, ω_c and ε are known in the Chebyschev case, the coefficients of the filter can be obtained by computing the pole positions by means of either Eqs (15.102) or (15.103).

The problem is now to investigate the relationship between the order x of the filter N, the pass band deviation δ, and the transition bandwidth Δf, defined by the cutoff frequency f_c and the frequency f_a at which the squared magnitude frequency response is less than or equal to $1/A^2$.

Let us now consider the three types of design specifications.

In the first case (Δf and δ fixed), the design procedure has to start with the evaluation of the order of the filter necessary to meet the specifications in terms of the desired attenuation, transition bandwidth and pass band deviation.

The pass band deviation can be controlled in the case of Chebyschev filters be ε. In any case, having defined f_c and f_a (i.e. transition bandwidth), the desired value $1/A^2$ of $|H(e^{j\omega})|^2$ at f_a and ε in the Chebyschev case, it is possible to determine N iteratively, starting from a first order filter and increasing the order of the filter to the point where the attenuation at f_a is greater than the desired value. At this point the design is completely determined.

In the second case (N and Δf fixed), the design is completely determined for the Butterworth filter case by obtaining the value of the attenuation at f_a directly.

In the third case (N and δ fixed), the filter is completely specified and the transition band width is directly obtainable during the design procedure.

A computer program is presented which designs Butterworth and Chebyschev filters by means of the above relation. It also computes the coefficients of their cascade structure. The inputs to the program are the critical frequencies f_c and f_a, the values of the desired attenuation of the filters at f_a and the value of the maximum pass band ripple if a Chebyschev filter is to be designed. The order of the filter is computed iteratively.

Two examples of the filter design with this program are shown in Figs 15.40 (a) and (b).

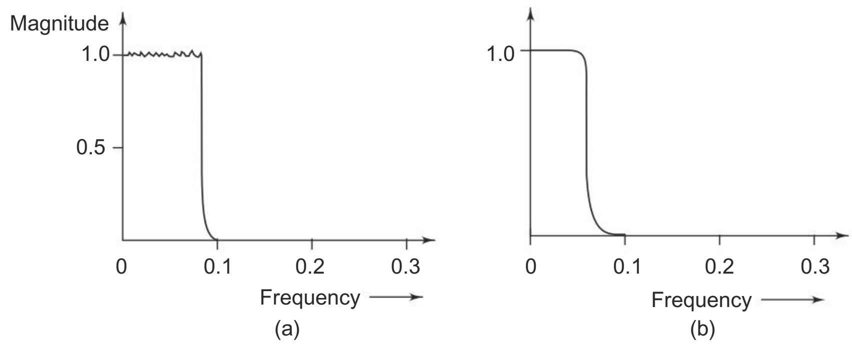


Fig. 15.40 (a) Example of IIR Chebyschev filter (N = 12, f_c/f_s = 0.1)
(b) Example of IIR Butterworth filter N = 8, f_c/f_s = 0.05

Assuming a realization structure by means of second order sections, only even order filters are designed by the program. However, it is quite simple to modify the program to design odd filters by replacing Eq. (15.102) with Eq. (15.103) and introducing a first order section in the structure.

The program presented here can be used to design only low pass filters. To design other types of filters, such as band pass, high pass and band stop, it is possible to start with the design of a normalized filter and then apply the appropriate frequency transformation, as shown in Table 15.6. A simple routine (TRASF) to perform these transforms is presented.

Table 15.6 Frequency Transformations from a Low Pass Digital Filter Prototype of Cutoff Frequency β

Filter Type	*Transformation*	*Associated Design Formulas*
Low Pass	$\dfrac{Z^{-1}-\alpha}{1-\alpha Z^{-1}}$	$\alpha = \dfrac{\sin\left(\dfrac{\beta-\omega_c}{2}\right)T}{\sin\left(\dfrac{\beta+\omega_c}{2}\right)T}$
High Pass	$\dfrac{-Z^{-1}+\alpha}{1+\alpha Z^{-1}}$	$\alpha = \dfrac{-\cos\left(\dfrac{\beta-\omega_c}{2}\right)T}{\cos\left(\dfrac{\beta+\omega_c}{2}\right)T}$
Band Pass	$\dfrac{-Z^{-2}-\dfrac{2\alpha KZ^{-1}}{K+1}+\dfrac{K-1}{K+1}}{\dfrac{K-1}{K+1}Z^{-2}-\dfrac{2\alpha KZ^{-1}}{K+1}+1}$	$\alpha = \cos\omega_o T = \dfrac{\cos\left(\dfrac{\omega_2+\omega_1}{2}\right)T}{\cos\left(\dfrac{\omega_2-\omega_1}{2}\right)T}$ $K = \cot\dfrac{(\omega_2-\omega_1)}{2}T\tan\dfrac{\beta T}{2}$

(Contd.)

Table 15.6 (Contd.)

Band Stop	$\dfrac{Z^{-2} - \dfrac{2\alpha}{1+K} Z^{-1} + \dfrac{1-K}{1+K}}{\dfrac{1-K}{1+K} Z^{-2} - \dfrac{2\alpha}{1+K} Z^{-1} + 1}$	$\alpha = \dfrac{\cos\left(\dfrac{\omega_2+\omega_1}{2}\right)T}{\cos\left(\dfrac{\omega_2-\omega_1}{2}\right)T} = \cos \omega_0 T$ $K = \tan \dfrac{(\omega_2-\omega_1)}{2} T \tan \dfrac{\beta T}{2}$

It is usually assumed that one has a low pass filter of a definite cut off frequency, say β rad/s, from which other low pass, high pass, band pass or band stop filters are required to be derived.

The low pass digital filter is said to be normalized when

$$\beta = \frac{1}{4}\left(\frac{2\pi}{T}\right) \tag{15.104}$$

PROGRAM FOR THE DESIGN OF BUTTERWORTH AND CHEBYSCHEV IIR DIGITAL FILTERS BY MEANS OF THE BILINEAR TRANSFORMATION 15.23

```
C  =      Program Ricor
C  =      Design of Butterworth and Chebyschev IIR digital filter
C  =      By means of the Bilinear transformation
C  =      Inputs:
C  card 1 –  Frequency transformation
C  J type =  0 – No transformation
             1 – Low pass to low pass
             2 – Low pass to High pass
             3 – Low pass to Band pass
             4 – Low pass to Band stop
C  F1 =   Lower Band-edge after transformation
C  F2 =   Upper Band-edge after transformation
C  In the 1 and 2 cases only F1 is used
C  card 2    –  Low Pass Prototype
C  I type    =  1 – Butterworth
                2 – Chebyschev
C  FT        =  Cut off frequency
C  FATT      =  Point of the frequency axis where the attenuation must be at least
                “ATTN”, dB.
C  RIPPL     =  Band pass ripple expressed as % of the magnitude value
C  ATTEN     =  Attenuation (In dB) at the FATT frequency
C  All the frequency values are expressed in the F/F’s scale
C  The coefficients of the cascade implementation of the prototype
```

```
C     Filter are first computed. In the low pass to band pass and band stop
C     Transformation fourth order sections are obtained
C     Required sub program - TRASF
      Compiler Double Precision
Dimensional Pol re (40), Polim (40), A (5, 20), B (5, 20), CEB(41)
Open 12, 'SYSOUT', ATT = 'AP'
Call FOPEN (9,'TMP')
IL    =       9
IS    =       12
Read Free (IL), Jtype,, F1, F2
Read free (IL), Itype, FT, FATT, RIPPL, ATTEN
PIE   =       3.141592653589
NU    =       3
FCA=          TAN (PIE * FT)
FATA          =   TAN (PIE * FATT)
FF = FATA/FCA
IF (1 type.EQ 2) EPS = SQRT [1./(1.-RIPPL/100) * 2 - 1]
ATT =         10 ** (ATTEN/10)
C     —       Computation of the order of the filter
N     =       1
2     GO TO (4,6), type
4     Res = 1/(1 + FT ** (2 * N)
      GO TO 8
6     CEB (1) = 1
      CEB (2) = FF
      IF (N.LP.1) GO TO 7
      F = FF + FF
      DO 5,I = 2,N
5     CEB (I + 1) = F* CEB(I) - CEB (I - 1)
7     RES = 1./[1. + CEPS * CEB(N + 1) ** 2]
8     IF (Res.LT.ATT) GO TO 9
      N = N + 1
      GO TO 2
9     If (Mod (N,2).EQ,1) N = N + 1
C     =       Pole position evaluation
      AT =    FCA
      BT =    FCA
      If (I type.EQ.1) GO TO 10.
AA    =       [(1. + SQRT(1. + EPS * * 2)/EPS] * * (1./Float(N)
AT    =       (AA – 1./AA) * AT/2
BT    =       (AA + 1./AA) * BT/2
10    NMax = N/2
      DO 20 JX = 1,N Max
      Tetap = PIE *(2 * Float (JX) - 1)/(2.*Float (N))
```

```
      Polre(JX) = - AT * Cos (Tetap)
20    Polim(JX) = BT * Sin (Tetap)
      DO 30 JZ = 1,N Max
      P1 = Polim(JZ) * * 2
      P2 = (1.- Polre(JZ)) * * 2
      P3 = Polre(JZ) * * 2
      Polre(JZ) = (1. - P3 - P1)/(P2 + P1)
```

$$30 \quad \text{Polim (JZ)} = \frac{(2.*\text{Polim}(J\,Z))}{(P_2 + P_1)}$$

```
      DO 40 I = 1,N Max
C     –       Evaluation of the second order section coefficient
      B(1,I) = 1
      B(2,I) = -2 * Polre(I)
40    B(3,I) = Polre(I) * Polre(I) + Polim(I) * Polim(I)
      If (Itype.EQ.1) Cost = 1
      IF (Itype.EQ.2) Cost = (1. - Rippl/100) * * (1./Float (N Max))
      DO 50 I = 1,N Max
      TOT = 4./[B(1,I) + B(2,I) + B(3,I)]
      CC = TOT/COST
      A(1,I) = 1./CC
      A(2,I) = 2./CC
50    A(3,I) = I./CC
      IF (J type .EQ.0) GO TO 60
C     –       Frequency transformation
      Call TRASF (A,B,5,20,N Max, NU,FT,F1,F2,J type)
60    Write (IS, 111) N
111   Format (1H1,1X, Filter order N =,I3)
      Write (IS, 112)
112   Format (///,1X , Numerator coefficients //)
      DO 70 I = 1,N Max
70    Write (IS,114) (A (IL,I),IL = 1,NU)
      Write (IS,113)
113   Format (/// IV, Denominator coefficients //)
      DO 80 I = 1,N Max
80    Write (IS, 114), (B (IL,I),IL = 1,NU)
114   Format (5 (4X,E21,14))
      STOP
      END
```

Subroutine for the frequency transformation of IIR digital filter

```
Compiler Double precision
Subroutine TRASF (A,B,NA,NB,NCEL,NU,FP,F1,F2,J TYPE)
C     –       Routine for the frequency transformation of IIR digital filters.
Dimension A(NA,NB),B(NA,NB),AN(5,3),FIN A(3),FIN B(3)
```

```
      PIE = 3.141592653589
      GO TO (10,20,30,40), J type
10    AL =   Sin (PIE * (FP – F1))/Sin (PIE * (FP + F1))
      A0 =   – AL
      A1 =   1.
      A2 =   0
      B0 =   A1
      B1 =   A0
      B2 =   0
      NU =   33
      GO TO 50
20    AL =   – Cos (PIE * (FP + F1)/Cos (PIE * (FP – F1))
      A0 =   –AL
      A1 =   –1
      A2 =   0
      B0 =   –A1
      B1 =   –A0
      B2 =   0
      NU =   3
      GO TO 50
30    AL =   Cos (PIE * (F2 + F1)/Cos (PIE * (FP – F1))
      CAPPA = TAN (PIE * FP)/TAN (PIE * (FP – F1))
      A0 =   – (CAPPA – 1)/(CAPPA + 1)
      A1 =   2.* AL * CAPPA/(CAPPA + 1)
      A2 =   –1
      B0 =   –A1
      B1 =   –A0
      NU =   5
      GO TO 50
40    AL =   Cos (PIE * (F2 + F1))/Cos (PIE * (F2 – F1))
      CAPPA    = TAN (PIE * FP) * TAN (PIE * (F2 – F1))
      A0 =   (1. – CAPPA)/(1. + CAPPA)
      A1 =   –2. * AL/(CAPPA + 1)
      A2 =   1
      B0 =   A2
      B1 =   A1
      B2 =   A0
      NU =   5
50    AN(1,1)   =   B0 * B0
      AN(2,1)   =   2.* B0 * B1
      AN(3,1)   =   B1 * B1 + 2. * B0 * B2
      AN(4,1)   =   2. * B1 * B2
      AN(5,1)   =   B2 * B2
      AN(1,2)   =   A0 * B0
```

```
      AN(2,2)   =  A0 * B1 + A1 * B0
      AN(3,2)   =  A0 * B2 + A1 * B1 + A2 * B0
      AN(4,2)   =  A1 * B2 + A2 * B1
      AN(5,2)   =  A2 * B2
      AN(1,3)   =  A0 * A0
      AN(2,3)   =  2. * A0 * A1
      AN(3,3)   =  A1 * A1 + 2. * A0 * A2
      AN(4,3)   =  2. * A1 * A2
      AN(5,3)   =  A2 * A2
      DO 60 L   =  1 ,NCEL
      DO 65 K   =  1 ,3
      FIN A(K)  =  A(K,L)
65    FIN B(K)  =  B(K,L)
      DO 60 K   =  1,5
      A (K,L)   =  0
      B (K,L)   =  0
      DO 60 I   =  1,3
      A(K,L)    =  AN(K,I) * FIN A(I) + A(K,L)
60    B(K,L)=  AN(K,I) * FIN B(I) + B(K,L)
      DO 70 K   =  1,NCEL
      RN     =  B (1,K)
      DO 70 I   =  1, NU
      A (I,K)=  A(I,K)/RN
70    B (I,K)=  B(I,K)/RN
      RETURN
      END
```

This routine can be used to transform second order sections obtained in the design procedure to produce second or fourth order sections if band pass and band stop filters are required to be designed.

MICROPROCESSOR BASED DIGITAL FILTER 15.24

The implementation of any digital filtering consists of a series of multiplications of samples of the input and/or output signals by constants, and of the additions of these products. Thus the main building blocks which have to be present in digital filtering hardware are the following:

(i) Memory cells
(ii) Adders
(iii) Multipliers
(iv) A programming unit which controls the sequence of operation

A general block diagram of a digital filtering hardware is presented in Fig. 15.41. It is composed of the following.

(i) A memory for the input samples
(ii) A memory for the output samples, if IIR operations have to be performed
(iii) A coefficient memory (ROM)

(iv) ALU, which is capable of performing multiplications and additions
(v) Control unit for the control of the sequence of operation.

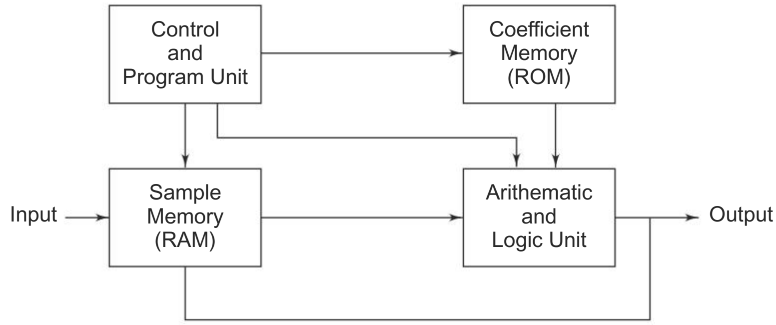


Fig. 15.41 Block diagram of a typical digital filtering processor

Several possibilities are now available for the implementation of the individual parts of this block diagram, which differ in speed of operation, level of integration, power consumption, etc. It is not possible to discuss all these possibilities, and all the structures which have been proposed for the imple-mentation of digital filtering hardware.

The most general and powerful concept that needs to be considered is that of multiplexing. There are a lot of applications where it is necessary to process several low frequency signals in parallel, e.g. in the case of spectral analysis performed by means of a bank of filters on a low frequency signal. In this case, if a fast arithmetic unit is available, all the filtering operations can be performed with the same arithmetic unit, which is used sequentially to perform the operations required for these different filters.

Thus the processors for low frequency data can share the same arithmetic unit between all the necessary channels. In the opposite case, when the signal to be processed is a very high frequency one, the operations to be performed can be shared between different arithmetic units which work in parallel.

As far as the control part is concerned, two possibilities can be considered. The first is its implementation as a logical network which generates all the signals necessary to control the processor. The second is to use a microprogra-mmed structure, where the processing unit is constructed to accept a set of instructions, and sequences of such instructions are memorized on a suitable, generally high speed, memory and executed during processor operation. This method is much more flexible than the fixed network considered before, bec-ause here the basic structure can be reprogrammed if different operations have to be performed by the processor.

APPLICATIONS OF DIGITAL FILTERS 15.25

The application of digital signal processing techniques in general, and of digital filtering in particular, has expanded to many important areas, such as speech signal processing, digital telephoning and communications, fascimile and TV image processing, radar-sonar systems, biomedicine, space research and operative system, geoscience, etc.

There are several reasons for the all embracing nature of this subject

1. The use of digital representation of signals and images in processing and transmission systems due to greater efficiency. This efficiency is due to accuracy, stability, noise immunity and other desirable effects.
2. The higher efficiency with which digital filtering can fulfill any realizable frequency response requirement, with great flexibility and adaptivity.
3. The decreasing cost of hardware and software implementation due to the production of highly reliable VLSI circuits and the availability of readily realizable software on minicomputers and microprocessors.

15.25.1 Typical Digital Filtering Operations

Digital filters in fact can be applied in many different parts of a signal and image processing transmission system. Two typical examples of systems in which digital filtering can be inserted are

(i) Local digital processing systems (Fig. 15.42), and
(ii) Digital communication system (Fig. 15.43)

In the first system of Fig. 15.42, an analog signal $x(t)$ is converted to digital form by an analog to digital converter, and thereby 1-D $x(n)$ or 2-D $x(n_1, n_2)$ data are obtained. The 1-D or 2-D digital filter in such systems performs the required operations which may correspond, for example, to some specific frequency requirements. This needs a suitable frequency transfer function $H(j\ \omega)$ or $H(j\ \omega_1, j\ \omega_2)$ and giving output data that can be immediately locally utilized, i.e. displayed or stored. This output data can be converted back to analog form by a digital to analog converter to produce a signal or an image reconstruction for direct use and presentation, e.g. on an oscilloscope, a plotter, display terminals, etc. It is interesting to observe that the same digital filter, or a suitable section of it, can also be used to give accuracy data interpolation to produce a reconstructed signal, or an image of better quality.

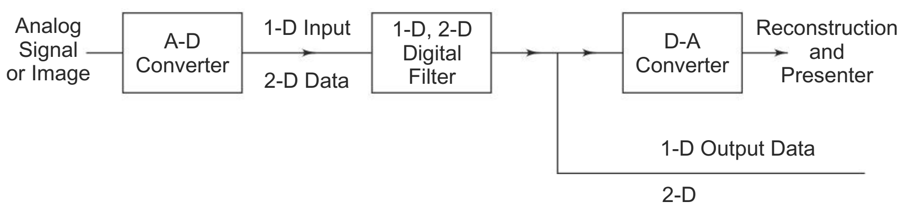


Fig. 15.42 General structure of a local digital processing system

In the second system of Fig. 15.43, an analog signal or image is converted to digital form as in Fig. 15.42, and the data so obtained is processed by a 1-D or 2-D digital filter as required, to specify and delimit the band widths or to perform suitable corrections on the signals.

The filtered data are hence sent through a modulator and a transmitter in the communication channel. The purpose of the modulator-transmitter is to represent the binary data in terms of AM, FM or phase modulation variations of a high frequency signal (carrier), suitable for transmission through the communication channel.

At the receiving terminal a receiver and demodulator reproduce the data which is usually not identical to that transmitted, but altered due to the noise and disturbances introduced in it. For this reason, data are recovered through a data reconstructor-synchronizer. In this way digital filtering can be performed on the received data, so as to extract pertinent information within specific frequency bands and reduce noise and other disturbances.

Alternatively, as shown by dashed lines in Fig. 15.43, one can apply digital filtering directly to the signals given by the demodulator converted to digital form. In this case digital filtering can perform the task of reconstructing data by extracting it from noise.

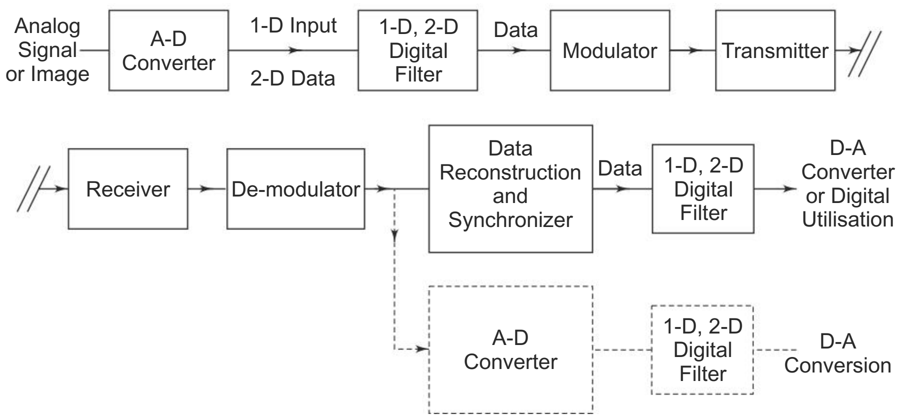


Fig. 15.43 General structure of a digital communication system

15.25.2 Multiple Filtering, Band Pass Analysis and Spectral Estimation

There exist many practical situations in which multiple filtering and band pass analysis are required. In multiple filtering, one specific operation (for example low pass, high pass, band pass, etc.) is required to be performed on many signal channels presented as input data, e.g. in multiplex systems. In band pass analysis, many band pass filtering operations, which are in general adjacent to one another so as to cover the entire signal spectrum, are performed on a single signal channel present as input data.

Figure 15.44 shows the typical structure of a digital filtering processor performing multiple filtering operations on N different input data channels. The different channel data are sent through a multiplexer to an input memory (RAM) in which the data are organised in the most convenient way, subject to the actual signal processing to be performed. From the memory the data are sent to a fast ALU, where they are processed using the appropriate filter coefficients given by another memory (ROM). The output data can again be obtained on N separate channels through a demultiplexer. All operations are under the timing and control of a general program control. If recursive type (IIR) digital filtering is to be performed, the additional output memory, shown with the dashed lines in Fig. 15.44, is required.

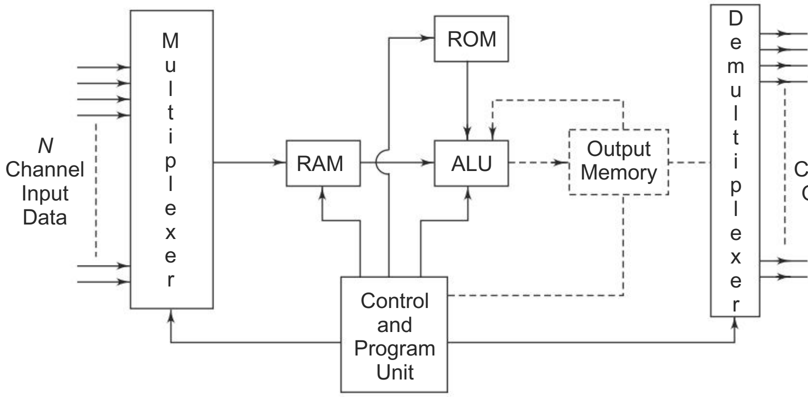


Fig. 15.44 Typical structure of a digital filtering processor performing multiple filtering operation on N input data channels

15.25.3 Applications in Radar Systems

Radar systems are of great importance for both military and civil applications.

The main problem of a typical radar system is the detection of return pulses; this substantially corresponds to determining the presence or absence of a pulse in noise. For such radar systems, the band width is usually chosen as the reciprocal of the pulse width to maximize the received signal to noise ratio (S/N). A matched digital filter which can maximize the S/N ratio is applied in this case.

The block diagram of a modern radar system is shown in Fig. 15.45. There is an A/D converter in the receiver to permit the processing of the return signals in digital form. The digital data is sent to the matched filter, followed by a digital detector. It is interesting to observe that subsequent digital processing

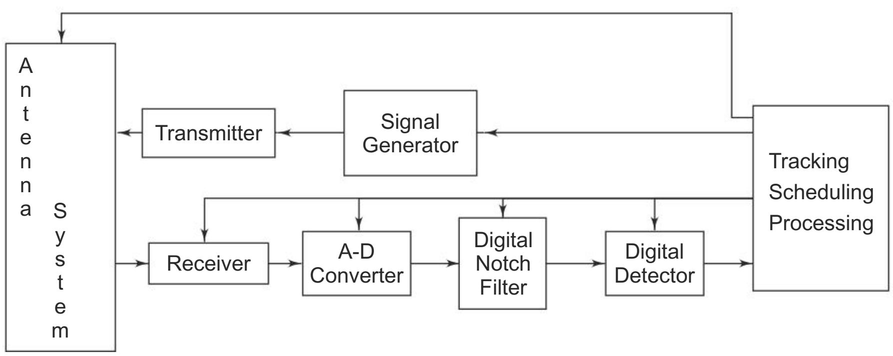


Fig. 15.45 Block diagram of a modern radar system

and interpretation can be done by a radar computer (software or hardware implemented), which also carries out the supervision and control of the whole radar system (in particular with procedures of tracking-scheduling it controls and adjusts the performance of the antenna, receiver, A/D conversion, digital matched filtering and digital detection). The radar system is thus a sophisticated digital adaptive system.

15.25.4 Applications in Biomedicine

Biomedicine is one area in which the use of digital signal processing techniques had great impact.

Computer Tomography A recent advance in diagnostic analysis is represented by computer tomography (CT) scanning. While in standard radiography (X-rays) a 3-dimensional object (part of the body) is represented on a plane (film), in CT the 3-D structure can be recovered. The basic idea of CT scanning is shown in Fig. 15.46.

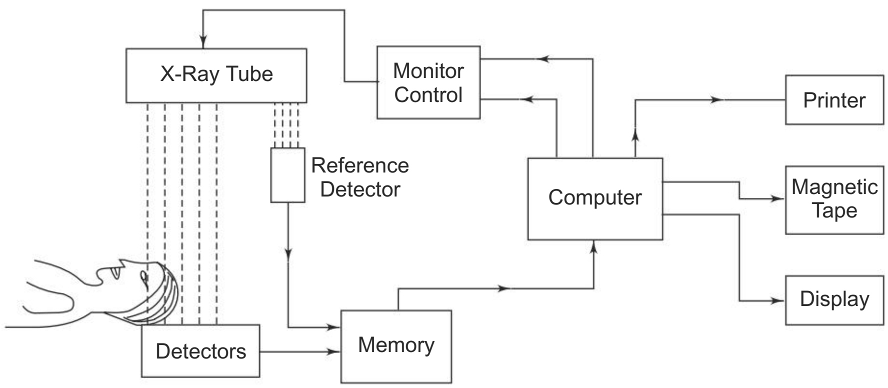


Fig. 15.46 Basic principles of CT scanning

A moving scanning system, including an X-ray source, collimators and detectors, rotates around a part of the patient's body, taking many readings of X-ray transmissions across the analysed part (e.g. the head). These readings are then sent along with the output of a reference detector to a large memory disk and to a computer, which through a suitable software program reconstructs a cross-section of the analysed part. The processing done by the computer is essentially a 2-D (image) reconstruction from a 1-D projection.

The use of digital filtering and FFT techniques is vital for various reconstruction techniques of images from their projections, known as Analytical reconstruction techniques.

Review Questions

1. Define a filter.
2. What are filter networks?
3. What do you mean by pass band and attenuation band?

4. What do you mean by cut-off frequencies?
5. List different types of filters.
6. State the fundamental theorem on filters.
7. Prove mathematically the fundamental theorem on filters.
8. Explain in brief passive filters.
9. What do you understand by active filters?
10. Compare between passive and active filters.
11. What do you understand by prototype filter or constant-K filter?
12. Why is a prototype filter also called a constant-K filter?
13. Draw the ideal responses of various filters.
14. Draw the circuit,frequency response and derive the cutoff frequency for an *RC* low pass filter.
15. Draw the circuit, frequency response and derive the cutoff frequency for an *RC* high pass filter.
16. Compare *RC* low and high pass filters.
17. Explain in brief active filters. How are active filters classified?
18. State the advantages and disadvantages of an active filter over a passive filter.
19. List the most commonly used active filters.
20. Draw and explain the frequency response of various active filters.
21. What are the important characteristics of an all pass filter?
22. Explain in brief Butterworth filter.
23. Explain with a diagram and frequency response the operation of a Butterworth low pass filter.
24. What do you understand by the order of active filters?
25. Explain with diagram the operation of a first-order Butterworth low pass filter.
26. Derive the gain magnitude and phase angle for a first-order low pass Butterworth filter.
27. Explain with a diagram and frequency response the operation of a second-order Butterworth low pass filter.
28. State the difference between a first-order and second-order Butterworth.
29. Derive the gain magnitude and phase angle for a second-order low pass Butterworth filter.
30. State the important types of active filters. Explain each type briefly.
31. Explain with a diagram the operation of a first-order Butterworth high pass filter.
32. Derive the gain magnitude and phase angle for a first-order high pass Butterworth filter.
33. Explain with a diagram and frequency response the operation of a second-order Butterworth high pass filter.
34. Derive the gain magnitude and phase angle for a second-order high pass Butterworth filter.
35. State the difference between a second-order Butterworth low pass and high pass filter.
36. What do you understand by higher order filters?
37. Explain with a diagram and frequency response the operation of a third-order low pass filter.
38. Explain with a diagram and frequency response the operation of a fourth-order low pass filter.
39. Compare third-order and fourth-order low pass filter.
40. What do you understand by band pass filters?
41. Explain with a diagram and frequency response the operation of a wide band pass filter.
42. Explain with a diagram and frequency response the operation of a second-order band pass filter.
43. Explain with a diagram and frequency response the operation

of a narrow band pass filter using multiple feedbacks.

44. Derive the gain of a narrow band pass filter.
45. What do you understand by band stop filters?
46. What is a notch filter? How can twin –T be used as a notch filter?
47. Explain with a diagram and frequency response the operation of a wide band reject filter.
48. Explain with a diagram and frequency response the operation of a passive notch filter.
49. Explain with a diagram and frequency response the operation of an active notch filter.
50. What do you mean by an all pass filter?
51. Explain with a diagram and wave-forms the operation of an all pass filter.
52. Derive the expression for the output voltage of an all pass filter with the output lagging.
53. What do you mean by universal filter? Why are they widely preferred?
54. Explain with a block diagram the operation of a FLT-U2 universal filter.
55. Explain with a block diagram how an FLT-U2 is used as a second-order low pass, high pass, and band pass filter. State the applications of a universal filter.
56. Explain in brief Chebyschev, Bessel and Elliptic filter.
57. Compare Butterworth and Chebyschev filter.
58. Why are Bessel filters preferred even though they have a ripple in the pass and stop bands?
59. What are digital filters?
60. Define discrete functions. How can they be used in digital filters?
61. State 1-D and 2-D sampling theorems.
62. State the fundamental properties of 1-D and 2-D digital systems.
63. Explain frequency domain representation in digital filters.
64. Derive the transfer functions of 1-D linear systems.
65. Explain in brief, finite impulse response (FIR) filters.
66. Explain in brief, infinite impulse response (IIR) filters.
67. State the differences between FIR and IIR filters.
68. What do you understand by zeros and poles?
69. Explain recursive realization with a diagram.
70. Explain non-recursive realization with a diagram.
71. Compare recursive and non-recursive realization.
72. Explain fast fourier transform (FFT) in brief.
73. How would you design an IIR digital filter by means of bilinear transform?
74. Explain with block diagram the operation of microprocessor based digital filter. State the applications of digital filter.
75. Explain with block diagram the working of the general structure of a local digital processing system.
76. Explain with a block diagram the working of a general structure of a digital communication system.
77. Explain with block diagram the working of a typical structure of a digital filtering processor using multiple filtering operations.
78. Describe with a block diagram the operation of modern radar systems.
79. Explain how digital filters can be used in a basic CT scanning.

Multiple Choice Questions

1. A network designed to attenuate certain frequencies and allow others to pass without attenuation is called a/an
   (a) amplifier (b) rectifier
   (c) oscillator (d) filter.
2 All passive filters are constructed using
   (a) purely resistance elements
   (b) purely reactive element
   (c) purely impedance element
   (d) none of the above
3. The constant –K filter are connected by the relationship
   (a) $Z_1 + Z_2 = R_o^2$ (b) $Z_1 - Z_2 = R_o^2$
   (c) $Z_1/Z_2 = R_o^2$ (d) $Z_1 \cdot Z_2 = R_o^2$
4. A low pass filter allows
   (a) high frequency to pass
   (b) low frequency to pass
   (c) mid band frequency to pass
   (d) upper and lower frequency to pass
5. A high pass filter allows
   (a) high frequency to pass
   (b) low frequency to pass
   (c) mid band frequency to pass
   (d) upper and lower frequency to pass
6. *RC* filters are commonly used for
   (a) RF or high frequency operation
   (b) audio or low frequency operation
   (c) VHF operation
   (d) UHF operation
7. LC filters are commonly used for
   (a) RF or high frequency operation
   (b) audio or low frequency operation
   (c) VHF operation
   (d) UHF operation
8. The gain of RC active filter falls by
   (a) 5 dbs (b) 3 dbs
   (c) 10 dbs (d) 1 dbs
9. The Butterworth filter also called a flat-flat filter has a
   (a) flat pass band and flat stop band
   (b) ripple pass band and flat stop band
   (c) flat pass band and ripple stop band
   (d) ripple pass band and ripple stop band
10. The Chebyschev filter has a
    (a) flat pass band and flat stop band
    (b) ripple pass band and flat stop band
    (c) flat pass band and ripple stop band
    (d) ripple pass band and ripple stop band
11. The elliptic filter has a
    (a) flat pass band and flat stop band
    (b) ripple pass band and flat stop band
    (c) flat pass band and ripple stop band
    (d) ripple pass band and ripple stop band
12. In the Butterworth first order filter, the gain rolls off at the rate
    (a) –3 dbs/decade
    (b) –10 dbs/decade
    (c) –20 dbs/decade
    (d) –40 dbs /decade
13. In the Butterworth second order filter, the gain rolls off at the rate
    (a) –3 dbs/decade
    (b) –10 dbs/decade
    (c) –20 dbs/decade
    (d) –40 dbs /decade
14. A band pass filter is called a wide band pass if the Q-factor is
    (a) less than 10
    (b) greater than 10
    (c) equal to 10
    (d) none of the above
15. A band pass filter is called a narrow band pass if the *Q*-factor is
    (a) less than 10
    (b) greater than 10
    (c) equal to 10
    (d) none of the above
16. A notch filter has a *Q*-factor that is
    (a) less than 10
    (b) greater than 10
    (c) equal to 10
    (d) none of the above

Practice Problems

1. Design a low pass filter having a cut-off frequency of 3 kHz and pass band gain of 3.
2. Design a low pass filter having a cutoff frequency of 10 kHz and pass band gain of 3.
3. Referring to Fig 15.9, calculate the value of resistance for a cutoff frequency of 50 kradian/s and C = 0.05 μF.
4. Design a second order low pass filter at a high cutoff frequency of 1.75 kHz.
5. Design a second order low pass filter at a high cutoff frequency of 5 kHz and $C_2 = C_3 = .0047$ μF.
6. Design a second order high pass filter at a low cutoff frequency of 100 Hz.
7. Design a second order high pass filter at a low cutoff frequency of 200 Hz and $R_2 = R_3 = 33$ kΩ.
8. Design a wide band pass filter with $f_1 = 150$ Hz, $f_H = 2$ kHz and a pass band gain of 4. Also, calculate the Q value of this filter.
9. Design a wide band pass filter with $f_L = 500$ Hz, $f_H = 5$ kHz and a pass band gain of 5. Also, calculate the Q value of this filter.
10. (i) Design a narrow band pass filter having a center frequency of $f_c = 500$ Hz, $Q = 4$ and $A_F = 6$. (Assume $C_1 = C_2 = C_3 = 0.033$ μF.)
    (ii) Change the frequency f_c to 1 kHz keeping gain A_F and bandwidth constant
11. (i) Design a narrow band pass filter having a center frequency of $f_c = 2$ kHz, $Q = 5$ and $A_F = 10$. (Assume $C_1 = C_2 = C_3 = 0.05$ μF.)
    (ii) Calculate the value of resistance required to change the frequency f_c to 5 kHz keeping gain A_F and bandwidth constant.
12. Design a wide band reject filter with $f_1 = 1$ kHz, $f_H = 150$ Hz.
13. Design a 100 Hz active notch filter. (Assuming $C = 0.047$ μF)
14. Determine the phase angle of an all pass filter, if the input frequency is 3 kHz. (Assume the data required).
15. Design an FLT-U2 as a second order inverting Butterworth low pass filter with a dc gain of 8, a cut off frequency of 1 kHz and $Q = 10$. (Determine the external values)
16. Design an FLT-U2 as a second order inverting Butterworth high pass filter with a dc gain of 6, a cut off frequency of 2 kHz and $Q = 8$. (Determine the external values)
17. Design an FLT-U2 as a notch filter with a notch out frequency of 2 kHz and $Q = 8$.

Further Reading

1. *The Royal Signals Handbook of Line Communications*, Produced by School of Signals, London, published by Her Majesty's Stationery Office, London, 1947.
2. C.S. Rangan, G.R. Sharma and V.S.V. Mani, *Instrumentation Devices and Systems,* Tata McGraw-Hill, 1983.
3. Ramakant A. Gayakwad, *Opamps and Linear Integrated Circuit Technology*, Prentice-Hall of India, 1987.
4. Coughlin and Driscoll, *Opamps and Linear Integrated Circuits*, 3rd Edition, Prentice-Hall of India, 1987.
5. Y. Cappellini, A.C. Constantinides and P. Emiuani, *Digital Filters and Their Applications*, Academic Press, 1978.
6. Malmstadt, Enke and Others, *Instrumentation for Scientists, W.A. Benjamin Inc., 1974.*

Chapter 16

Measurement Set-up

INTRODUCTION 16.1

Whatever the nature of applications, intelligent selection and use of measurement set-up (instrumentation) depends on a broad knowledge of what is available and how the performance of the equipment may be best described in terms of the job to be done. New devices is continuously being developed, but these basic devices have proved their usefulness in broad areas and undoubtedly will be widely used for many years. A representative cross-section of such devices is discussed in this chapter. These devices are of great interest in themselves, and also serve as a vehicle for the presentation and development of general techniques and principles needed in handling problems in measurement set-up. In addition, these general concepts are useful in treating any devices that may be developed in future.

MEASUREMENTS OF MICROWAVE FREQUENCIES 16.2

Measurement at microwave frequencies can be done by the use of wavemeters.

A wavemeter is an adjustable resonant circuit provided with a calibration that gives the resonant frequency in terms of the settings of the tuning adjustment.

Wavemeters are used to measure frequency when the higher accuracy of a primary or secondary standard is not required, and where simplicity and portability are important.

A wavemeter may employ any type of resonant circuit that is convenient for the frequency range to be covered. Hence resonant circuit based on lumped constants, co-axial two wire lines and cavities, all are used in a wavemeter.

Wavemeters may employ absorption, reaction or transmission devices.

In the absorption type, the wavemeter is equipped with means for indicating the current induced in it. The wavemeter is then loosely coupled to the oscillator whose frequency is to be determined, and is adjusted for maximum response.

In the reaction type, the adjustment of the wavemeter corresponding to the frequency being measured is determined from the reaction produced by the wavemeter upon the system being measured. For example, the resonant frequency of a low power oscillator may be readily measured when loosely coupled to the tank circuit of the oscillator. The dc grid current will drop abruptly when the coupled wavemeter is tuned through resonance with the frequency being generated by the oscillator.

In the transmission type, the wavemeter is used as a coupling device in a system that transmits power from a generator to a load or indicator. Such a system is proportional to the energy transfer, such that appreciable transmission of the energy to the load occurs only when the wavemeter is tuned to the frequency of the energy involved. Transmission wavemeters find considerable use at microwave frequencies.

RESONANT CO-AXIAL LINES 16.3

A co-axial line such as illustrated in Fig. 16.1 and operating as a resonant system can be used to measure frequencies in the range of 600 – 10,000 MHz.

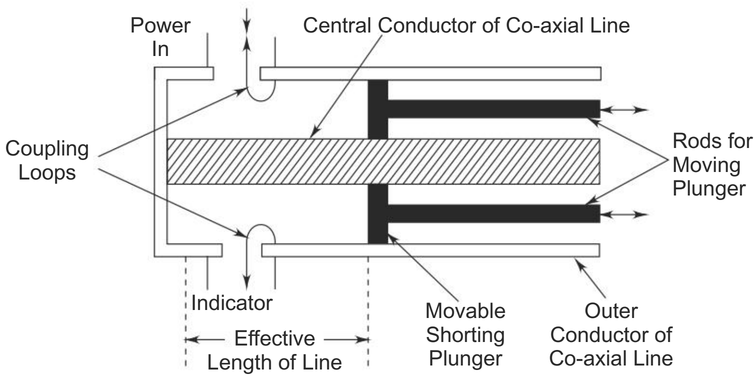


Fig. 16.1 Co-axial wavemeter

At lower frequencies the cavity becomes excessively long and at higher frequencies the dimensions become very small.

The arrangement illustrated is of the transmission type. It has two coupling loops, one for feeding power into the line through a co-axial cable, and one for coupling a crystal rectifier indicator to the oscillators in the cavity.

Tuning is accomplished by varying the position of the short-circuit plunger, using a lead screw. Provision is made for accurately measuring the resulting displacement of the plunger.

At shorter wavelengths, where the total line length available is one wavelength or more, frequency can be measured by determining the displacement Δl between adjacent maxima using the equation

$$f = \frac{150,000,000}{\Delta l} \tag{16.1}$$

where Δl is in metres.

At lower frequencies, the co-axial line is ordinarily adjusted to be exactly half a λ long. In this case, it is necessary to make a calibration of frequency as a function of line length.

The accuracy of co-axial wavemeters can be quite high (values of 0.05%). This accuracy is a result of the fact that the Q of a co-axial resonator is very high.

CAVITY WAVEMETERS 16.4

Resonant cavities find extensive use as wavemeter at microwave frequencies. They enjoy high accuracy, mechanical simplicity, and a large physical size in proportion to the wavelength being measured.

Practical cavity wavemeters are always in the form of cylinders which are tunable by means of a piston that varies the length of the cavity, as in Fig. 16.2.

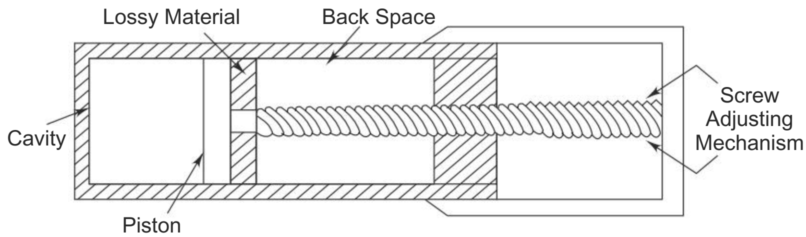


Fig. 16.2 Simple cylindrical cavity

Coupling to a wavemeter is accomplished by a loop or an orifice. A loop is used when co-axial systems are used to excite a cavity or carry power from it, while an orifice is normally used in a waveguides system.

The accuracy with which frequency can be determined using a well made cavity wavemeter is quite high. Precision of the order of 1 part in 1000 is easy to obtain, and 1 part in 100,000 can be obtained by careful mechanical design combined with temperature compensation and correction.

RF/UHF FIELD STRENGTH METER (METHODS FOR MEASURING THE STRENGTH OF RADIO WAVES) 16.5

There are two general methods available for determination of field strength.

One is the standard antenna method. In this, use is made of an antenna in which the relationship between the field strength of the radio wave and the equivalent voltage that the wave induces in the antenna. The field strength is then determined by measuring this equivalent induced voltage and making use of the relationship between it and the field strength.

The second method of measuring field strength is known as the standard field generator method or the substitution method. It consists of comparing the strength of the radio wave with a field of known strength produced by a standard field generator.

At frequencies above 300 MHz, it is more practical to deal with power than with voltage or current. Loop antennas used are not satisfactory at these frequencies and therefore it is customary to use either half wave dipoles, or a directional antenna such as a simple array, a horn or a parabola.

At very high frequencies, such antennas can be assumed to have negligible loss resistance, so that a substantial amount of power delivered is radiated and virtually all of the power that is absorbed from a passing radio wave is delivered to the load impedance associated with the antenna.

These considerations cause field strength measuring equipment intended for use at very high frequencies to be modified somewhat, as compared to the corresponding equipment intended for use at short wave and lower frequencies.

In the standard field generator method, the standard field can be determined from the radiated power and the antenna gain.

In the substitution method shown in Fig. 16.3, the input to the co-axial line can be switched from the receiving antenna to a signal generator, which is then adjusted to deliver the same power to the co-axial line as does the actual radio wave of unknown field strength.

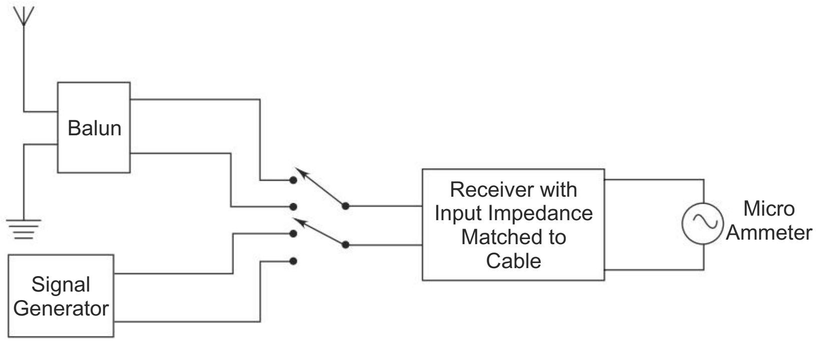


Fig. 16.3 Substitution method

The relationship between the power which the signal generator delivers to the line, and the field strength of the radio wave is then given by the equation

$$\varepsilon = \sqrt{\frac{480\pi^2 P_r}{\lambda^2 G}} \tag{16.2}$$

where ε – field strength
P_r – load power
G – antenna power gain in terms of the receiving antenna.

This arrangement requires that the signal generator be calibrated in terms of the power that it delivers to a terminated co-axial line, instead of in terms of the voltage that it develops across a resistance.

MEASUREMENT OF SENSITIVITY 16.6

The sensitivity of a radio receiver is its ability to amplify weak signals. It is often defined in terms of the voltage that must be applied to the receiver input terminals to give a standard output power, measured at the output terminals. For AM broadcast receivers, many of the relevant quantities have been standardised. Hence 30% modulation by a 400 Hz sine wave is used and the signal is applied to the receiver through a standard coupling network, known as a dummy antenna.

The standard output is 50 mW; for all types of receivers the loud speaker is replaced by a load resistance of equal value.

Sensitivity is often expressed in μ volts or in decibles below 1 V, and measured at three points along the tuning range when a production receiver is lined up. It can be seen from the sensitivity curve in Fig. 16.4, that sensitivity varies over the tuning band. At 1000 Hz, this particular receiver has a sensitivity of 12.7 μ V or – 98 db V (db below 1 V).

The most important factors for determining the sensitivity of a superheterodyne receiver are the gains of the IF and RF amplifiers.

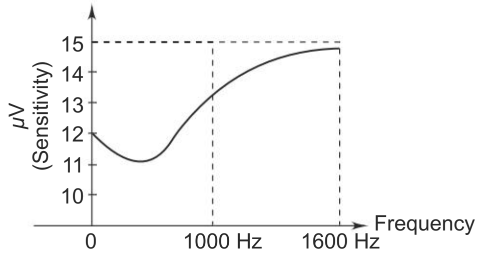


Fig. 16.4 Sensitivity curve

MEASUREMENT OF SELECTIVITY 16.7

The selectivity of a receiver is its ability to reject (adjacent) unwanted signals. It is expressed as a curve, such as the one in Fig. 16.5, which shows the attenuation that the receiver offers to signals at frequencies near the one to which it is tuned. Selectivity is measured at the end of a sensitivity test under the same conditions, except that now the frequency of the generator is varied to either side of the frequency to which the receiver is tuned. The output of the receiver obviously falls (attenuation increases), since the frequency is incorrect. Hence the input voltage must be increased until the output is the same as it was originally. The ratio of the voltage required at resonance to the voltage required when the generator is tuned to the receivers frequency is calculated at a number of points and then plotted in decibels, to give the curve shown in Fig. 16.5.

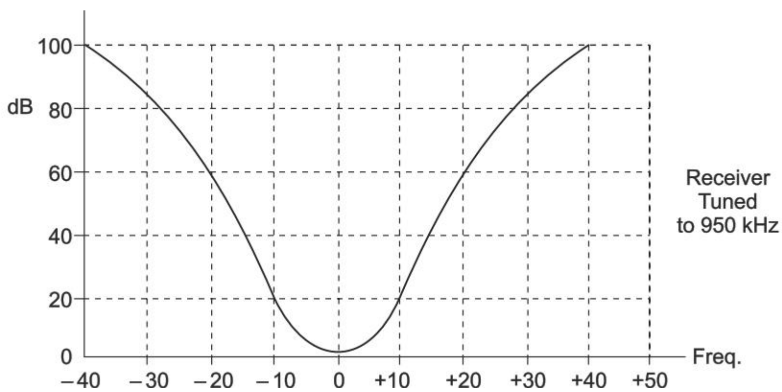


Fig. 16.5 Typical Selectivity Curve

From the curve, it can be seen that at 20 kHz below the receiver tuned frequency, an interfering signal would have to be 60 db greater than the wanted signal, to come out with the same amplitude.

Selectivity varies with receiving frequency if ordinary tuned circuits are used in the IF section, and becomes somewhat worse when the receiving frequency is raised.

INTERMODULATION METHOD OF MEASURING NON-LINEAR DISTORTION 16.8

16.8.1 SMPTE Method

One of the methods of measuring non-linear distortion consists of simultaneously applying two sinusoidal voltages of different frequencies to the amplifier input and observing the sum, difference and various combination frequencies that are produced by the non-linearity of the amplifier. This method of determining the distortion is called the intermodulation method.

The two signals may be combined under many different conditions, but two cases are important.

In the first case, a low frequency test signal and a high frequency test signal of somewhat smaller amplitude are simultaneously applied to the amplifier. If non-linear distortion and its effects are present, then the amplification that the high frequency component f_2 experiences will vary at the frequency f_1 of the low frequency signal, as shown in Fig. 16.6.

Hence the signal of frequency f_2 in the amplifier output is modulated at the low frequency f_1. The amount of distortion is then expressed in terms of the modulation experienced by the high frequency wave, according to the relation

$$\text{Intermodulation distortion (SMPTE method) \%} = 100 \times \sqrt{m_1^2 + m_2^2 + \ldots} \qquad (16.3)$$

where m_1 = degree of modulation that the output component of frequency f_2 experiences at frequency f_1

m_2 = corresponding modulation of f_2 at a frequency $2f_1$.

This particular method of employing two signals to determine non-linear distortion is widely used in AF systems. It is commonly referred to as the SMPTE intermodulation method, since it has been standardised by the Society of Motion Pictures and Television Engineers (SMPTE).

(An alternative method of explaining this situation is to consider that the non-linear action of the amplifier produces sum and difference frequencies, $(f_1 + f_2)$ and $(f_1 - f_2)$ respectively. These frequency components, together with f_2, can be thought of as representing the side bands and a carrier respectively, thus forming a wave of carrier frequency f_2 modulated at frequency f_1. If the non-linear distortion is also sufficient to produce second order combination frequencies $f_2 + 2f_1$, the modulation envelope then contains a component that is the second harmonic of f_1. The modulation envelope is then not sinusoidal.)

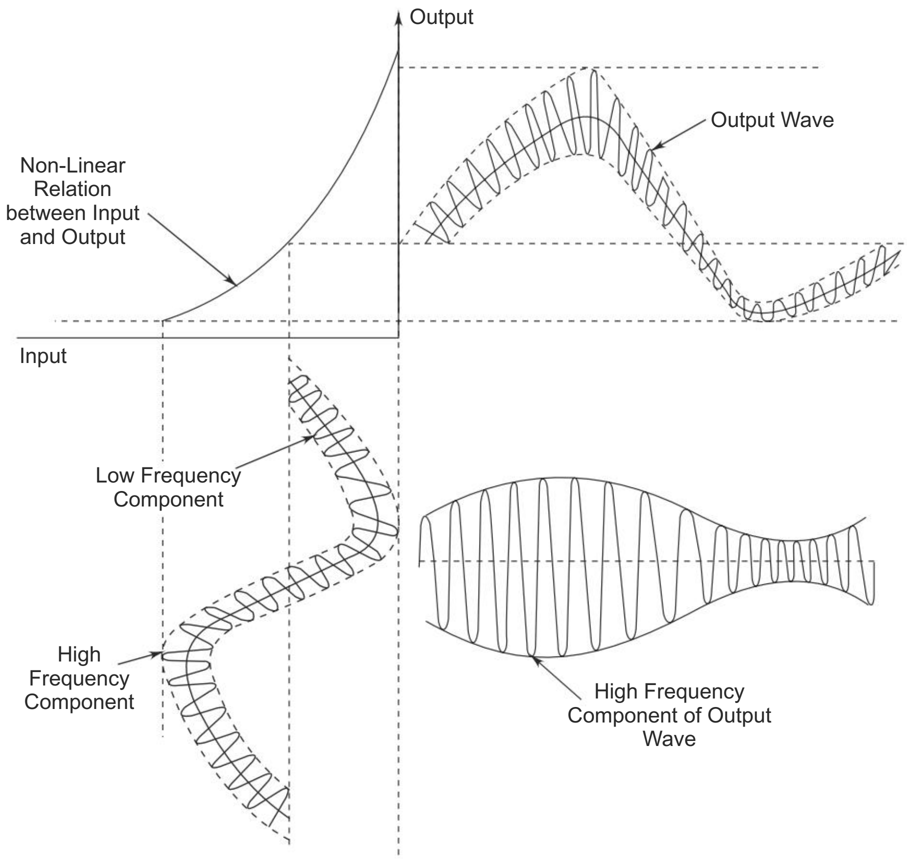


Fig. 16.6 Non-linear relationship between input and output (results in modulation)

The block diagram for the setup of determining intermodulation by the SMPTE method is shown in Fig. 16.7.

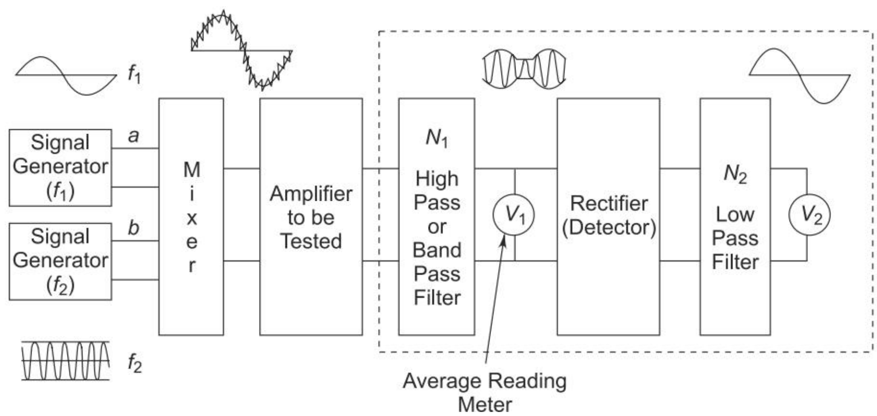


Fig. 16.7 Block diagram for setup of determining intermodulation by the SMPTE method

The system consists of two signal generators, generating frequencies f_1 and f_2 respectively, and combining these two outputs from the signal generators. The combining systems may take any of a number of forms, such as a transformer with centre tapped primary, or two amplifier active devices with separate inputs but common output load impedance. (Alternatively, it is possible to use a two frequency signal generator specially designed for intermodulation test.)

The network N_1 separates the high frequency (f_2) and its side bands from other frequency components that may be present in the amplifier output. Voltmeter V_1 measure the amplitude of the resulting modulated wave, i.e. it measures the carrier amplitude E_2.

The modulation envelope is determined by rectifying (detecting) the modulation wave, separating the modulation frequency component of the envelope by a low pass filter N_2, and determining the amplitude of the modulation component by voltmeter V_2. The dotted block indicates a system for measuring the characteristics of a modulated wave. Calibration can be carried out by applying known voltages of frequencies f_1 and f_2 to terminals *a* and *b*.

In such a measuring system, shown in Fig. 16.7, the normal practice is to make the low frequency test signal at least four times as large as the high frequency test signal applied to the system. The low frequency used is about 50 – 100 Hz, while the high frequency is typically about 5 kHz or more.

The low pass filter N_2, must have a sufficient band width to pass second order side bands, that is, a cutoff frequency greater than $2f_1$. Band pass filter N_1 must have a corresponding band width greater than $4f_1$.

The voltmeter V_1 used must be an average-reading type and voltmeter V_2 must be a square law device. An average reading meter, such as a copper oxide rectifier instrument gives sufficiently accurate results for ordinary requirements. In no case should voltmeter V_2 be a peak reading meter.

The value of the intermodulation distortion determined by the SMPTE method depends basically on the non-linearity met by the low frequency test signal, because this signal is much larger under standard test conditions. The observed intermodulation distortion is largely dependent on the value of the high frequency f_2, but depends at least slightly on the value of f_1.

16.8.2 CCIF Method

In the second method of carrying out an intermodulation test, the two test signals used have equal amplitude, and relatively high but slightly different frequencies.

When non-linear distortion is present, a difference frequency component appears in the output and is used as a measure of the amount of distortion present.

Figure 16.8 illustrates the frequency relations, where f_a and f_b represent the two test frequencies of amplitude E_a and E_b, and E_d is the amplitude of the resulting difference frequency component f_d.

(Second order combination frequencies E_{d2} and E'_{d2} are also shown in Fig. 16.8, and are a measure of the cubic type of distortion. These components can however be detected only by the use of a highly selective wave analyzer, and are generally not made use of in this form of the intermodulation test.)

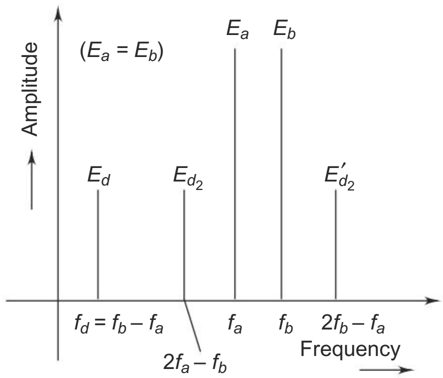


Fig. 16.8 Frequency relation existing in intermodulation distortion (CCIF method)

The intermodulation distortion is expressed by the relation

$$\text{Intermodulation distortion (CCIF) \%} = \frac{E_d}{E_a + E_b} \times 100 \qquad (16.4)$$

This form of intermodulation test has been recommended by the International Telephonic Consultative Committee, and is referred to by the initials CCIF.

The basic block diagram setup for measuring intermodulation distortion tests by the CCIF method is shown in Fig. 16.9.

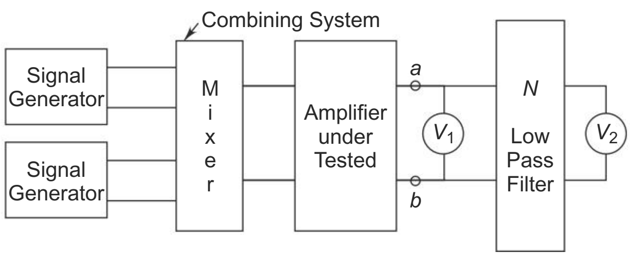


Fig. 16.9 Block diagram for setup of determining intermodulation by the CCIF method

The difference frequency component E_d is separated by a low pass filter N and indicated separately by the voltmeter V_2. The amplitude of test signals E_a and E_b can be determined with the help of V_1. If V_1 is the peak reading instrument, then the peak voltage observed at terminals $a - b$ will be very close to the peak of $E_a + E_b$.

Alternatively, E_a and E_b can be determined individually if test voltages of frequencies f_a and f_b are applied one at a time to the input terminals of the amplifier under test. (Both these procedures involve approximation as a result of failure to allow for other frequency components that are simultaneously present at terminals $a - b$, as a result of the distortion of the amplifier. For exact determination of E_a and E_b, a highly selective wave analyzer is used. However, the approximation procedure is adequate for most practice applications.)

For intermodulation test by the CCIF method, the test frequencies f_a and f_b are generally centered towards the high frequency end of the response range, and are sometimes placed purposely in the region where the high response begins to fall off appreciably. The difference frequency $f_d = f_b - f_a$, between the two test frequencies, is generally a moderate value from 50 – 200 Hz. This difference frequency should not be so small as to be below the normal low frequency response range of the amplifier, as the amplifier circuits tends to suppress the difference frequency output.

The intermodulation distortion observed in the CCIF method are indicative of the non-linear distortion present in the system in the frequency range f_a to f_b. In particular, the CCIF method is able to measure the result of the non-linear distortion present at the upper frequency limit of the amplifier, where the high response starts to fall off.

The observed value of the intermodulation distortion is substantially independent of the exact values of f_a and f_b, and of the difference frequency $(f_b - f_a = f_d)$ of the two tests, provided this difference frequency is a small fraction of the test frequencies.

MEASURING FREQUENCY RESPONSE IN AUDIO AMPLIFIERS 16.9

The frequency response measurement is essential for filters, couplings circuits, amplifier stages and overall audio circuits. An amplifier stage is used as an example. A test setup for this amplifier stage is shown in Fig. 16.10.

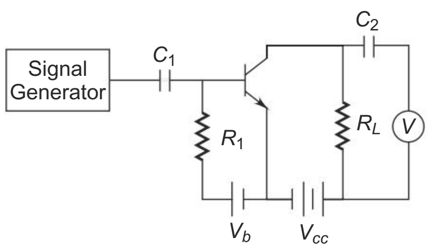


Fig. 16.10 Measuring AF response

This measurement is used to determine the band of frequencies that this stage can efficiently handle. Since it is an audio stage, a flat response is expected from about 50 Hz to about 20 kHz. A well designed stage may have a relatively flat response from 15 kHz to more than 50 kHz.

The signal generator is coupled to the input of the amplifier stage and the electronic voltmeter is connected across the output load resistor R_L. To obtain good results, the amplifier should be in its own operational circuit. If this is not possible, the test circuit should duplicate the operational circuit as closely as possible.

The signal generator is set to a low frequency sinusoidal wave output and the amplitude is set to a reference value at a reasonable audio frequency. This amplitude is held constant as the frequency is slowly increased.

The voltmeter is set on ac and gives no indication until the signal feeds through the amplifier. As the sinusoidal wave frequency approaches the audio band, a small voltage indication appears on the voltmeter. The voltage levels off to a steady reading as the frequency reaches a value where the amplifier has the maximum response (around about 20 Hz).

The constant voltage reading holds until the frequency approaches the top of the response band. As the frequency increases above this value, the voltage slowly drops off.

A plot of voltage and frequency, as shown in Fig. 16.11, gives the complete frequency response of this amplifier stage.

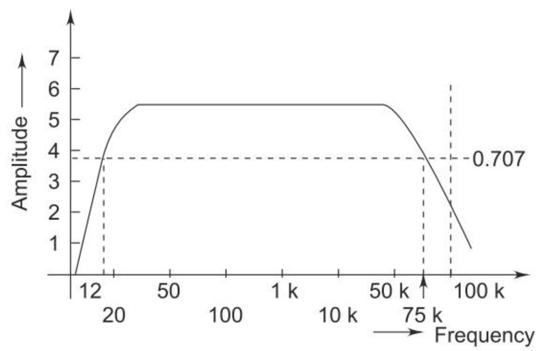


Fig. 16.11 Frequency response of an AF stage

The actual response of this stage is the same as the band pass, the frequencies being between the low and high half power points that is 0.707 times the voltage. The graph in Fig. 16.11 shows a frequency response from 12 Hz to 75 kHz and the top of the curve is reasonably flat.

MODULATION 16.10

Modulation is the process of impressing information on a carrier wave by altering either its amplitude of frequency. In amplitude modulation, the degree is the percentage of amplitude change from the unmodulated value of the carrier.

In frequency modulation, the frequency deviation is the change in frequency in one direction from the centre frequency of the carrier. The degree of modulation is important, because it has a direct bearing on output power and band width.

A carrier wave has 100% modulation when the total amplitude change from the crest to the trough is equal to twice the amplitude of the unmodulated carrier. 50% modulation is present when the amplitude of the trough is half the amplitude of the unmodulated carrier. When the amplitude of the trough of a wave reaches zero amplitude and remains there for a finite time period, the wave is over 100% modulated. These three conditions are illustrated in Fig. 16.12.

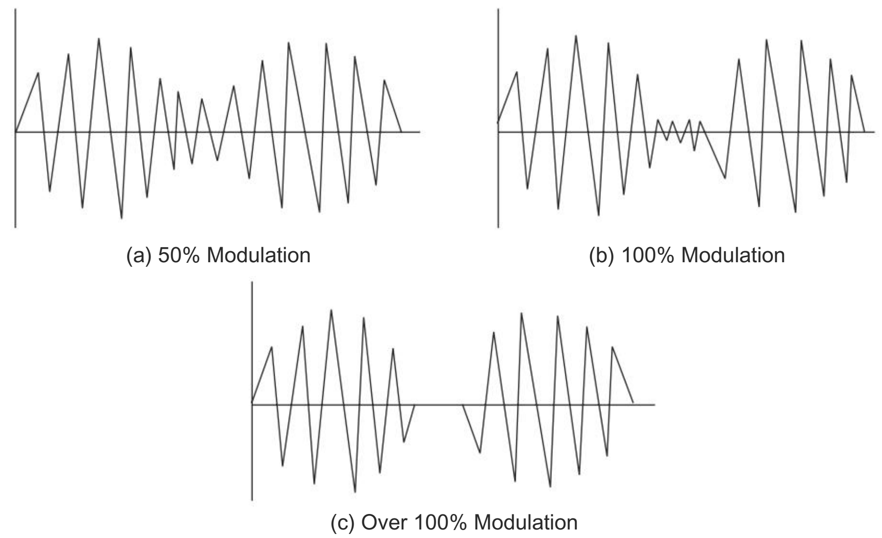


Fig. 16.12 Degree of modulation

The percentage modulation of a modulated carrier is calculated by dividing the change in amplitude by the amplitude of the unmodulated carrier, multiplied by 100.

One method of measuring the degree of modulation makes use of a wave component separation and comparison test circuit. This circuit rectifies and filters the modulated waves and compares the ac and dc components of the resulting waveform. This process is illustrated in Fig. 16.13.

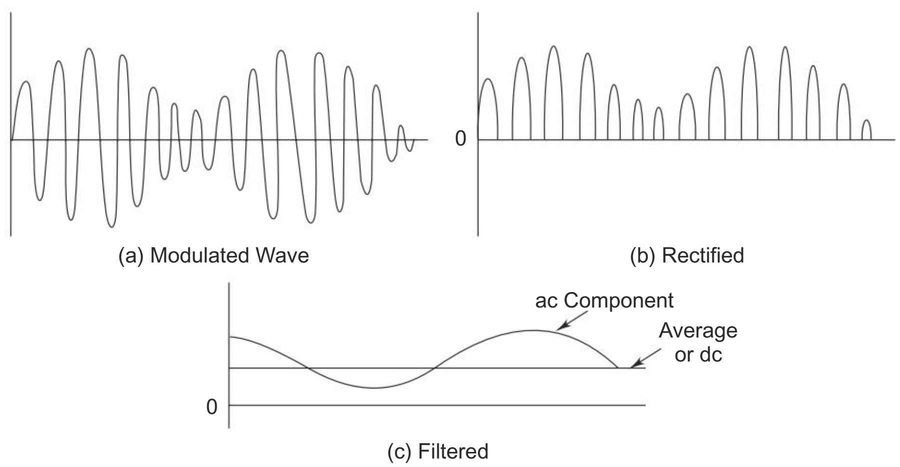


Fig. 16.13 Waveforms for measurement of the degree of modulation

The amplitude of the dc component is proportional to the carrier amplitude. The sine wave amplitude is proportional to the amplitude of the modulated signal.

A basic test circuit arrangement for measuring the degree of modulation is shown in Fig. 16.14.

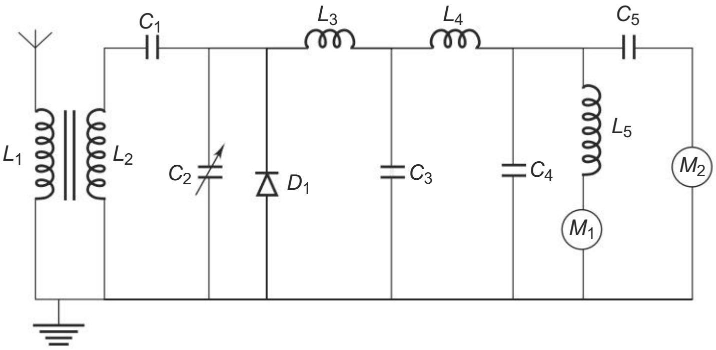


Fig. 16.14 Single rectifier test circuit

The modulated RF carrier is picked up on the antenna and rectified by the diode. The coils and capacitors filter out the RF components and leave an audio signal riding on the dc component. The meters indicate the values of the ac and dc components. M_1 is the dc meter, whose indication is proportional to the amplitude of the unmodulated carrier. M_2 is an ac meter whose indication is proportional to the amplitude of the modulating signal.

Since the ac meter reads effective values, 100% modulation causes the reading on M_2 to be 0.707 of the reading on M_1. This shows the modulating changes in amplitude is equal to twice the amplitude of the unmodulated carrier.

With 50% modulation, the carrier amplitude is equal to the modulating change in the amplitude; in this case the reading on meter M_1 is nearly thrice the value of M_2.

The circuit of Fig. 16.14 can be modified to provide a continuous reading directly in percentage of modulation, as illustrated in Fig. 16.15.

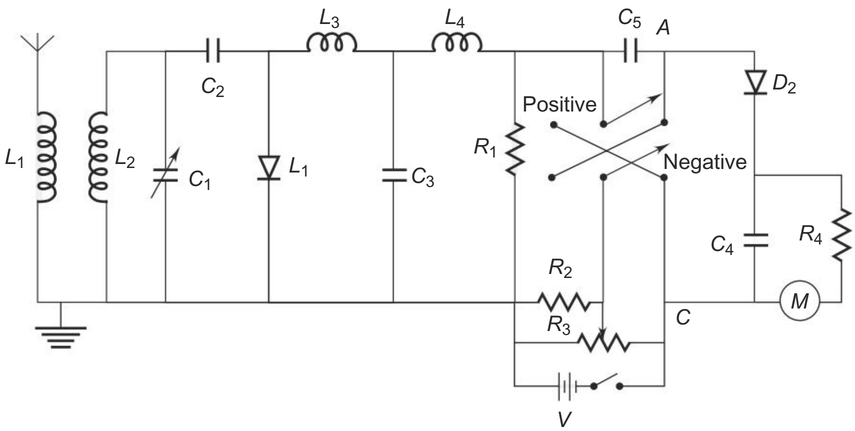


Fig. 16.15 Modulation meter

This circuit rectifies the difference in voltage between points *A* and *C*. This causes a current through the meter proportional to the percentage of modulation. The meter is calibrated directly in percentage modulation.

MEASURING FREQUENCY MODULATION 16.11

In FM modulation, the carrier amplitude remains constant and the modulated signal causes the carrier frequency to change.

Using the unmodulated carrier frequency as the reference, the modulated signal frequency varies about the reference frequency. The rate of deviation is the number of total changes per second, and is the same as the frequency of the modulating signal. The degree of deviation is the amount of frequency shift to either side of the reference frequency. The degree of deviation determines the band width of the transmitted signal and the number of sidebands it contains. In FM, all the intelligence is in the sidebands and it is desirable to retain as many sidebands as possible. The Federal Communication Commission allows a maximum deviation of 75 kHz on entertainment broadcast bands. This is narrow band FM transmission. Broad band FM transmission is allowed a much wider frequency deviation, up to several hundred MHz. The maximum allowable deviation is often referred to as 100% modulation.

MEASURING FREQUENCY DEVIATION WITH A RADIO RECEIVER 16.12

The simplest method of measuring frequency deviation utilises an FM receiver with a BFO as the measuring instrument. The only coupling used is the air waves between the transmitting and receiving antenna, as shown in Fig. 16.16.

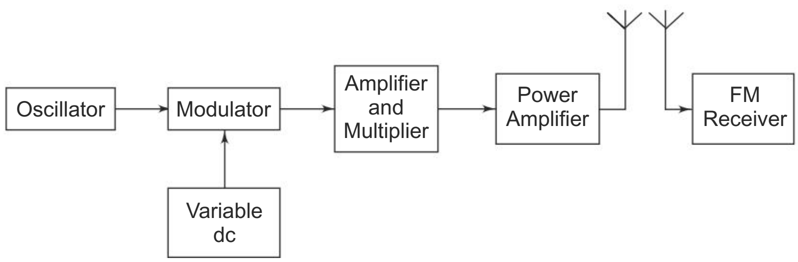


Fig. 16.16 Measuring frequency deviation

In this setup, the modulating signal is replaced by a variable dc voltage. Since the amplitude of the modulating signal determines the degree of deviation, the frequency is of no importance. The FM receiver is tuned to a frequency corresponding to either the minimum or maximum allowable excursion (swing). The dc voltage is then applied and slowly increased until a zero beat is obtained in the receiver. The amplitude of the dc voltage is now equal to the maximum allowable peak amplitude of the modulating signal.

(Since transmission of a carrier during tuning and testing operations is an undesirable practice, a dummy antenna should be replaced by a regular antenna.) This reduces the output to too low a value for proper reception, a pickup loop may be used to replace the receiving antenna. A suitable signal can then be obtained by loosely coupling the output of the power amplifier to the pickup loop.

Once the degree of deviation has been determined, the modulation index can be calculated by using the equation

$$M_i = \frac{f_d}{f_m} \tag{16.5}$$

where f_d = frequency deviation
f_m = frequency of the modulating signal

Example 16.1 *With a frequency deviation of ± 75 kHz and a modulation frequency of 5 kHz, calculate the modulation index.*

Solution

$$M_i = \frac{f_d}{f_m} = \frac{75\text{ kHz}}{5\text{ kHz}} = 15$$

(It can be seen from the modulation index table shown in Table 16.1 that for modulation index = 15.

(i) There are 20 effective sidebands above and below the centre frequency.
(ii) The band width is 40 times the modulation frequency, hence the band width is:

$$\text{BW} = 40 \times f = 40 \times 5\text{ kHz} = 200\text{ kHz}$$

Under these conditions the effective side bands are not transmitted. Side bands becomes ineffective when their amplitude drops to 1% of the amplitude of the modulation carrier.)

MEASURING AMPLITUDE MODULATION USING CRO 16.13

The CRO is widely used as an AM measuring instrument. It presents the waveform for visual monitoring and is fairly accurate in measurements of modulation percentage. The type of CRO pattern observed depends on how the CRO is connected. The various patterns obtained are (i) wave envelope (ii) trapezoidal, and (iii) a double ellipse.

16.13.1 Wave Envelope Pattern

The wave envelope pattern is obtained by using the CRO sawtooth sweep voltage on the H-plates and applying the RF signal to the vertical plates. The setup for this measurement is illustrated in Fig. 16.17.

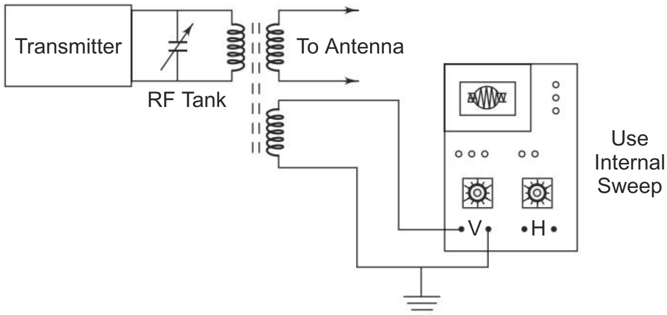


Fig. 16.17 Setup for obtaining wave envelope pattern

In order to obtain actual measurements, it is necessary to obtain a stationary pattern. This occurs only when the sweep frequency is the same as the modulation frequency or some sub-multiple of it. At 100% modulation, the peak of the modulating carrier is twice the amplitude of the unmodulated carrier; the minimum amplitude reaches zero but does not remain there for a significant period of time. This waveform is shown in Fig. 16.18.

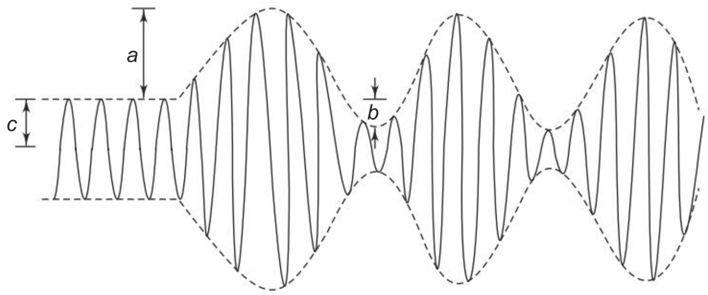


Fig. 16.18 Carrier 100% modulation

The modulating signal in this case is a symmetrical sine wave. With 100% modulation, amplitudes *a*, *b* and *c* are all equal. With any symmetrical wave, the percentage of modulation is equal to either *a* or *b* divided by *c*. If the modulating wave is not symmetrical, *a* and *b* will be of different amplitudes. In this case, the larger of the two amplitudes is divided by *c* to obtain the percentage of modulation.

16.13.2 Trapezoidal Pattern

Trapezoidal patterns may be obtained using the modulating signal on the horizontal deflecting plates and a modulated carrier wave on the vertical plates.The basic setup and the possible patterns are shown in Fig. 16.19 and Fig. 16.20.

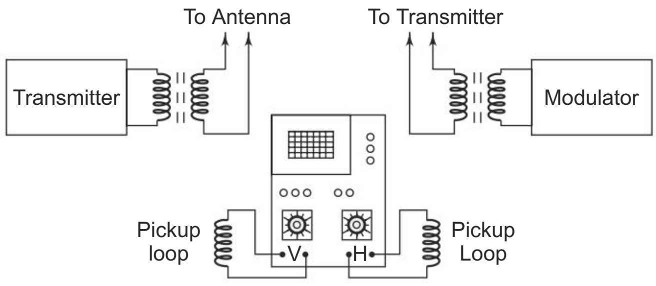


Fig. 16.19 Oscilloscope setup for obtaining trapezoidal pattern

The pattern shown in Fig. 16.20, is the unmodulated carrier and the carrier with a different percentage of modulation. The amplitude of modulated carrier (a) is a reference to illustrate amplitude changes due to the modulating signal. The narrow end (b) of the trapezoidal pattern represents the modulation troughs, and the wide end (c) represents the modulation peaks. These amplitude need not be reduced to voltage measurements; the number of divisions of the deflection is sufficient for calculations. The percentage modulation may be calculated as

$$\text{Modulation \%} = \left(\frac{c-b}{c+b}\right) \times 100 \tag{16.6}$$

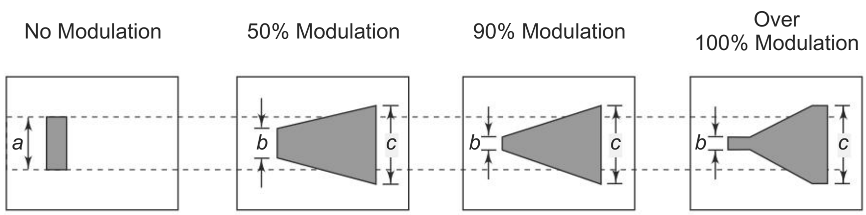


Fig. 16.20 Different trapezoidal patterns

With 50% modulation, the ratio between minimum and maximum amplitudes is 1:3, regardless of the divisions used.

$$\therefore \qquad \text{\% Modulation} = \left(\frac{c-b}{c+b}\right) \times 100$$

$$= \frac{3-1}{3+1} \times 100 = 50\%$$

In the case of 100% modulation, the narrow end of the trapezoidal pattern flattens out, becomes elongated and grows brighter than the remainder of the pattern. This indicates that the minimum amplitude is going all the way down to 0 and remaining their for a finite period of time. If the maximum amplitude is checked against the unmodulated carrier amplitude under these conditions, it will be more than twice the amplitude of the unmodulated carrier.

16.13.3 Double Ellipse Method

A double ellipse pattern for measuring amplitude modulation can be obtained by connecting the oscilloscope into the receiver circuit, as shown in Fig. 16.21(a).

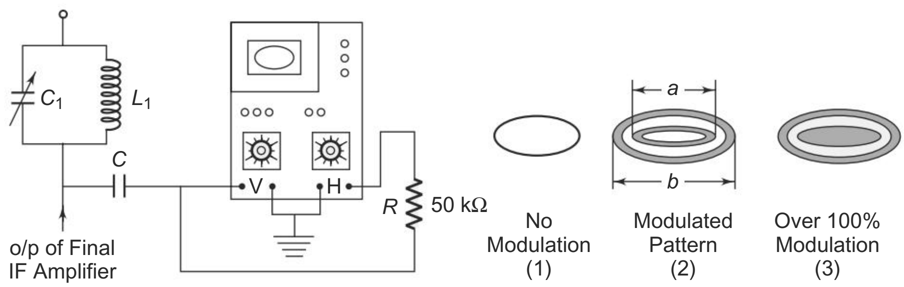


Fig. 16.21 (a) Setup for obtaining ellipse pattern (b) Obtaining double ellipse pattern

An ellipse consisting of a single sharply defined line is produced by the unmodulated carrier. The shape is due to the phase difference between the signals applied to the H and V deflection plates. The phase shift is obtained by the 50 kΩ resistor, along with the input capacitor in the Horizontal amplifier of the oscilloscope.

When the carrier is modulated, the pattern becomes a double ellipse in a ribbon pattern. Any degree of modulation less than 100% will produce a blank spot inside the centre ellipse. As the percentage of modulation increases, the centre blank space decreases, and it completely disappears at 100% modulation. Over 100% modulation causes a bright spot to appear in the centre ellipse.

Distance a in Fig. 16.21(b), represents the minimum amplitude, while distance b represents the maximum amplitude.

% modulation is calculated as follows:

$$\% \text{ Modulation} = \frac{b-a}{b+a} \times 100 \tag{16.7}$$

The ratio of a to b in the figure is 1 : 2, which represents about 33% modulation.

Table 16.1 Modulation Index Table

Modulation Index	*Number of Effective*	*Sidebands*	*Bandwidth*
	Above Carrier	*Below Carrier*	*f = Modulating frequency*
0.1	1	1	$2f$
0.2	1	1	$2f$
0.3	1	1	$2f$

(Contd.)

Table 16.1 (Contd.)

Modulation Index	*Number of Effective*	*Sidebands*	*Bandwidth*
	Above Carrier	*Below Carrier*	*f = Modulating frequency*
0.4	1	1	$2f$
0.5	2	2	$4f$
1	3	3	$6f$
2	4	4	$8f$
3	6	6	$12f$
4	7	7	$14f$
5	8	8	$16f$
6	9	9	$18f$
7	10	10	$20f$
8	12	12	$24f$
9	13	13	$26f$
10	14	14	$28f$
11	16	16	$32f$
12	17	17	$34f$
13	18	18	$36f$
14	19	19	$38f$
15	20	20	$40f$

Review Questions

1. How can measurements at microwave frequency be done?
2. What do you understand by a wavemeter. State the different types of wavemeters.
3. State the difference between absorption and transmission-type wavemeters.
4. Explain in brief absorption type wavemeter.
5. Describe briefly reaction type wavemeter.
6. Explain briefly transmission type wavemeter.
7. What do you understand by a coaxial wavemeter.
8. Explain with a diagram the working of a coaxial wavemeter.
9. What do you understand by a cavity wavemeter?
10. Explain with a diagram the operation of a simple cylindrical cavity wavemeter.
11. Compare resonant coaxial wavemeter and cavity wavemeter.
12. What do you understand by field strength?
13. State the two general methods for measurements of field strength.
14. Explain the standard antenna method for determination of field strength.
15. Explain with a diagram the operation of a standard field generator method or substitution method.
16. Define sensitivity and selectivity of a radio receiver.

17. Explain the procedure of measurement of sensitivity.
18. Explain with a diagram the sensitivity curve.
19. State the procedure of measurement of selectivity of a radio receiver.
20. What do you understand from the selectivity curve?
21. What is intermodulation distortion? State the working principle of intermodulation.
22. Explain with a graph the procedure to determine distortion using intermodulation method.
23. Explain with a block diagram the set-up and working of an SMPTE method to determine intermodulation.
24. Explain with a block diagram and frequency response the working of CCIF method to measure intermodulation.
25. Compare SMPTE method and CCIF methods of measurement of intermodulation.
26. Explain with a diagram the procedure of how frequency response of an audio amplifier can be measured.
27. Comment on the frequency response of an audio amplifier.
28. Define modulation. State the different types of modulation.
29. What do you understand by degree of modulation? Define percentage modulation of a modulated carrier.
30. Describe with a diagram and waveform how is degree of modulation of an AM receiver measured.
31. Explain with a diagram the operation of a modulation meter.
32. Briefly explain FM modulation.
33. Describe with a block diagram how frequency deviation can be measured.
34. State the different methods by which amplitude modulation can be measured using a CRO.
35. Explain with set-up and waveform the wave envelope pattern method to measure percentage modulation of AM.
36. Describe the set-up and the possible trapezoidal patterns used for measuring percentage modulation of AM.
37. Describe the set-up and double ellipse patterns obtained for measuring percentage modulation of AM.
38. Compare wave envelope method, trapezoidal pattern and double ellipse method.
39. Comment on which is best suitable for percentage modulation. Why.

Multiple Choice Questions

1. A wavemeter is used to measure
   (a) frequency (b) power
   (c) voltage (d) current
2. A wavemeter uses the principle based on
   (a) resonant circuit
   (b) amplifier circuit
   (c) attenuation circuit
   (d) non-resonant circuit
3. At lower frequencies the coaxial line is adjusted exactly to
   (a) λ (b) $\lambda/2$
   (c) $\lambda/4$ (d) $3\lambda/2$
4. The Q factor that the resonator uses in coaxial wavemeter is
   (a) very low (b) very high
   (c) medium
   (d) none of the above
5. Cavity wavemeters are used at
   (a) AF (b) RF
   (c) microwave frequency
   (d) VHF

6. Practical cavity wavemeters are in the form of
   (a) square (b) rectangular
   (c) cylinder (d) pyramidal
7. Precision of the cavity wavemeter can be obtained in the order of
   (a) 1 part in 100
   (b) 1 part in 1000
   (c) 1 part in 10000
   (d) 1 part in 100000
8. At frequencies above 300 MHz, it is more practical to deal with
   (a) temperature (b) power
   (c) voltage (d) current
9. Sensitivit of a radio receiver is defined in terms of
   (a) impedance (b) power
   (c) voltage (d) current
10. SMPTE method of measurement of nonlinear distortion consists of simultaneously applying
    (a) two sinusoidal voltages of different frequency
    (b) two sinusoidal voltages of same frequency
    (c) two sinusoidal voltages of different amplitude
    (d) two sinusoidal voltages of different phase
11. The method of determining nonlinear distortion is widely used in
    (a) RF (b) VHF
    (c) AF (d) VHF
12. In CCIF method of measurement of intermodulation, the two test signals used have
    (a) equal amplitude
    (b) equal frequencies
    (c) different phases
    (d) different amplitudes
13. In the SMPTE method, the normal practice is to make low frequency test signal
    (a) 4 times the HF test signal
    (b) 2 times the HF test signal
    (c) 10 times the HF test signal
    (d) equal to the HF test signal
14. The difference frequency between the two-test frequencies is generally in the range
    (a) 50 Hz – 200 Hz
    (b) 500 Hz – 2 kHz
    (c) 5 kHz – 20 kHz
    (d) 50 kHz – 200 kHz

Further Reading

1. F.E. Terman and J.M. Petit, *Electronic Measurements*, 2nd edition, McGraw-Hill, New York.
2. Vestor Robinson, *Hand Book of Electronic Instrumentation*, *Testing and Troubleshooting*, Reston Publishing, 1979.
3. George Kennedy, *Electronic Communication Systems*, McGraw-Hill, 1987.
4. D. Roddy and J. Collen, *Electronic Communication*, 3rd edition, Prentice-Hall of India, 1984.

Data Acquisition System (DAS)

Chapter 17

INTRODUCTION 17.1

A typical data acquisition system consists of individual sensors with the necessary signal conditioning, data conversion, data processing, multiplexing, data handling and associated transmission, storage and display systems.

In order to optimise the characteristics of the system in terms of performance, handling capacity and cost, the relevant sub systems can be combined together. Analog data is generally acquired and converted into digital form for the purpose of processing, transmission, display and storage.

Processing may consist of a large variety of operations, ranging from simple comparison to complicated mathematical manipulations. It can be for such purposes as collecting information (averages, statistics), converting the data into a useful form (e.g., calculations of efficiency of motor speed, torque and power input developed), using data for controlling a process, performing repeated calculations to separate signals buried in the noise, generating information for display, and various other purposes.

Data may be transmitted over long distances (from one point to another) or short distances (from test centre to a nearby PC).

The data may be displayed on a digital panel or on a CRT. The same be stored temporarily (for immediate use) or permanently for ready reference later.

Data acquisition generally relates to the process of collecting the input data in digital form as rapidly, accurately, and economically as necessary. The basic instrumentation used may be a DPM with digital outputs, a shaft digitiser, or a sophisticated high speed resolution device.

To match the input requirements with the output of the sensor, some form of scaling and offsetting is necessary, and this is achieved by the use of amplifier/attenuators.

For converting analog information from more than one source, either additional transducers or multiplexers are employed. To increase the speed with which information is accurately converted, sample-hold circuits are used. (In some cases, for analog signals with extra-wide range, logarithmic con-version is used.)

A schematic block diagram of a General Data Acquisition System (DAS) is shown in Fig. 17.1.

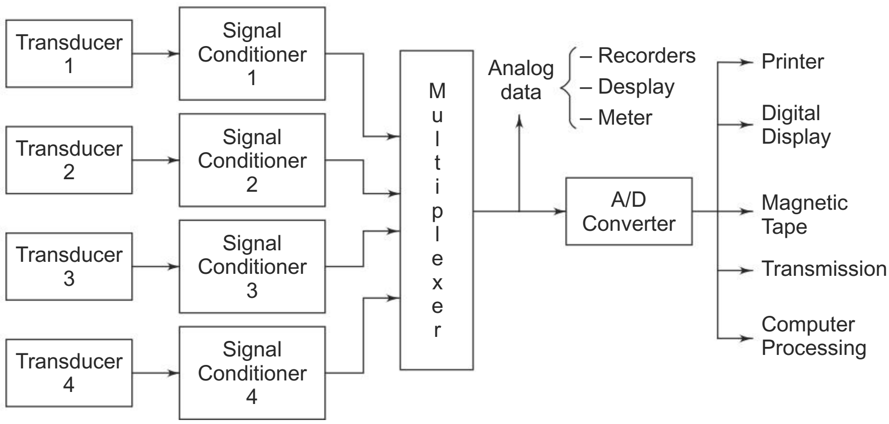


Fig. 17.1 Generalised data acquisition system

The characteristics of the data acquisition system depend on both the properties of the analog data and on the processing carried out.

Based on the environment, a broad classification divides the DAS into two categories.

1. Those suitable for favourable environments (minimum RF interference and electromagnetic induction)
2. Those intended for hostile environments

The former category may include, among other, laboratory instrument applications, test systems for collecting long term drift information on zeners, high calibration test instruments, and routine measurements in research, as mass spectrometers and lock-in amplifiers. In these, the systems are designed to perform tasks oriented more towards making sensitive measurements than to problems of protecting the integrity of analog data.

The second category specifically includes measure, protecting the integrity of the analog data under hostile conditions. Such measurement conditions arise in aircraft control systems, turbovisous in electrical power systems, and in industrial process control systems.

Most of these hostile measurement conditions require devices capable of a wide range of temperature operations, excellent shielding, redundant paths for critical measurements and considerable processing of the digital data.

On the other hand, laboratory measurements are performed over a narrow temperature range with much less electrical noise, employing high sensitivity and precision devices for higher accuracies and resolution.

The important factors that decide the configuration and sub systems of the data acquisition system are as follows.

1. Accuracy and resolution
2. Number of channels to be monitored

3. Analog or digital signal
4. Single channel or multichannel
5. Sampling rate per channel
6. Signal conditioning requirements of each channel
7. Cost

The various general configurations include the following.

1. Single channel possibilities
   (i) Direct conversion
   (ii) Pre-amplification and direct conversion
   (iii) Sample and hold, and conversion
   (iv) Pre-amplification, signal conditioning and any of the above
2. Multi channel possibilities
   (i) Multiplexing the outputs of single channel converters
   (ii) Multiplexing the output of sample-hold circuits
   (iii) Multiplexing the inputs of sample-hold circuits
   (iv) Multiplexing low level data

OBJECTIVE OF A DAS 17.2

1. It must acquire the necessary data, at correct speed and at the correct time.
2. Use of all data efficiently to inform the operator about the state of the plant.
3. It must monitor the complete plant operation to maintain on-line optimum and safe operations.
4. It must provide an effective human communication system and be able to identify problem areas, thereby minimising unit availability and maximising unit through point at minimum cost.
5. It must be able to collect, summarise and store data for diagnosis of operation and record purpose.
6. It must be able to compute unit performance indices using on-line, real-time data.
7. It must be flexible and capable of being expanded for future requirements.
8. It must be reliable, and not have a down time greater than 0.1%.

SIGNAL CONDITIONING OF THE INPUTS 17.3

Since all the data that have to be acquired, do not generally originate from identical sources, signal conditioning becomes necessary in some cases.

A simple attenuator, is used to scale down the input gains, this is to match the input signal level to the converter' full scale range.

Linearisation of the data, for example from the thermocouple Wheatstone's bridge, is performed by analog techniques using either linear approximation, or smooth series approximation using a low cost IC amplifier.

Alternately linear approximation can be performed digitally after data acquisition and conversion by the use of ROMs (storing a suitable linearisation table or programme initially).

Analog differentiation, precision rectification and averaging, phase detec-tion, logarithmic conversion, ratio computation using dividers and many other types of processors are used, before DAS.

Two methods of signal conditioning which are particularly applicable with advantage to data acquisition are (i) ratiometric conversion, and (ii) logarithm conversion.

17.3.1 Ratiometric Conversion

Consider a transducer using four strain gauges in a Wheatstone's bridge. The output voltage is a function of the change in resistance of the strain gauge in each of the four arms, and the excitation voltage of the bridge.

If the strain gauges are under maximum but constant unbalance and if the excitation varies by ± x%, the output of the bridge also varies by ± x%. However, if the bridge output is conditioned in such a way that the output of the signal amplifier is a voltage proportional to the strain only and independent of the excitation voltage, the system accuracy improves, since variations in the excitation voltage do not affect the sensitivity of the system.

The analog method of achieving this is to incorporate an analog divider to which the amplifier output and excitation voltage are fed, so that the output voltage of the divider is the ratio of the amplifier output voltage to the excitation voltage.

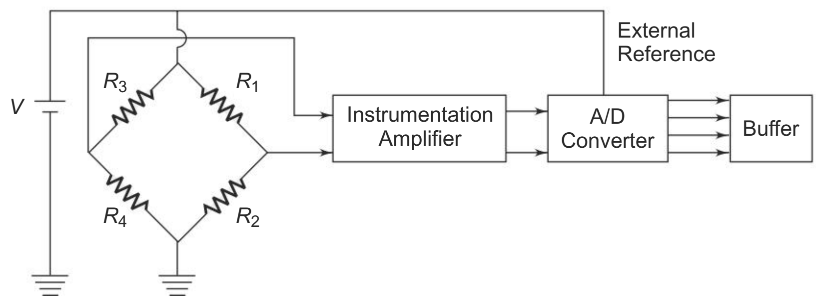


Fig. 17.2 Ratiometric conversion

One method, as shown in Fig. 17.2, is to feed the bridge excitation voltage as an external reference voltage to the analog/digital converter, in which the conversion factor is proportional to the reference voltage. The system sensitivity is therefore independent of the excitation voltage.

17.3.2 Logarithm Compression (Conversion)

A logarithm compression circuit enables the measurement of a fractional change in the input as a percentage of the input magnitude rather than a percentage of a range.

For example, for an input in the rang of 100 μV to 100 mV, the output voltage may correspond to 0 for 100 μV and 3 V for 100 mV, if the logarithm conversion gain is 1% per decade.

Consider now a change of 1%, i.e. the input changes from 100 mV to 101 mV. The output of the logarithm amplifier would change by

$$\Delta V = (\log 1.01) \times 1 \text{ V} = 4.3 \text{ mV}$$

where 1.01 is the ratio of the inputs, 101 mV/100 mV.

Since the output change is related to the ratio of the input, it is evident that the change in output is the same, i.e. 4.3 mV, whether the input changes from 10.0 mV to 10.1 mV or from 100 μV to 101 μV.

If the logarithm amplifier output is converted into digital output using a 12 bit BCD converter, the resolution of the converter would be 3 V/1000 = 3 mV for a 3 V full scale, provided the output of the logarithm amplifier is scaled properly.

With this resolution of the converter, it is possible to monitor and record changes as low as 1 μV for an input of 100 μV or 10 μV for 1 mV. If no logarithm amplifier had been used, the resolution would have been 100 μV (100 mV/1000 = 100 μV). Hence a 110 to 1 improvement is possible using a logarithm amplifier.

SINGLE CHANNEL DATA ACQUISITION SYSTEM 17.4

A single channel data acquisition system consists of a signal conditioner followed by an analog to digital (A/D) converter, performing repetitive conversions at a free running, internally determined rate. The outputs are in digital code words including over range indication, polarity information and a status output to indicate when the output digits are valid.

A single channel DAS is shown in Fig. 17.3. The digital outputs are further fed to a storage or printout device, or to a digital computer device, or to a digital computer for analysis. The popular Digital panel Meter (DPM) is a well known example of this. However, there are two major drawbacks in using it as a DAS.

1. It is slow and the BCD has to be changed into binary coding, if the output is to be processed by digital equipment.
2. While it is free running, the data from the A/D converter is transferred to the interface register at a rate determined by the DPM itself, rather than commands beginning from the external interface.

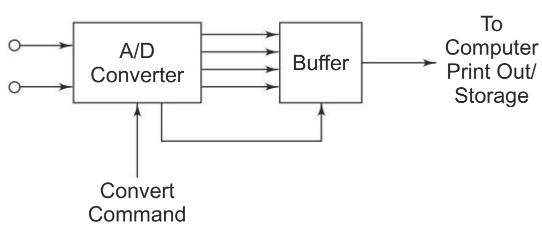


Fig. 17.3 Single channel DAS

17.4.1 Analog to Digital Converters (A/D)

Analog to digital converters used for DAS applications are usually designed to receive external commands to convert and hold. For dc and low frequency signals, a dual slope type converter is often used. The advantage is that it has a linear averaging capability and has a null response for frequencies harmonically related to the integrating period.

(Generally, the integrating time is selected equal to the period of the line frequency, since a major portion of the system interference occurs at this frequency and its harmonics.)

A/D converters based on dual slope techniques are useful for conversion of low frequency data, such as from thermocouples, especially in the presence of noise. The most popular type of converter for data system applications is the successive approximation type (refer to Ch. 5), since it is capable of high resolution and high speed at moderate cost. (For a conversion time of 10 μS, the maximum dv/dt for full scale and 0.1% resolution is about 1 V/ms, which is a considerable improvement.)

Higher speeds are obtained by preceding the A/D converter by a sample hold (S/H). The sample hold is particularly required with successive approximation type A/D converters, since at higher rates of input change the latter generates substantial non-linearity errors because it cannot tolerate changes during the conversion process.

Direct digital conversion carried out near the signal source is very advantageous in cases where data needs to be transmitted through a noisy environ-ment. Even with a high level signal of 10 V, an 8 bit converter (1/256 resolution) can produce 1 bit ambiguity when affected by noise of the order of 40 mV.

17.4.2 Pre-amplification and Filtering

Many low resolution (8/10 bit) A/D converters are constructed with a single ended input and have a normalised analog input range of the order of 5–10 V, bipolar or unipolar. For signal levels which are low compared to input requirements, amplification may be used in order to bring up the level of the input to match converter input requirements, so that optimum use can be made in terms of accuracy and resolution. The amplifier used has a single ended input or a differential input, as shown in Fig. 17.4.

If the signal levels are below a tenth of an mV, or when resolution of 14 bits or 16 bits is needed, the use of differential amplifiers can become a necessity.

When differential output has to be handled from a bridge network, instrumentation amplifiers are employed.

The accuracy, linearity and gain stability specifications should be carefully considered, to ensure the system is not affected by any limitations.

If the input signals are to be physically isolated from the system, the conductive paths are broken by using a transformer coupled or an optocoupled isolation amplifier. These techniques are advantageous in handling signals from high voltage sources and transmission towers. In biomedical applications such isolation becomes essential.

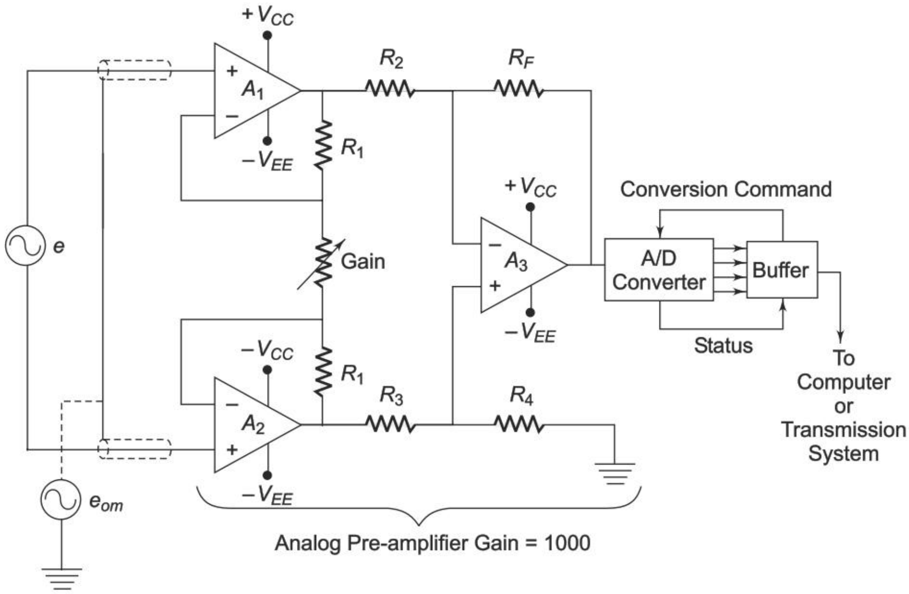


Fig. 17.4 DAS with pre-amplification

Pre-amplifiers can be coupled with active filters before processing of data, in order to minimise the effect of noise carriers and interfering high frequency components. They effective compensate for transmission sensitivity loss at high frequency and hence enable measurements over an enhanced dynamic frequency range.

Special purpose filters, such as tracking filters, are used for preserving phase dependent data.

MULTI-CHANNEL DAS 17.5

The various sub-systems of the DAS can be time shared by two or more input sources. Depending on the desired properties of the multiplexed system, a number of techniques are employed for such time shared measurements.

17.5.1 Multi-Channel Analog Multiplexed System

The multi-channel DAS has a single A/D converter preceded by a multiplexer, as shown in Fig. 17.5.

The individual analog signals are applied directly or after amplification and/ or signal conditioning, whenever necessary, to the multiplexer. These are further converted to digital signals by the use of A/D converters, sequentially.

For the most efficient utilisation of time, the multiplexer is made to seek the next channel to be converted while the previous data stored in the sample/hold is converted to digital form.

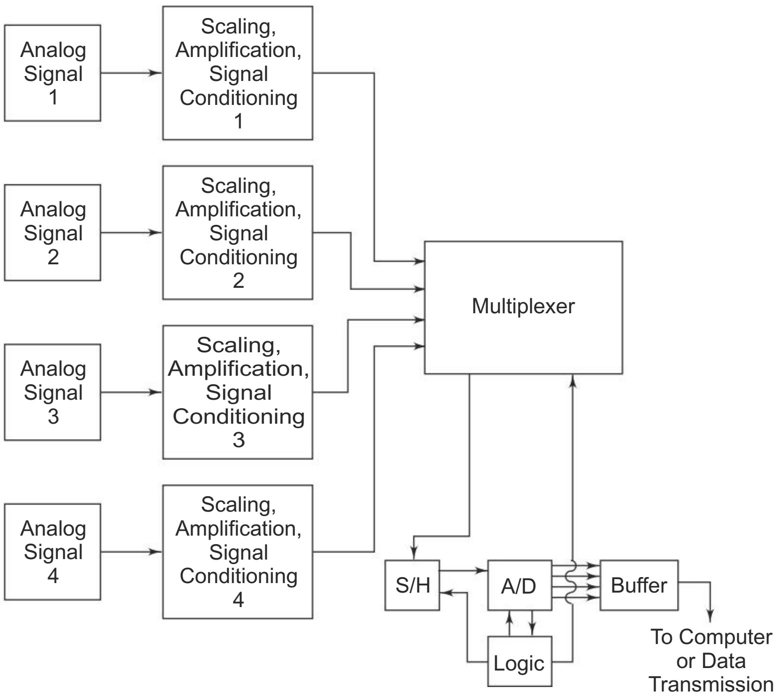


Fig. 17.5 Multi-channel DAS (A/D preceded by a multiplexer)

When the conversion is complete, the status line from the converter causes the sample/hold to return to the sample mode and acquires the signal of the next channel. On completion of acquisition, either immediately or upon command, the S/H is switched to the hold mode, a conversion begins again and the multiplexer selects the next channel. This method is relatively slower than systems where S/H outputs or even A/D converter outputs are multiplexed, but it has the obvious advantage of low cost due to sharing of a majority of sub-systems.

Sufficient accuracy in measurements can be achieved even without the S/H, in cases where signal variations are extremely slow.

17.5.2 Multiplexing the Outputs of Sample/Hold

When a large number of channels are to be monitored at the same time (synchronously) but at moderate speeds, the technique of multiplexing the outputs of the S/H is particularly attractive.

An individual S/H is assigned to each channel as shown in Fig. 17.6, and they are updated synchronously by a timing circuit.

The S/H outputs are connected to an A/D converter through a multiplexer, resulting in a sequential readout of the outputs.

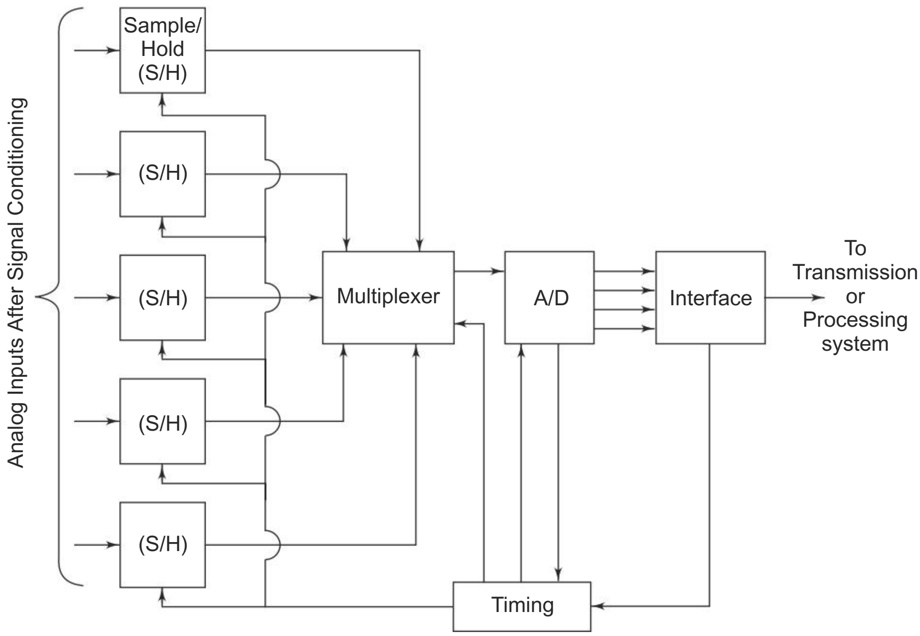


Fig. 17.6 Simultaneous Sampled System Multiplexer

(Applications that might require this approach include wind tunnel measurements, seismographic experimentation, radar and fire control systems. The event to be measured is often a one-shot phenomenon and information is required at a critical point during a one-shot event.)

17.5.3 Multiplexing After A/D Conversion

It is now economically feasible to employ an A/D converter for each analog input and multiplex the digital outputs.

Since each analog to digital converter (A/D) is assigned to an individual channel, the conversion rate of the A/D need only be as fast as is needed for that channel, compared to the higher rates that would be needed if it were used as in a multi channel analog multiplexed system.

The parallel conversion scheme shown in Fig. 17.7 provides additional advantages in industrial data acquisition systems where many strain gauges, thermocouples and LVDTs are distributed over large plant areas. Since the analog signals are digitised at the source, the digital transmission of the data to the data centre (from where it can go on to a communication channel) can provide enhanced immunity against line frequency and other ground loop interferences. The data converted to digital form is used to perform logic operations and decisions. Based on the relative speed at which changes occur in the data, the scanning rate can be increased or decreased.

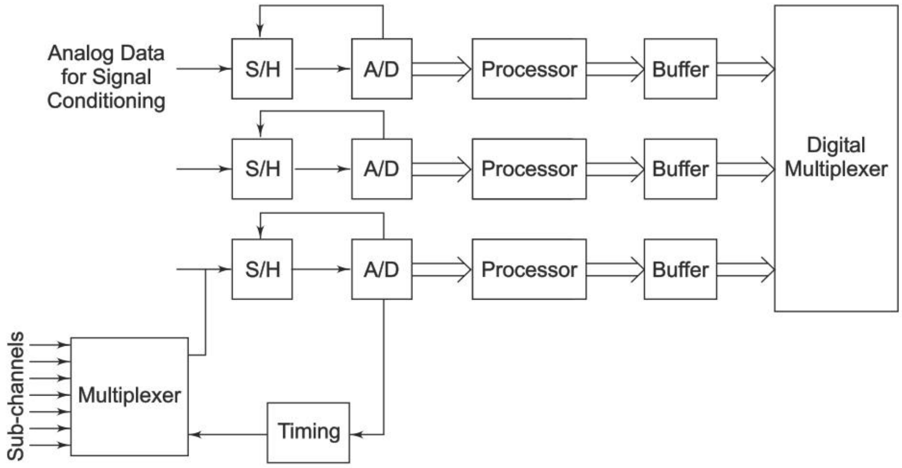


Fig. 17.7 Multi-channel DAS using digital multiplexing

Alternatively, input channels having slowly varying data can be pre-multiplexed in any of the forms suggested earlier, so that a set of sequentially multiplexed sub channels can then replace one channel of the main digital multiplexed system, as indicated in Fig. 17.7.

17.5.4 Multiplexing Low Level Data

A low level data multiplexing system, as shown in Fig. 17.8, enables the use of a single high quality data amplifier for handling multichannel low level inputs.

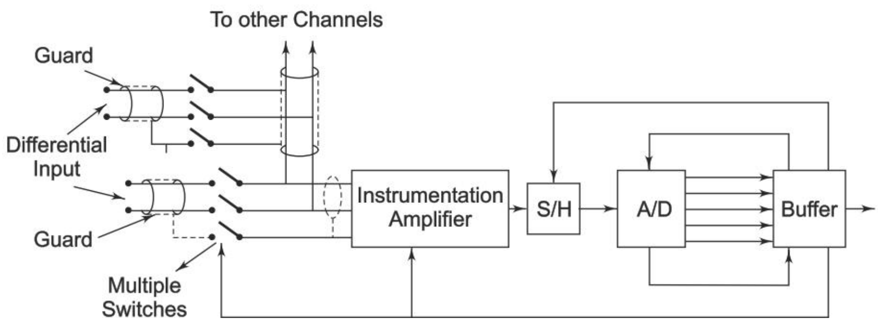


Fig. 17.8 Low level multiplexing

Individual amplifiers are used for each low level signal. Low level multiplexing can be attractive when a large number of channels (25), all having low level outputs, need to be used at moderate speeds. The use of individual channels is possible because of the availability of high quality amplifiers at moderate cost. (A typical application is a 200 channel stress measurement system in a transmission tower set up.)

Several factors have to be considered to accomplish low level multiplexing successfully. Guarding may have to be employed for every channel, and each individual guard may have to be switched, so that the appropriate guard is driven by the common mode pertaining to that channel.

Problems of pickup gets more complicated and have to be taken care of, to preempt the possibility of signal-to-signal, and even common mode-to differential mode signal cross-talk.

Capacitance balance may need to be carried out. When the number of channels to be multiplexed increases, the problems of stray capacitances and capacitive balance are worsened.

In the specific case of a 48 channel system, the input channels are subdivided into groups of eight channels in the first tier. Each of these six subgroups are in turn multiplexed by a six channel multiplexer on the second tier. The main advantage of using this is the reduction of capacitance effects.

COMPUTER BASED DAS 17.6

If a large number of inputs are to be measured, some equipment is needed to measure them and display the results in a meaningful and operationally useful fashion. All this is possible with DAS, which utilises a computer driven visual display unit (CRT) as an operator aid.

A screen display can be obtained within two seconds by pressing a button. Information may be displayed only when called up. The screen display can be designed in several ways, using a combination of graphical and numeric displays, so as to be of maximum utility to the operator.

DAS aids operate in the following manner.

1. Display information instantly in condensed, understandable and legible manner so that it can be easily assimilated.
2. Display spatial as well as time variation.
3. Display vital parameters grouped together logically and concisely, eliminating the need of looking at many scattered instruments.
4. Display CRT graphic displays of plant sub-systems.
5. Display short trends on a long and short term basis, as required.
6. Analyse the data and present the highest priority problem first, and display operator guidance messages.
7. Analyse the data and present the derived data; do performance calculations to depict the performance of several equipments and plants.
8. Display alarms, indicating abnormal plant operating conditions on the CRT.
9. Provide trending of analog variables on strip chart recorders, in the form of a histogram on the CRT, and provide dynamic updating of parameters.
10. Produce a hard copy record of all plant operating events and various plant logs.
11. Provide a recording of the sequence of events, whenever an emergency occurs.

DIGITAL TO ANALOG (D/A) AND ANALOG TO DIGITAL (A/D) CONVERTERS 17.7

Digital to analog and analog to digital conversion form two very important aspects of digital data processing.

D/A involves translating digital information into equivalent analog information. For example, the output of a digital system might be changed to analog form to drive a pen recorder. Similarly, a signal might be required for the servomotors which drive the arms of a plotter. D/A can also be considered as a decoding device, since it operates on the output of a digital system.

A/D converters are used for the reverse process of changing analog signals to equivalent binary signals. A/D might be used to change analog output signals from transducers (measuring temperature, pressure, vibration, etc.) into equivalent digital signals. An A/D is often referred to as an encoding device.

A D/A converter is usually an integral part of any A/D conversion.

17.7.1 Variable Resistor Network

The basic problem in converting a digital signal into an equivalent analog signal is to change the *n* digital voltage levels into one equivalent analog voltage. This can be achieved most easily by designing a resistive network which changes each of the digital levels into an equivalent binary weight voltage (or current).

To understand the meaning of equivalent binary weight consider the truth table for the 3 bit binary signal shown in Table 17.1.

Table 17.1 Truth Table for a 3-bit Binary

2^2	2^1	2^0
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Suppose we wish to change the 8 possible states of digital signals into equivalent analog voltages. The smallest number represented by 000 is 0V and the largest number represented is 111. Let us make this signal equal to +7V. This then establishes the range of the analog signal which will be developed. Now, between 000 and 111 there are seven discrete levels to be defined. Therefore, it is convenient to divide the analog signal into seven levels.

The smallest increment change in the digital signal is represented by the LSB (2^0). Hence we would like to have this bit cause a change in the analog output equal to 1/7 of the full scale analog output voltage. The resistive divider will then be designed such that a 1 in the 2^0 position causes $+ 7 \times 1/7 = 1$ V at the output.

Since $2^1 = 2$ and $2^0 = 1$, it can be seen that the 2^1 number is twice the size of the 2^0 bit. Therefore a 1 in the 2^1 bit position must cause a change in the analog output voltage which is twice the size of the LSB. The resistive divider must then be designed such that a 1 in the 2^1 bit position causes a change of $+ 7 \times 2/7 = +2$ V in the analog voltage.

Similarly, $2^2 = 4 = 2^1 \times 2 = 4 \times 2^0$ this 2^2 bit must cause a change in the output voltage which is 4 times that of the LSB. The 2^2 bit must then cause an output voltage change of $+ 7 \times 4/7 = + 4$ V.

The process can be continued and it will be seen that each successive bit must have a value which is twice that of the preceding bit. Thus the LSB is given a binary equivalent weight of 1/7, or 1 part in 7. The next LSB is given a weight of 2/7, which is twice the LSB or 2 parts in 7. The MSB (in the case of a 3 bit system) is given by 4/7, which is 4 times the LSB or 4 parts in 7.

The total sum of the weights must be equal to 1. Hence

$$1/7 + 2/7 + 4/7 = 7/7 = 1$$

In general, the binary equivalent weights assigned to the LSB is $\dfrac{1}{2^n - 1}$ where n is the number of bits.

A resistive divider having three digital inputs and an analog output is shown in Fig. 17.9.

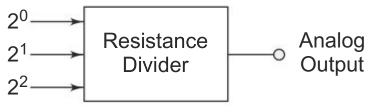


Fig. 17.9

In this case, assume digital levels 0 = 0 V and 1 = + 7 V.

For an input of 001, the output is + 1 V; similarly for 010, the output is + 2 V and for 100, the output is + 4 V. Now digital input of 011 is seen to be a combination of signals 001 and 010. If + 1 V from the 2^0 bit is added to + 2 V from the 2^1 bit, the output is + 3 V for an input of 011. Similarly, other levels are determined by an additive combination of voltages, as shown in Table 17.2.

Hence the resistive ladder must do two things in order to change the digital input into an equivalent analog output.

1. 2^0 bit must be changed to +1 V, 2^1 bit to +2 V and 2^2 bit to + 4 V.
2. The three voltages representing the digital bits must be summed together to form the analog output voltage.

A resistive ladder which performs the above functions is shown in Fig. 17.10.

Table 17.2 Analog Outputs Levels for a 3-bit Digital Input

Digital inputs	*Analog outputs*
000	0
001	+ 1V
010	+ 2V
011	+ 3V
100	+ 4V
101	+ 5V
110	+ 6V
111	+ 7V

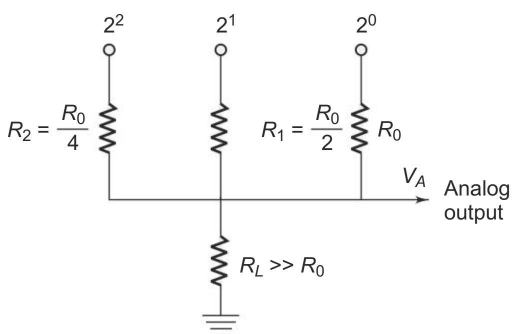


Fig. 17.10 Resistive ladder

The resistors R_0, R_1 and R_2 form the divider network. R_L is the load to which the divider is connected and is large enough not to load the divider network.

Assume that the digital input signal 001 is applied to this network. Using the levels as before, 0 = 0 V and 1 = +7 V. The equivalent circuit is shown in Fig. 17.11. R_L is considered very large, and hence neglected.

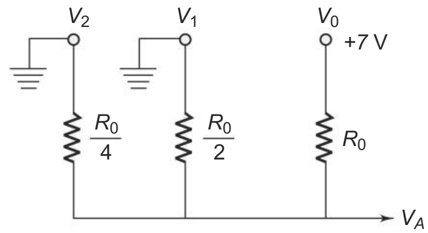


Fig. 17.11 Equivalent circuit for 001

The analog output voltage V_A can be easily determined by the use of Millman's theorem.

Millman's theorem states that the voltage appearing at any node in a resistive network is equal to the summation of the current entering the node (found by assuming the node voltage is 0) divided by the summation of the conductances connected to the node.

In equation form, Millman's theorem states

$$V = \frac{E_1/R_1 + E_2/R_2 + E_3/R_3 + \cdots}{1/R_1 + 1/R_2 + 1/R_3 + \cdots}$$

Applying Millman's theorem to Fig. 17.11

$$V_A = \frac{V_0/R_0 + V_1/(R_0/2) + V_2/(R_0/4)}{1/R_0 + 1/(R_0/2) + 1/(R_0/4)}$$

$$= \frac{7/R_0}{1/R_0 + 2/R_0 + 4/R_0} = \frac{7}{7} = +1 \text{ V}$$

In general, the following hold.

1. There is one input resistance for each digital bit.
2. Beginning with the LSB, each following resistor is half the previous resistance.
3. The full scale output voltage is equal to the +ve voltage of the digital input signal (negative voltages work equally well).
4. The change in output voltage due to a change in the LSB is equal to $V/(2^n - 1)$, where V is the digital input voltage level.
5. The LSB has a weight of $1/(2^n - 1)$, where n is the number of input bits.
6. The output voltage V_A can be found for any digital input signal by using the following modified form of Millman's theorem.

$$V_A = \frac{V_0 \cdot 2^0 + V_1 \cdot 2^1 + V_2 \cdot 2^2 + \cdots + V_{n-1} \cdot 2^{n-1}}{2^n - 1}$$

where V_0, V_1, V_2, ..., V_{n-1} are the digital input voltage levels (0 and + V) and n is the number of input bits.

The resistive network has two basic drawbacks.

1. Each resistance in the network has a different value. Since the dividers are usually constructed using precision resistors, their cost factor increases.
2. The resistance used in the MSB is required to handle a much larger current than the LSB resistor. (For example, the current through the MSB may be 500 times larger than that through the LSB.)

For these reasons, a second type of resistive network, called a ladder is used.

Example 17.1 *For a 5 bit resistive divider, determine the following*

1. *The weights assigned to the LSB.*
2. *The weights assigned to the 2nd and 3rd LSB.*
3. *The change in output voltage due to the change in the LSB, 2nd LSB and 3rd LSB.*
4. *The output voltage for a digital input of 11011 and 10110. (Assuming 0 = 0 V and 1 = + 10 V.)*

Solution

1. The LSB weight = $1/2^n - 1 = 1/2^5 - 1 = 1/31$.
2. The 2nd LSB weight is 2/31 and the 3rd LSB weight is 4/31.
3. The LSB causes a change in the output voltage of 10/31 V.
   The 2nd LSB causes a change in the output voltage of 20/31 V.
   The 3rd LSB causes a change in the output voltage of 40/31 V.
4. The output voltage for a digital input of 11011 is

$$V_A = \frac{10 \times 2^4 + 10 \times 2^3 + 0 \times 2^2 + 10 \times 2^1 + 10 \times 2^0}{2^5 - 1}$$

$$= \frac{160 + 80 + 0 + 20 + 10}{31} = \frac{270}{31} = 8.71 \text{ V}$$

The output voltage for a digital input of 10110 is

$$V_A = \frac{10 \times 2^4 + 0 \times 2^3 + 10 \times 2^2 + 10 \times 2^1 + 0 \times 2^0}{31}$$

$$= \frac{160 + 0 + 40 + 20 + 0}{31} = \frac{220}{31} = 7.09 \text{ V}$$

17.7.2 Ladder Type D/A Converter

Looking into any load from the left side we always see a resistor R. Similarly looking from the right side for any load, one always sees a resistor 2 R. This impedance phenomenon is the key to analyzing a D/A converter, as shown in Fig. 17.12.

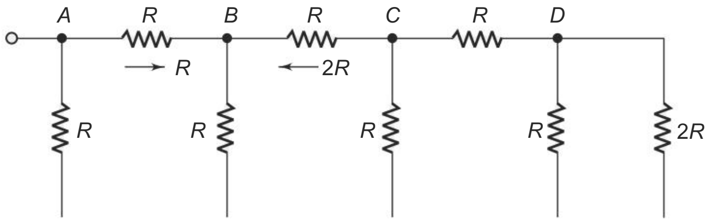


Fig. 17.12 *R*–2*R* Ladder Network

When the switch is to the right (B), as shown in Fig. 17.13, the upper ground is connected. When the switch is to the left (A), it steers the current to the lower ground which collects the output of the staircase.

$$\therefore \qquad I_{out} = (D_3 + 2^{-1} D_2 + 2^{-2} D_1 + 2^{-3} D_0) \times \frac{I_{ref}}{2}$$

The output of a 4 bit ladder is from 0 – $(15/16) I_{ref}$, and each increment is of $(1/16)I_{ref}$.

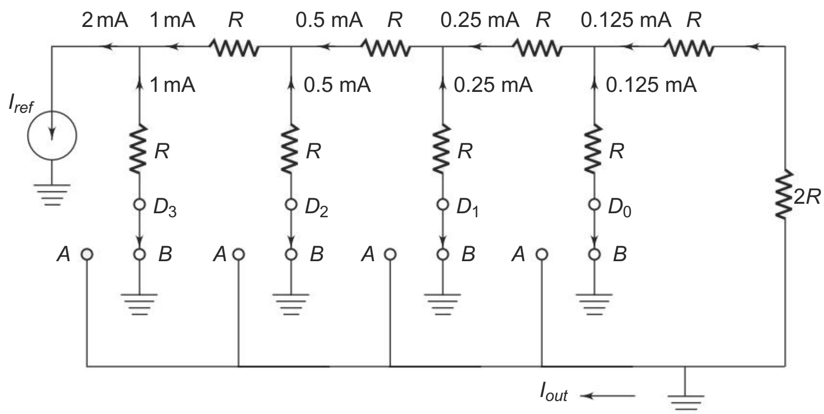


Fig. 17.13 *R*—2*R* ladder network

Similarly an 8 bit ladder produces a maximum output of $(255/256)I_{ref}$. The LSB increment is then $(1/255)I_{ref}$.

The maximum resistance used is always twice the minimum. At each node, there is a constant voltage or current; only the ground is changed. Due to this fact, stray capacitances have little effect and transients, that is exponential charge and discharge, are eliminated. This reduces the settling time and increases speed. Therefore, the ladder converter can be manufactured in bulk. Table 17.3 shows the various analog output levels for a 4-bit digital input.

Table 17.3 Analog Outputs for a 4-bit Digital Input

Digital Inputs				
D_3	D_2	D_1	D_0	*Output Voltage*
0	0	0	0	0
0	0	0	1	1/16
0	0	1	0	2/16
0	0	1	1	3/16
0	1	0	0	4/16
0	1	0	1	5/16
0	1	1	0	6/16
0	1	1	1	7/16
1	0	0	0	8/16
1	0	0	1	9/16
1	0	1	0	10/16
1	0	1	1	11/16
1	1	0	0	12/16
1	1	0	1	13/16
1	1	1	0	14/16
1	1	1	1	15/16

17.7.3 Binary Ladder

A binary ladder is constructed of resistors having only two values and thus overcomes the disadvantages of weighted resistors. The left end of the ladder is terminated in 2 R, as shown in Fig. 17.14.

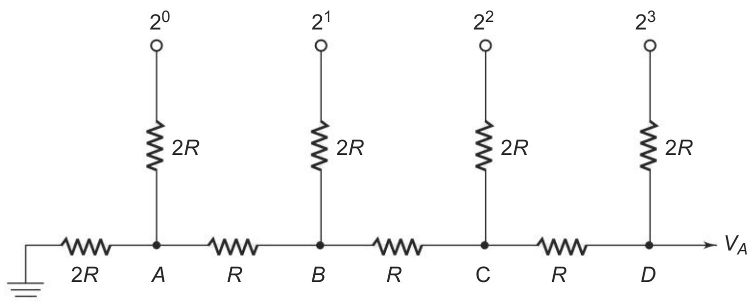


Fig. 17.14 Binary ladder

Assuming that all digital inputs are at ground, beginning at point A of Fig. 17.14, the total resistance looking into the terminating resistors is $2R$, as seen from Fig. 17.15(a). The total resistance looking out towards the 2^0 input is $2R$. These two resistors combine to form a value of R. At node B or at 2^1, the input is still $2R$, as seen from Fig. 17.15(b). It is clear from Fig. 17.15(c) that the resistance looking back towards node C is $2R$, as is the resistance looking at the 2^3 input.

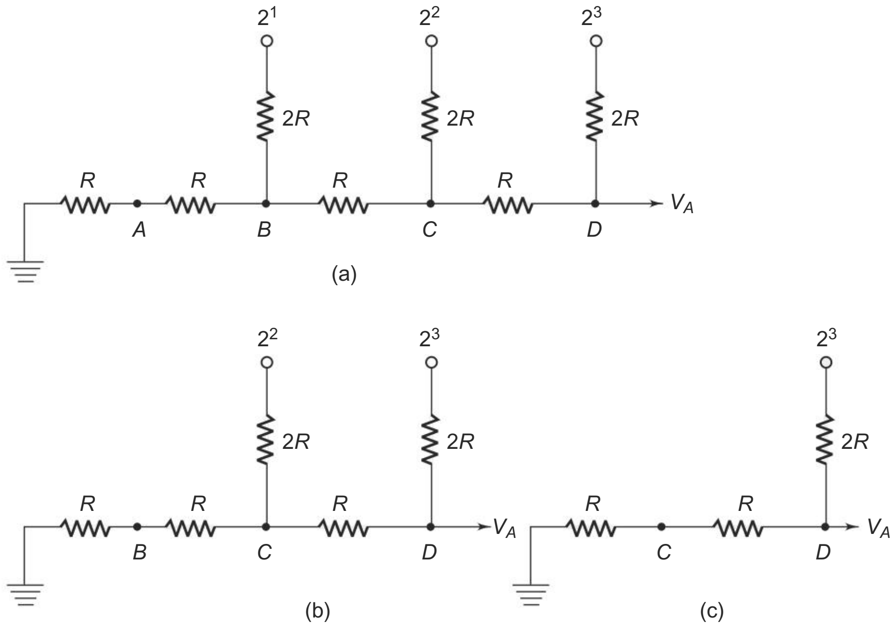


Fig. 17.15 (a) Resistance looking at point A with respect to ground
(b) Resistance looking at point B with respect to ground
(c) Resistance looking at point C with respect to ground

From this we conclude that the resistance looking back from any node towards the terminating resistance or out toward the digital input is $2R$. We can use this to determine the various digital inputs. First, assume that the digital input signal is 1000, as shown in Fig. 17.16.

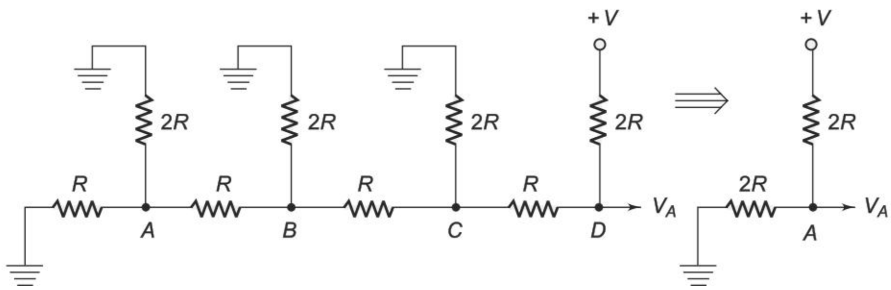


Fig. 17.16 Equivalent circuit for binary number 1000

Since there is no voltage source to the left of the node D, it can be replaced by $2R$.

therefore $$V_A = \frac{2\,R}{2\,R + 2\,R} \times V = \frac{+V}{2}$$

Hence a 1 in the MSB position provides an output voltage of $+V/2$.

To determine the output voltage due to the 2nd MSB, assume an input of 0100, as shown in Fig. 17.17(a).

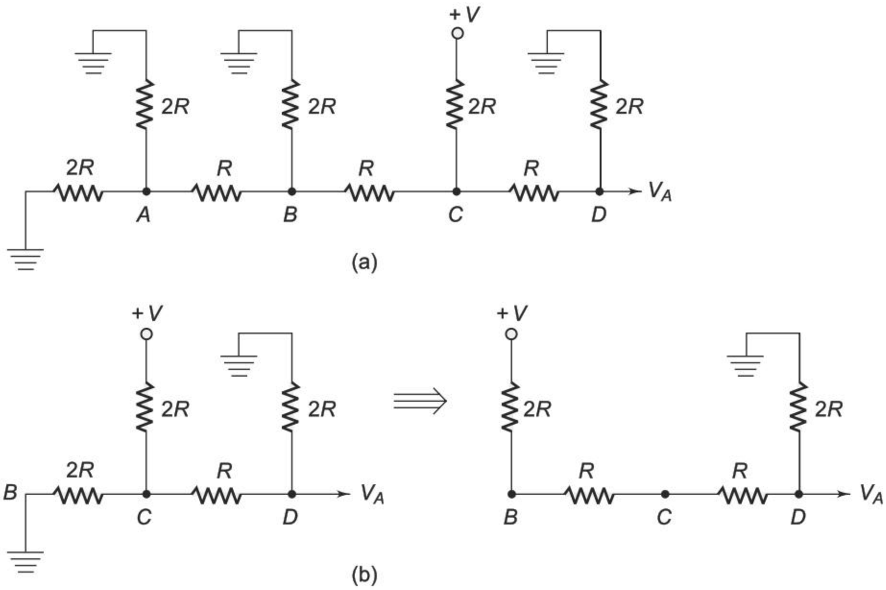


Fig. 17.17 (a) R–$2R$ ladder network for binary 0100
(b) Equivalent circuit for binary 0100

Therefore, the left side network of node C with its Thévenin equivalent, is clearly a resistance R in series with a voltage source of $+V/2$. The final equivalent circuit, the Thévenin's equivalent include, is

$$V_A = \frac{2\,R}{2\,R + 2\,R} \times \frac{+V}{2} = \frac{+V}{4}$$

Hence the 2nd MSB equals $+V/4$. This process can be continued, and it can be shown that the 3rd MSB gives an output voltage of $+V/8$, the 4th MSB gives an output voltage of $+V/16$ and so on.

The various output voltages for the corresponding MSB are given in Table 17.4.

Table 17.4 Various Output Voltage for Corresponding MSB

Bit Position	*Binary Weight*	*Output Voltage*
1st MSB	1/2	$V/2$
2nd MSB	1/4	$V/4$
3rd MSB	1/8	$V/8$
4th MSB	1/16	$V/16$
5th MSB	1/32	$V/32$
6th MSB	1/64	$V/64$
–	–	–
–	–	–
–	–	–
nth MSB	$1/2^n$	$V/2^n$

Example 17.2 *For a 5 bit ladder, if the input levels are 0 = 0 V and 1 = + 10 V. What are the output voltages for each bit?*

Solution

MSB	V_A	$= V/2$	$= 10/2$	$= 5$ V
2nd MSB	V_A	$= V/4$	$= 10/4$	$= 2.5$ V
3rd MSB	V_A	$= V/8$	$= 10/8$	$= 1.25$ V
4th MSB	V_A	$= V/16$	$= 10/16$	$= 0.625$ V
5th MSB	V_A	$= V/32$	$= 10/32$	$= 0.3125$ V

Since this ladder is composed of linear resistors, it is a linear network and the principle of superposition can be used. This means that the total voltage due to a combination of input digital levels can be found by simply taking the sum of the output voltage levels caused by each of the digital inputs.

Therefore, the output voltage is given by

$$V_A = \frac{V}{2} + \frac{V}{4} + \frac{V}{8} + \frac{V}{16} + \frac{V}{32} + \cdots + \frac{V}{2^n}$$

Where n is the total number of bits.

$$V_A = \frac{V_0 2^0 + V_1 2^1 + V_2 2^2 + V_3 2^3 + V_4 2^4 + \cdots + V_{n-1} \cdot 2^{n-1}}{2^{n-1}}$$

17.7.4 Digital to Analog Converters (D/A) (Using OpAmp)

The basic circuit of an operational amplifier used as an inverter is shown in Fig. 17.18(a). The input signal is applied to the inverting terminal of the opamp. The input impedance is almost infinity, and almost no current can enter the input terminals. The input voltage is fixed to almost zero at its input terminals; the upper terminal is at nearly the same potential as the ground terminal. Therefore, the upper terminal is called a virtual ground.

The difference between virtual ground and normal ground is zero, while sinking any amount of current. Virtual ground is for voltage, not for current. Therefore, virtual ground point is at 0 V, i.e. it sinks current.

As $\quad I_{in} = \dfrac{V_{in}}{R_{in}} \quad$ therefore $I_{in} \cong I_F$

Because no current can enter the virtual ground (because of ∞ impedance) all the current has to go through R_F with respect to the virtual ground. V_{out} is measured with respect to normal ground.

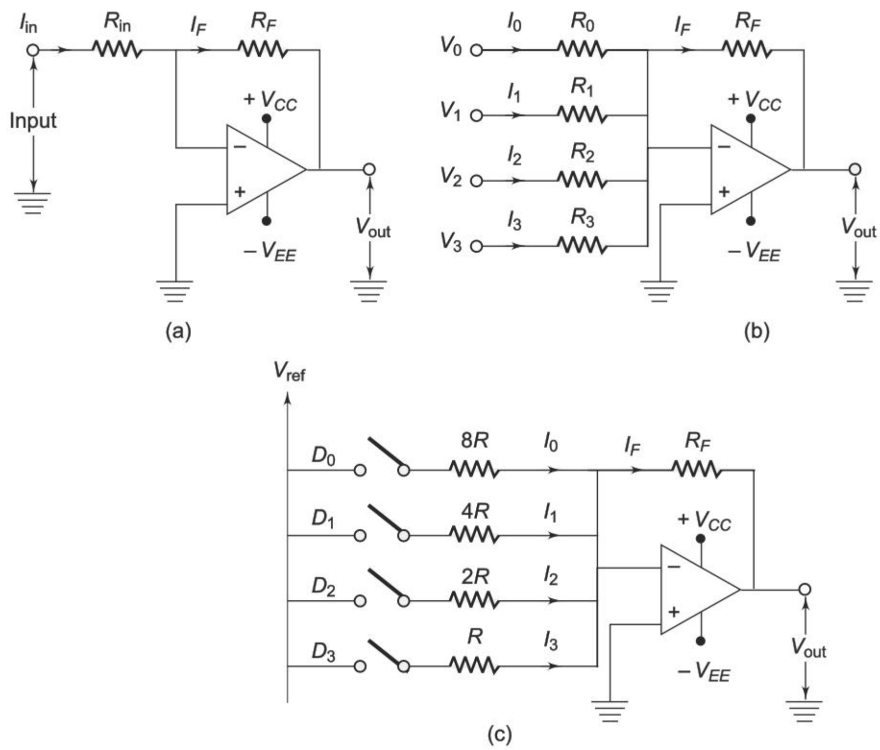


Fig. 17.18 (a) Basic Opamp circuit (b) Summing amplifier
(c) D/A using Opamp as a summing amplifier.

Example 17.3 *If V_{in} = 5 V, R_{in} = 2.5 k, R_F = 1 k, calculate the output voltage V_{out}.*

Solution

$$I_{in} = \frac{V_{in}}{R_{in}} = \frac{5}{2.5\text{ k}} = 2\text{ mA}$$

Therefore output voltage $V_{out} = -I_F \times R_F = -2\text{ mA} \times 1\text{ k} = -2\text{ V}$

A basic circuit of an operational amplifier used as a summing amplifier is shown in Fig. 17.18 (b). This circuit has four input voltages, V_0, V_1, V_2 and V_3 the circuit analysis is as follows.

$$I_{in} = I_0 + I_1 + I_2 + I_3$$

Where I_0, I_1, I_2 and I_3 are the individual currents of each input branch.

But $$I_0 = \frac{V_0}{R_0}, I_1 = \frac{V_1}{R_1}, I_2 = \frac{V_2}{R_2}, I_3 = \frac{V_3}{R_3}$$

and $$V_{out} = -I_F \times R_F$$

Therefore $$I_{in} = \frac{V_0}{R_0} + \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}$$

This principle can be used in digital to analog conversion (D/A), as shown in Fig. 17.18(c). The summing amplifier consists of four inputs having resistances 8*R*, 4*R*, 2*R* and *R*, corresponding to the 4 bit binary inputs D_0, D_1, D_2 and D_3 respectively. Switches are provided in each input, so that if a switch is open it corresponds to '0' and if closed it corresponds to '1'. The '1'or high input to the binary input is obtained from a reference source + V_{ref}.

For a 4-bit summing amplifier used as a D/A converter, the following is applicable.

$$I_3 = \frac{V_{ref}}{R},\ I_2 = \frac{V_{ref}}{2R},\ I_1 = \frac{V_{ref}}{4R},\ I_0 = \frac{V_{ref}}{8R}$$

But $$I_{in} = I_0 + I_1 + I_2 + I_3$$

$\therefore$ $$I_{in} = \frac{V_{ref}}{8R} + \frac{V_{ref}}{4R} + \frac{V_{ref}}{2R} + \frac{V_{ref}}{R}$$

$\therefore$ $$I_{in} = \frac{V_{ref}}{R}\left[\frac{1}{8} + \frac{1}{4} + \frac{1}{2} + 1\right]$$

$$= \frac{V_{ref}}{R}\ [0.125 + 0.25 + 0.5 + 1]$$

$$I_{in} = \frac{V_{ref}}{R}\ [1.875]$$

Hence for a 4 bit D/A converter

$$I_{in} = \frac{V_{ref}}{R}\ (D_3 + 2^{-1} D_2 + 2^{-2} D_1 + 2^{-3} D_0)$$

Hence the total voltage available at the output of the opamp is the total of the input voltage levels which represents the equivalent analog signal of the 4 bit digital input.

Example 17.4 *If V_{ref} = 5 V and R = 5 kΩ. Then the analog output current for various binary input for Fig. 17.18 (c) is shown in Table 17.5.*

Table 17.5 Various Equivalent Analog Outputs for Each Digital Input

D_3	D_2	D_1	D_0	*Output Current in mA*	*Fraction of Maximum*
0	0	0	0	0	0
0	0	0	1	0.125	1/15
0	0	1	0	0.250	2/15
0	0	1	1	0.375	3/15
0	1	0	0	0.500	4/15
0	1	0	1	0.625	5/15
0	1	1	0	0.750	6/15
0	1	1	1	0.875	7/15
1	0	0	0	1.000	8/15
1	0	0	1	1.125	9/15
1	0	1	0	1.250	10/15
1	0	1	1	1.375	11/15
1	1	0	0	1.500	12/15
1	1	0	1	1.625	13/15
1	1	1	0	1.750	14/15
1	1	1	1	1.875	15/15

17.7.5 Weighted Converter Using Transistor Switches (Current Switch)

Figure 17.19 (a) shows a circuit diagram of a weighted converter, using transistor switches.

When the D bit is high (logic 1), it produces enough base current to saturate the transistor. When the D bit is low (logic 0), the base current is zero and the transistor remains off. Therefore, each transistor is in either saturation or cutoff, and acts as an open circuit or closed switch. If the D bits are the output of a counter, the output is 0000, 0001, ..., 1111. The digital signal is converted into a continuous signal i.e. D/A conversion has taken place. One step is equal to an LSB increment. The resolution of the staircase is defined as the ratio of the LSB increment to the maximum output. Therefore

$$\text{Resolution} = \frac{1}{2^4 - 1} = \frac{1}{15}$$

This is sometimes given as 1 part in 15 and

$$\%\ \text{Resolution} = \text{Resolution} \times 100$$

For a 4 bit converter, the % resolution is $1/15 \times 100 = 6.67\%$.

The number of bits gives the resolution and an estimates number of steps. This converter is a weighted resistor converter. The accuracy of the output depends mainly on the tolerances of the weighted resistors. If they are exactly R, $2R$, $4R$ and $8R$, then all steps will be equal. If the resistances do not have these tolerances, the stairs (steps) will have non-equal steps. The various % resolutions for different numbers of bits are given in Table 17.6. The output staircase waveform is shown in Fig. 17.19(d).

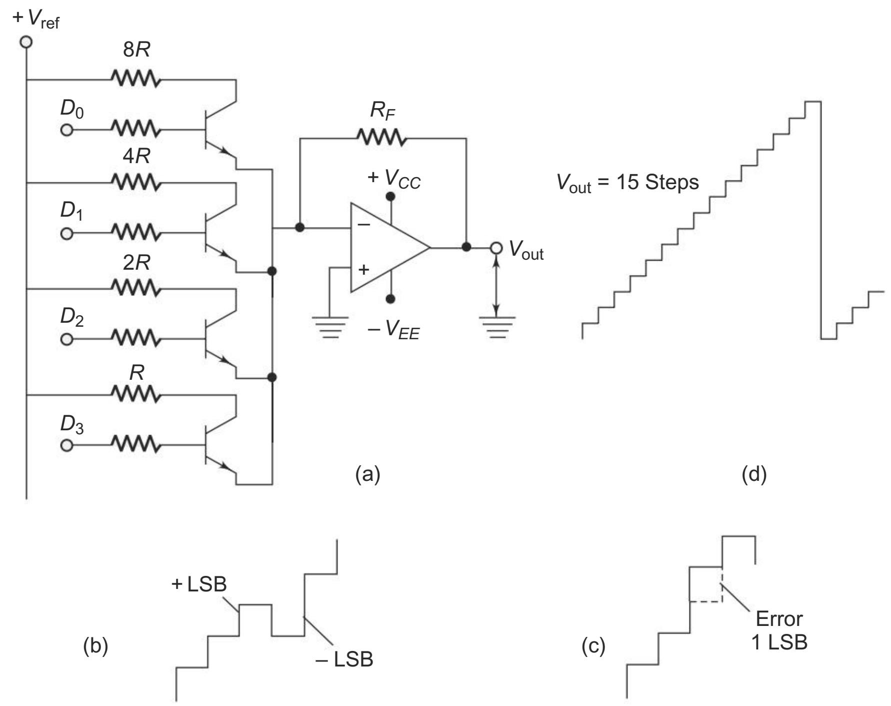


Fig. 17.19 (a) Weighted converter using transistor switches (b) Monotonic staircase (c) Non-monotonic staircase (d) Staircase waveform

Table 17.6 Resolutions for Different Number of Bits

Bits	*Resolution*	*% Resolution*
4	1/15	6.67%
6	1/63	1.59%
8	1/255	0.397%
10	1/1023	0.0978%
12	1/262143	0.00381%

When the weighted resistances used are of poor quality, the staircase produced will not be monotonic.

Monotonicity A monotonic D/A converter is one that produce an increment in output for each increment in the input. The staircases voltages are nonmonotonic shown in Fig. 17.19 (c). For a monotonic converter the error permissible is less than +1/2 LSB. Considering the worst case of +1/2 LSB error, if it is followed by a – 1/2 LSB error, a critical level is produced, where the monotonicity is not lost as shown in Fig. 17.19 (b).

Disadvantage of Weighted Converters The more the number of steps the converter has, the better the tolerances of the step, and the tolerances of resistance required. For example, if the staircase has 15 steps, the tolerance of the resistance should be less than + 6.67% (1/15). If the staircase has 255 steps, the resistance tolerances must be better than + 0.04% (1/255).

Therefore, for an 8 bit converter, resistance of a very high quality is required, which is a disadvantage. Another difficulty that arises with weighted resistors is that for an 8 bit we need resistances R, $2R$, $4R$, ... 128 R. Therefore, the largest resistance is 128 times the smallest. For a 12 bit converter, the largest resistance must be 2048 times the smallest. Because of this disadvantage, it is not possible to produce mass weighted resistor D/A converters. Each converter has to be made and tested individually.

17.7.6 Practical D/A Converter

The resistive divider or ladder can be used as the basis for a D/A converter. It is in the resistive network that the actual translation from a digital to an analog signal voltage takes place. However, there is a need for additional circuitry to complete the design of a D/A converter.

An integral part of a D/A converter is a register which can be used to store digital information. The simplest register is formed using an RS flip-flop, with one F/F per bit. There must also be a level amplifier between the register and the resistive network, to ensure that the digital signals presented to the network are all of the same levels and are constant. Finally there must be some form of gating on the input of the register, such that the F/F's can be set with the proper information from the digital system. A basic block diagram of a D/A converter is shown in Fig. 17.20, and a complete block diagram of a 4-bit D/A converter is shown in Fig. 17.21.

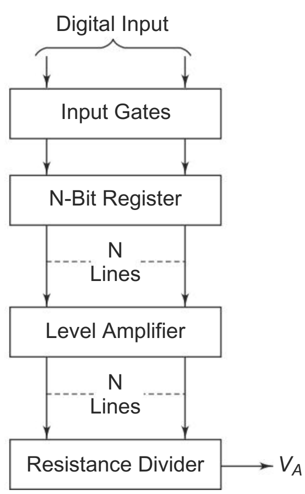


Fig. 17.20 Block diagram of a D/A converter

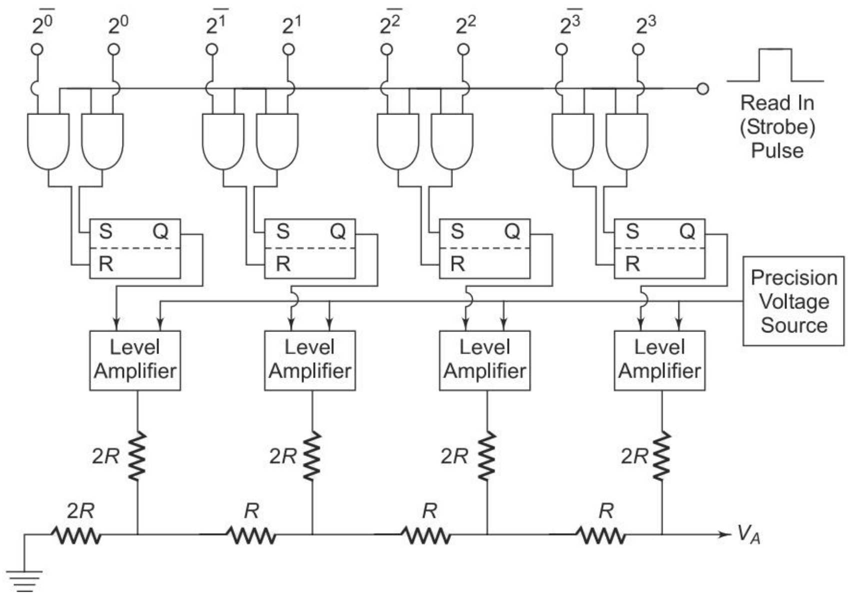


Fig. 17.21 4-Bit D/A converter

The resistive network used is of the ladder type. The level amplifiers have two input each. One input is the +10 V from the precision voltage source, and the other is from an F/F. The amplifier works in such a way that when the input from an F/F is high, the output of the amplifier is at +10, and when the input from the F/F is low, the output is 0 V.

Four F/Fs used form the register necessary for storing the digital information.

The F/F on the right represents the MSB and the F/F on the left LSB.

Each F/F is a simple RS type F/F and requires a + ve level at R or S inputs to reset or set it.

When the read line goes high, one of the two gates connected to the F/F is true (enabled) and the F/F sets or resets accordingly. Hence data are entered into the register each time the read in (strobe) pulse occurs.

Quite often it is necessary to decode more than one signal, e.g. $X-Y$ coordinates for a plotter. There are two methods of decoding these signals.

1. The first method, is to simply use a D/A converter for each signal, as shown in Fig. 17.22 (a). This has the advantage that each signal to be decoded is held in its register and the analog output voltage is then held fixed. The digital input lines are connected in parallel to each converter. The proper converter is then selected for decoding by the select lines.
2. The second method involves using only one D/A converter and switching its output. This is called multiplexing, and is shown in Fig. 17.22 (b). The disadvantage here is that the analog output signal must be held

between sampling periods and the outputs must therefore be equipped with sample hold circuits.

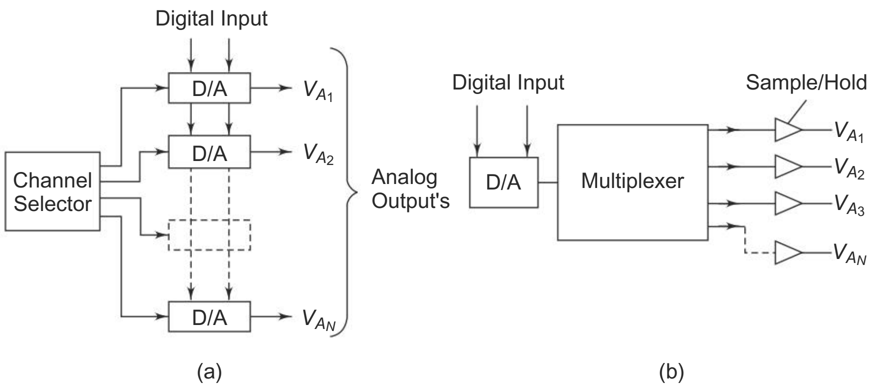


Fig. 17.22 (a) Channel selector (b) Multiplexer method

A sample hold amplifier can be approximated by a capacitor and high gain opamp, as shown in Fig. 17.23.

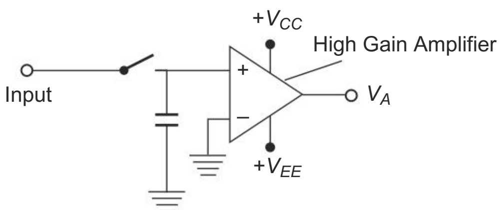


Fig. 17.23 Sample/hold circuit

When the switch is closed, the capacitor charges to the D/A converter output voltage. When the switch is opened, the capacitor holds the voltage level until the next sampling time. The opamp provides a large input impedance, so that the capacitor is not discharged appreciably and at the same time offers gain to drive external circuits.

When the D/A converter is used in conjunction with a multiplexer, the maximum rate at which the converter can operate must be considered. Each time data is shifted into the register, transients appear at the output of the converter. This is due to the different rise and fall of each F/.F. Hence, a settling time must be allowed between the time data is shifted into the register, and the time the analog voltage is read out. The settling time is the main factor in determining the maximum rate at which the output can be multiplexed. The worst case is when all bits change (i.e. from 1000 to 0111). The sampling rate is a function of the capacitors, as well as the frequency of the analog signal which is expected at the output of the converter.

DATA LOGGERS 17.8

The basic function of data loggers is to automatically make a record of the readings of instruments located at different parts of the plant. Data loggers measure and record data effortlessly as quickly, as often, and as accurately desired.

Measurement errors are eliminated and the log is permanent in a form most suited to ones requirements.

It can measure electrical output from virtually any type of transducer and log the value automatically. Since simple presentation of data for a large number of points at regular intervals of time is not sufficient, present day data loggers are designed to scan and record data very fast under fault conditions of automatic initiation. In addition, automatic data loggers are capable of giving plant performance computation and a logic analysis of alarm conditions during emergency.

A data logger processes the readings to render then immediately as recognisable scientific units (°C kg/cm^2, etc.). It can detect readings outside defined limits and initiates corrective action, and it can record all or selected readings on a variety of output devices, or pass them into a computer for further processing.

The measurement range, speed of measurement and operation for recorders can be different for blocks of channels or even for single channel and all this is achieved in data loggers by having preset programs. Since loggers handle the data in digital form, they can manipulate and process the reading without any loss of accuracy.

A logging sequence may be started manually by pressing a push button, or automatically by the clock. When a scan is started by the clock, the logger advances ahead by time. The clock will start logging sequences at intervals of one second, up to two hours. The log can be started and stopped on any channel.

Presently, data loggers are invariably used in power generation plants, petrochemical installations, continuous process plants, engine testing, component evaluation, etc.

Modern data loggers are tailor made systems, designed precisely from standard modules to meet a particular requirement. They generally possess the following characteristics:

1. *Modularity* Systems can be expanded whenever required, simply and efficiently, often entailing little or no interruption to the working system.

2. *Reliability and Ruggedness* Designed to operate continuously without interruption even in the worst industrial environments.

3. *Accuracy* The specified accuracy is maintained throughout the period of use.

4. *Management Tool* In addition to simple data acquisition, there are facilities to perform many other functions, and present the results in handy form.

5. *Easy to Use* These communicate with operators in a logical manner, are simple in concept, and therefore easy to understand, operate and expand.

17.8.1 Basic Operation of a Data Logger

For proper understanding of a data loggers, it is essential to understand the difference between analog and digital signals. For example, measurement of temperature by a milli voltmeter, whose needle shows a reading directly proportional to the emf generated by the thermocouple, is an analog signal.

However, digital equipment presents a digital output in terms of pulses and involves an electronic pulse counting equipment which counts the number of pulses. The pulses are generated such that each pulse corresponds to the smallest value of the parameter being measured.

These digital signals are precise at all times. Consider the example of temperature. In the case of analog measurements even the accuracy of the potentiometric method is limited by the precision with which the resistance can be subdivided. In the digital method, the electrical signal obtained from the thermocouple is subdivided by an electronic decade circuit and thus the thermocouple voltage can be measured to many places of decimal.

An analog device is capable of measuring with an error of ± 0.5% to ±1%,whereas a digital device can be obtained with an error of any ± 0.01%. An analog instrument responds to a change in input levels in times of the order of 0.25 to 1s while a digital instrument gives accurate readings in a few hundredth's of a second, and often many times faster.

One advantage of a digital instrument is that its reading can be recorded by suitable printer.

The data logger senses only digital signals and hence analog signals, if any, have to be converted to digital signals. The digital technique is employed because it measures very small (or large) signals accurately and fast.

The recording device may be a printed log or a punched paper tape. The printed output can be either line by line on a paper strip or on a type written page.

Time words are printed at the start of each sequence. Time is recorded in hours, minutes and seconds. Data consists of the channel identity number, followed by polarity indication (+ or –), the measured value (4 or 6 digits) and units of measurement. Sometimes the range may also be indicated.

Basic parts of a data logger

1. Input scanner
2. Signal conditioner
3. A/D converter
4. Recording equipment
5. Programmer

The block diagram of a data logger involving all these parts is shown in Fig. 17.24.

The input scanner is an automatic sequence switch which selects each signal in turn. Low level signals, if any, are multiplied to bring them up to a level of 5 V. If the signals are not linearly proportional to the measured parameter, these signals are linearised by the signal conditioner.

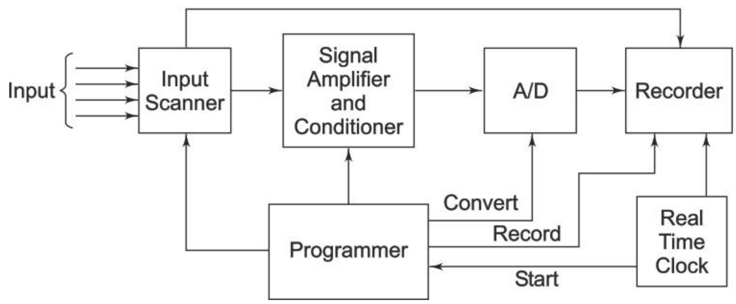


Fig. 17.24 Block diagram of a data logger

The analog signals are then converted to digital signals suitable for driving the recording equipment (printer or punched paper tape).

The programmer (serialiser) is used to control the sequence operation of the various items of the logger. It tells the scanner when to step to a new channel, and receives information from the scanner, converter and recorder. The real time clock is incorporated to automatic the system. The clock commands the programmer to sequence one set of measurements at the intervals selected by the user.

Input Signals The input signals fed to the input scanner of the data logger can be of the following types.

1. High level signals from pressure transducers
2. Low level signals from thermocouples
3. ac signals
4. Pneumatic signals from pneumatic transducers
5. On/off signals from switches, relays, etc.
6. Pulse train from tachometer
7. Digital quantities

The last three signals (5, 6 and 7) are of the digital type and are handled by one set of input scanners and the remaining signals are of the analog type and are handled by a different set of input scanner.

Low level dc signals are first amplified and then conditioned by the law network and finally fed to the A/D converter.

High level signals are fed straight to law network and converter.

The ac and pneumatic signals are first converted to electrical dc signals, conditioned and then converted.

In this manner, all types of signals are converted to a form, suitable for handling by the data logger.

The purpose of the conditioner is to provide a linear law for signals from various transducer which do not have linear characteristics.

Filters are used for noise and ripple suppression at the interface of the output of the transducers and the input of the signal conditioner, since these signals carried by the cables are of very low magnitude. Digital signals are then fed to the digital interface, whereas analog signals are first amplified, linearised and then brought to the analog interface. They are then converted into digital form and finally fed to the digital interface.

1. Input Scanners

Because the scanner select each input signal in turn, the data logger requires only one signal amplifier and conditioner, one A/D converter and a single recorder.

Modern scanners have input scanners which can scan at the rate of 150 inputs/s, but the rate of scanning has to be matched with the rate of change of input data, and the time required by the recorder and the output devices to print one output.

Sometimes it is desirable to scan certain parameters at a faster rate and some others at a longer intervals. For such mixed scan rates, the scanning equipment is designed for an interlaced scan operation, in which it is possible to log some parameters at 30 – 60 minutes interval, some every 5 minutes, and others every few seconds.

A scanner, in effect, is a multiway switch which is operated by a scanner drive unit for selecting the circuits. As the switch contacts have to continuously (24 hours/day) deal with low level signals at very high frequencies, the following requirements (desired characteristics) must be considered in the design of the contacts and their operations.

1. Low closed resistance
2. High open circuit resistance
3. Low contact potential
4. Negligible interaction between switch energising signal and input signals
5. Short operating times
6. Negligible contact bounce
7. Long operation life

Although it may not be possible to achieve all these characteristics in one switch, the arrangement selected must satisfy the maximum possible conditions.

The various switching elements available commercially are as follows.

1. Rotary selector switch
2. Electromagnetic operated relays
    (i) Dry Reed type
    (ii) Mercury wetted reed type
    (ii) Solid state switches

Scanner Drive The most common arrangement for selecting individual input one after another, is to use a matrix, as shown in Fig. 17.25.

The matrix is formed by using two energising lines, X and Y, corresponding to horizontal and vertical respectively, each having 10 contacts. Hence a 10×10 matrix is formed, giving 100 input channels per scanner unit or module. The only relay at the intersection of the energised X and Y lines is operated. The timing pulse thus consists of two signals, one for the X line and the other for the

Y line. Each relay has a diode in series with its coil, to prevent other relays being energised via other paths.

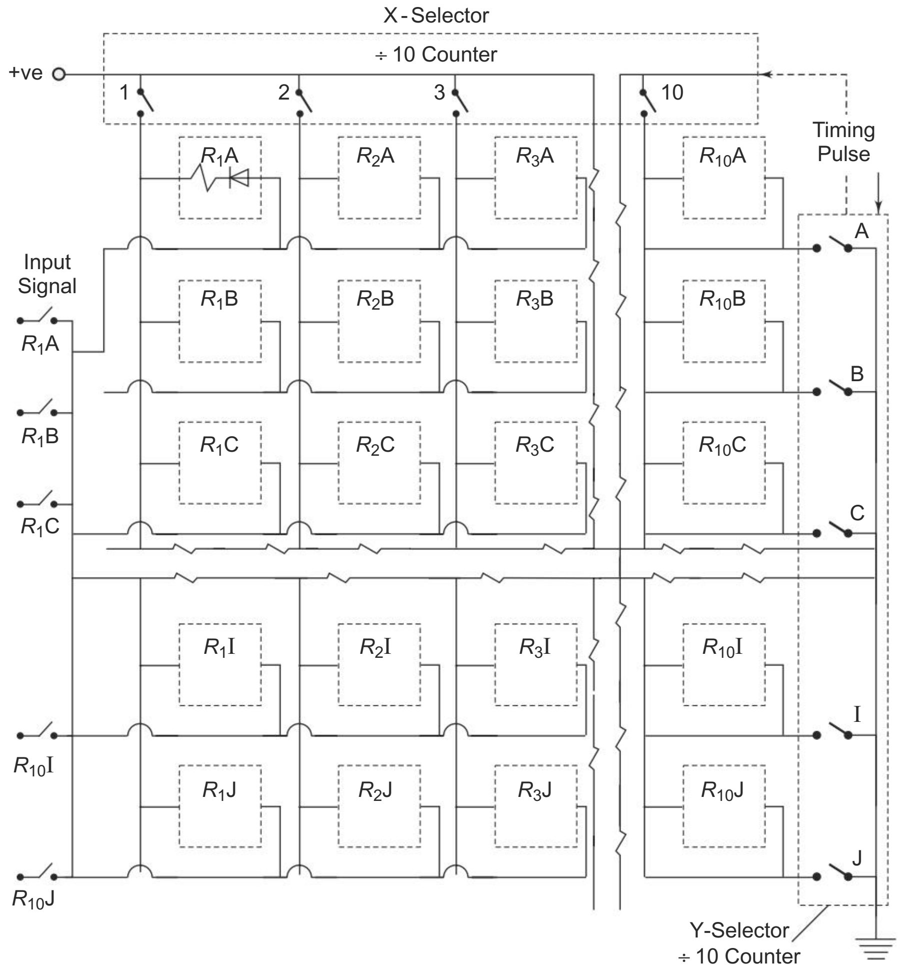


Fig. 17.25 Selector switch (Matrix)

The energising signals, i.e. the timing pulses are normally developed using counter circuits, which start off selecting one of the *X* lines, and then all the *Y* lines in sequence. After this cycle is completed, the next *X* line is selected and again all the *Y* lines are selected in turn. In this way each input is scheduled in turn. Generally, transistor switches are used to select the input relay.

As the measuring transducers and sensing elements are located at distances of about 300–400 m away from the scanner, the electrical interference, i.e. the electrical voltage induced in the signal lines can swamp an actual signal voltage.

The most common method to eliminate or reduce the effect of common mode noise is to use an amplifier with a floating input.

Although the effect of noise is practically eliminated as the low level signals, these signals have to be amplified further to a suitable level to drive the A/D converter. Hence the signal amplifier should have the following characteristics.

1. Precise and stable dc gain
2. High signal to noise ratio
3. Good linearity
4. High input impedance
5. High CMRR
6. Low output impedance
7. Low dc drift
8. Wide band width
9. Fast recovery time

2. (Signal) Input Conditioning Since data loggers give their readout in the units of measurements concerned, there are two requirements:

(i) Scaling linear transducers

(ii) Correcting the curvature of a non-linear transducer, such as a thermocouple

Linear inputs can be dealt with in two ways.

(i) The simplest is to provide individual resistance attenuation on each input in order to reduce the transducer output level, where the scale factor is an integral power of ten. For example, if a particular transducer has a full scale of 10 mV for a pressure of 500 kg/cm^2, we can reduce the value to one half by the use of an attenuator, such that 500 kg/cm^2 may be represented by 5 mV. If the system is to have a resolution of 1 kg/cm^2, the A/D converter must have a resolution of 10 μV. This technique is limited only by the sensitivity of the A/D converter.

(ii) The second method is to change the sensitivity of the A/D converter. But since each input may require a different scale factor, this is not convenient as an input attenuation technique.

The signal can be linearised at any one of the following three places.

(i) In the analog stage before conversion

(ii) In the conversion process

(iii) Digitally after conversion

The first method is not suited to low level voltages, as it requires some form of amplification. The signal conditioner may be placed between the scanner and the converter. But, each type of transducer requires individual linearising circuits.

The third method requires a storage capability and a computer processing technique. The most satisfactory is the second method, whereby linearisation is built into the conversion process.

3. A/D Converters These have been discussed in detail Chapter 5.

4. Recorders The output from the data logger can be printed on any of the following.

1. Typewriter
2. Strip printer and/or digitally recorded on punched tape or magnetic tape for further analysis in a digital computer.

The typewriter provides a conventional log sheet with tabulated results, and prints in two colours.

The signals obtained from the A/D converters are applied to the electromagnetic operated levers of a typewriter. Plus, Minus, characters which can be printed one at a time, decimal point shift, line shift, type colour and spacing are controlled by the EM solenoids which are energised from the programmer unit.

Punched paper tape or magnetic tape is used when the recorded data is to be further analysed or where the rate of data acquisition is too great for a printer.

5. Programmer This can be considered as an automatic sequence switch which controls the operation of all other units of the data logger. The sequential operations performed by a programmer are as follows.

(i) Set amplifier gain for individual input, i.e. gain of the amplifier has to be so adjusted that for a maximum value of input signal, the A/D converter records a full scale reading.

(ii) Set linearization factor so that the adjusted output from the signal amplifier is directly proportional to the measured quantity.

(iii) Set high and low alarm limit

(iv) Initiate alarm for abnormal condition

(v) Select input signal scanner switching is set normally by a timing pulse to select the reset input.

(vi) Start A/D conversion

(vii) Record reading channel identify and time (in order that the readings may be identified at a later stage, a number identifying that the input has been normally recorded, with the actual reading and the time during the beginning of each complete scan).

(viii) Display reading

(ix) Reset logger. (At the end of cycle the A/D converter sections of the logger are reset to their initial conditions and the cycle starts again.)

17.8.2 Summary of a Data Logger

In general, a data logger is a comprehensive and highly advanced DAS. It is made versatile and flexible, to render it suitable for widely varying applications, specific requirements being met simply by setting up a suitable program.

It can accommodate from 10 to 1000 analog signals, depending upon the capacity of the analog scanner selected. There may be three or four poles per unit (or a combination) and the signal conditioning options of the scanner may be used. A digital voltmeter is used as the system A/D cinverter. The logger programs the voltneter (range, mode, integration period) according to the requirements of the particular input channel. Generally, data loggers are designed to drive two output recorders completely independent of each other.

The basic assembly of data logger consists of main frames, front panel assembly, and power supply unit, with the following essential modules.

1. Scanner controller
2. Data exchange
3. Processor
4. Programmer
5. DVM interface

The optional modules, the need for which depends on the requirements of the system, include the following.

1. Data buffer store
2. Alarm store
3. Linearise
4. Additional programmers

The function of various modules are as follows.

1. Scanner Controller It is an interface between the logger and the analog scanner. It selects the channels for examination and receives the program location address from the scanner.

2. Data Exchange All data transfer within the logger is made via this module. It ensures correct transmission of data throughout the system.

3. Processor It controls the sequence of events within the logger by defining and producing commands for the modules.

4. Programmer This is used to set up the channel programs, so that the DVM functions, output functions and alarm limit can be arranged for individual channels or groups of channels.

5. DVM Interface The DVM functions (range, mode, integration periods, etc.) are selected by the program instruction and transmitted via the interface.

After the completion of a measurement, the interface receives the DVM measurement data, which is then supplied to other modules in the logger.

6. Data Buffer Store Measurement data can be held in a buffer store. Use of the store enables the logger to scan at speeds higher than the operating speeds of the output recorders.

7. Alarm Store During an alarm scan, and with an alarm store fitted to the logger, a channel is recorded only when it goes into or comes out of alarm. Hence repeated recording of the channel in alarm are avoided.

8. Lineariser

It provides scale factors and linearisation laws which convert the DVM reading into engineering units.

In additions to these modules, the basic assembly also provides facilities of digital clock, off limit detector and a display unit. The digital clock has an output of days (up to 99), hours, minutes and seconds which may be recorded on the output devices. It can command the logger to make measurements at practically any time interval.

The off limit detector allows any measured values of any desired channel to be compared to preset upper and lower limit values. When a channel is found to be outside limits, the fact may be indicated on the output record, or an external signal may be given, or the logger may automatically go into a different routine. The limit values are set up on the program pin board.

Control Facilities There are three main areas in which the operator can control and program the logger.

1. Front panel controls
2. Programmer pin boards on which channels programs are defined
3. Diode matrices and soldered wire links on modules by which output format and code are defined.

17.8.3 Compact Data Logger

A typical unit provides 60 channels of data in a $20 \times 40 \times 60$ cm box weighing about 20 kg. Most manufacturers offer local or remote add-on scanners to expand to about 1000 channels. Scan rates are modest usually (1 – 20 channels per second) and though versatile signal conditioning is provided, the signal processing capability is limited to simple functions such as (mx + b) scaling, time averaging of single channels, group averaging of several channels, and alarm signaling when preset limits are exceeded. However, most units do allow interfacing to computers, where versatile processing is possible.

Data loggers of this class utilise a built in microprocessor to control the interval of operations and carry out calculations through a single amplifier –A/D converter, which is automatically ranged or gain switched under program control to accommodate the signal level of each channel, as shown in Fig. 17.26.

This is not useful for applications in which fast changing signals must be observed, since a (typical) 5 channels per second scan rate takes 12 s to scan 60 channels before returning to any given channel. Also, the time skew of 12 s can cause a density error if the signals change too rapidly (for example, if gas density from a pressure on channel 1 and temperature on channel 60 is measured).

Often multiplexers (scanners) are available in both general purpose (two wire) and low level (two original wires plus shield) versions, since milli-volt level signals, such as from thermocouples, generally use a shielded, twisted pair of conductors. A three wire system scanner can reduce errors from about 10 to 1 μV. Electro-mechanical reed switches are used frequently in such scanners, since speed requirements are modest but low noise is important.

Since thermocouples are very common in data logger applications, reference junction compensation and linearization options are always available. Reference-junction compensation can be offered economically and accurately for any

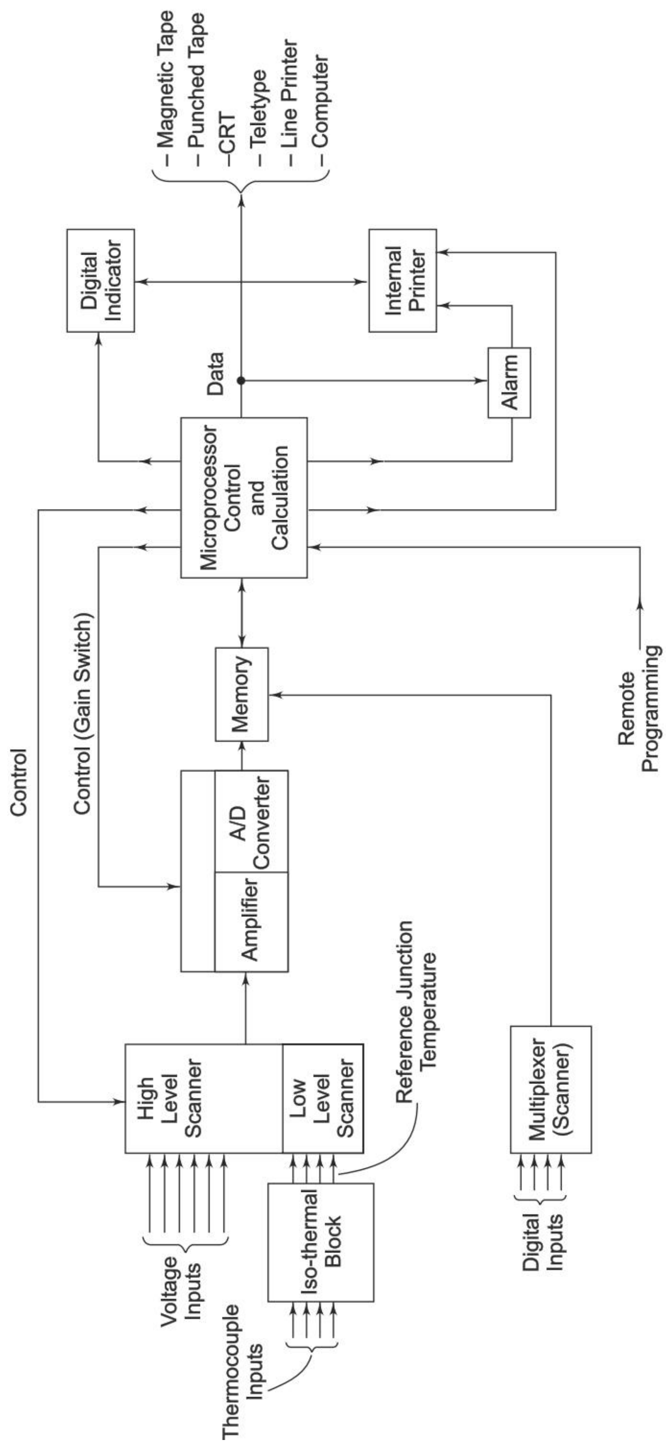


Fig. 17.26 Compact data logger configuration

mixture of thermocouple types by the use of an isothermal connection block. This thermocouple terminal block is designed to have an uniform (±0.05°C) temperature (the reference junction) over its length. The block temperature is allowed to drift with ambient conditions, but is measured (often with a junction semiconductor sensor, since these work well near room temperature). This reference-junction temperature is sent to the microprocessor, where the temperature/voltage data for each thermocouple being employed is stored and the necessary correction is calculated. The microprocessor also stores the equations which curve-fit the thermocouple tables (over the desired range) for each thermocouple type, providing software linearization. For resistance thermometers the data loggers provide constant-current excitation and software linearizations.

The system amplifier and A/D converter are the crucial elements for overall system accuracy. Since data logger inputs vary widely in voltage range, while the A/D input is typically fixed at +10 V, the microprocessor sets the amplifier gain at a proper value as each channel is sampled (some data loggers also provides automatic ranging). These range selections are entered from the front panel when the logger is programmed for the particular application. Programming of this and other functions is very simple, and do not require a knowledge of computer languages.

A typical set of ranges and resolutions would be ± 40 mV (1 μV resolution), ± 400 mV (10 μV resolution), ± 4 V (100 μV) and ± 40 V (1 mV) with a input impedance of 200 MΩ. (Except 10 MΩ on 40 V range). Zero drift is kept negligible by an automatic zero system.

A/D converters are often of the dual slope type or voltage to frequency type, since conversion speed is modest and these integrating converters give good noise rejection. Fast or slow scanning rates (say 15 versus 3 channels per second) may be selected, to allow a trade off between speed and accuracy, since as integration time is increased the integrating A/D noise rejection improves.

A readout obtained by means of a built in digital indicator and two colour printer (prints alarm in red), or channel number, date and time of day is a standard. When a built-in printer is used, the printer speed (2 to 4 channels per second) limits the overall speed, even though scanning without printing may be possible at 15 to 20 channels per second. The readout format is selected by front panel programming.

Some units provide a 5 years non-volatile program memory which preserves stored programs in case of power failure. Interface options for external magnetic tape, punched tape, CRT terminals, line printers and computers are usually obtainable.

SENSORS BASED COMPUTER DATA SYSTEMS 17.9

This section describes hardware/software which is commercially available at several levels of completeness, ranging from single board computer "front ends" to stand alone system and high level programming languages.

Even the simplest and least expensive devices require considerable electronics/computer expertise on the part of the user, access to a micro-computer development system and sufficient engineering time to integrate the interface and computer into a working overall system. Since micro-computer system design is specialised and beyond the range of this text, we give only a brief description of micro-computer interface boards, shown in Fig. 17.27 as an example. These are available from several manufacturers, and interfaced with most popular micro-computers.

To reduce program storage requirements and execution times (at the expense of memory address area), a memory mapped I/O is often employed. Note that analog inputs (up to 32 single ended inputs 16 differential inputs) are processed through a multiplexer, programmable gain amplifier (PGA), sample/hold (S/H), and A/D converter in a fashion very similar to that of a data logger.

However, since we wish to handle HF signals, a successive approximations (rather than dual-slope) A/D converter (maximum throughput rate 28 kHz) and an electronic (rather than reed switch) multiplexer are necessary.

It is possible to obtain two (optional) D/A converters for driving analog recorders, generating analog control signals, etc. The hardware problems are reduced to a minimum by the use of such an interface card. Also, the overall system throughput rate must be less (often much less) than the 28 kHz value given above, since software execution time must be added to the A/D conversion time.

Figure 17.28 shows a single board, micro-computer based data acquisition system designed to accept multichannel analog and digital inputs and provide digital output to a host computer (usually a mini-frame or Main frame supporting high level languages such as BASIC and FORTRAN) through a standard serial communications port (RS 232C or 20 mA current loop). The on-board micro-computer unburdens the host computer by allowing supervisory control.

It performs data acquisition control, linearlization, conversion to engineering units, limits checking, interface control, and data output formatting. The analog channels are scanned continuously (15 to 30 channels per second) and the resultant data are stored in the micro-computer memory (RAM). The data in the RAM is refreshed on a continuous basis (the latest data is kept in memory), so that requests for data from the host are serviced immediately. Upon receipt of a transit command, the micro-computer [via. the UART (Universal Asynchronous Receiver Transmitter)] begins transmitting a string of data in the ASCII format to the host. No programming of the micro-computer is necessary, since it is preprogrammed by the firmware to respond to host commands.

The 12 channels of the analog input are broken into 3 groups of 4, and convenient 4 channel plug-in modules for thermocouples, RTDs, strain gauge transducers, etc. are available. Up to three expander boards can be controlled by

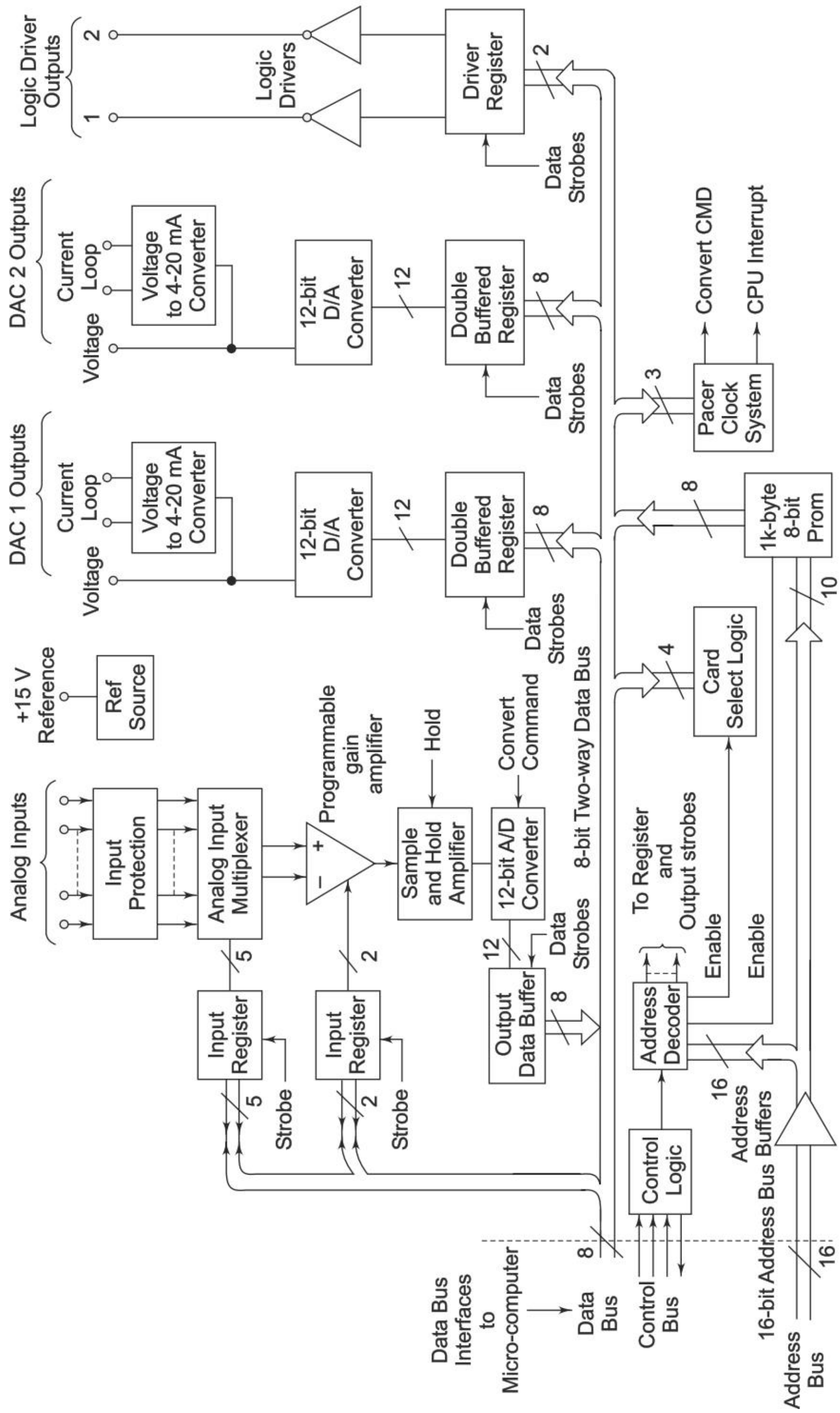


Fig. 17.27 Micro-computer interface board

a master board, creating a cluster as shown in Fig. 17.29. Up to 8 clusters can be operated from the same host, providing expansion to 384 channels.

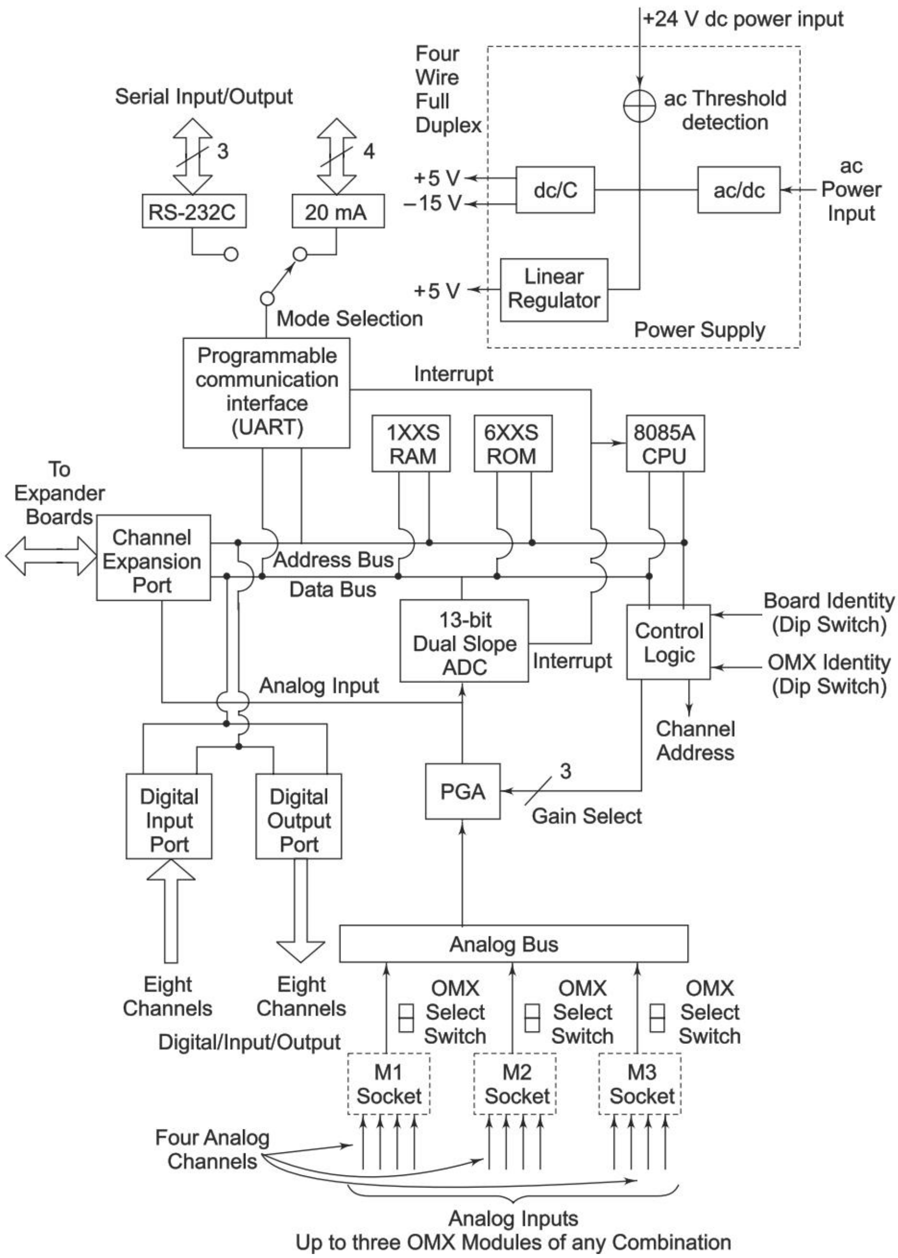


Fig. 17.28 Micro-computer based data acquisition system

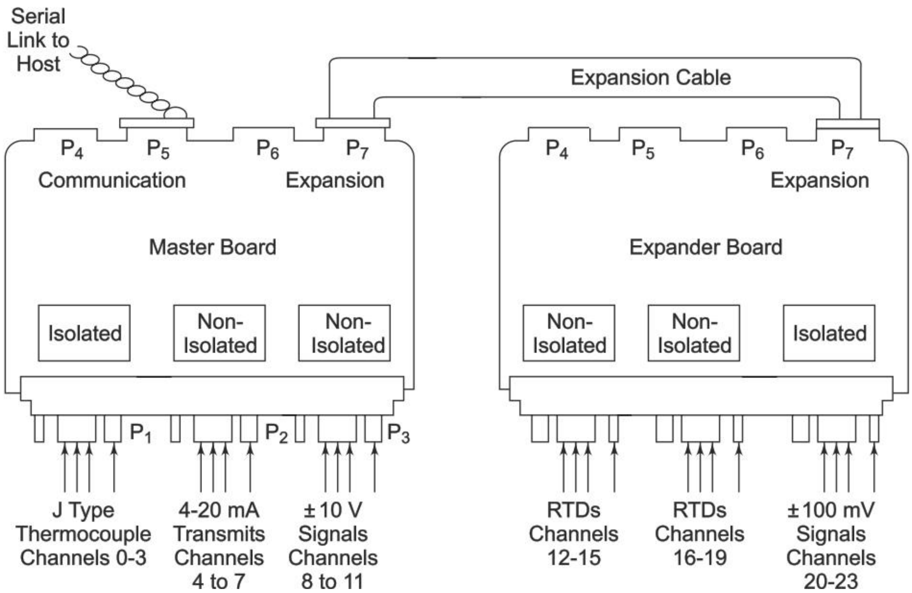


Fig. 17.29 System expansion by clustering

For those applications in which a sensor based measurement and control system with comprehensive and easy to use computer processing is designed, with a minimum of user engineering effort, complete stand alone systems such as that of Fig. 17.30 are available.

Analog and digital input/output is through a plug-in signal conditioning cards, space for 16 cards is provided (expandable up to 256 cards). A wide selection of cards functions are available, allowing easy interfacing with all kinds of sensors and control devices. Single cards are often themselves multichannel devices with a multiplexer on the card. Thus, two level of computer controlled multiplexing are present, slot multiplexing (chooses the card slot desired) and card channel multiplexing (selects the channel wanted on the chosen card). For example, a single digital input card provides 16 channels (bits) while the analog input card has 32 channels (single ended) or 16 channels (differential). This analog card has its own PGA (gain = 1, 16, 256) which combined with the PGA of the central controller (gain = 1, 2, 4, 8), allows versatile selection of channel gain under program control. Thermocouple cards are four channel units, which share a common reference-junction compensation circuit and have fixed gain. Lineraization is accomplished in software by a general purpose polynomial sub-routine. A fast (25 μS conversion time) successive approxi- mation A/D converter allows rapid scanning and storage of analog input (mixed channels at 2 kHz, single channel at 4 kHz). A mini computer specially designed for measurement control applications has 32,000 words of 16 bit MOS random access memory (RAM) augmented by 105 k bytes of cartridge tape mass storage.

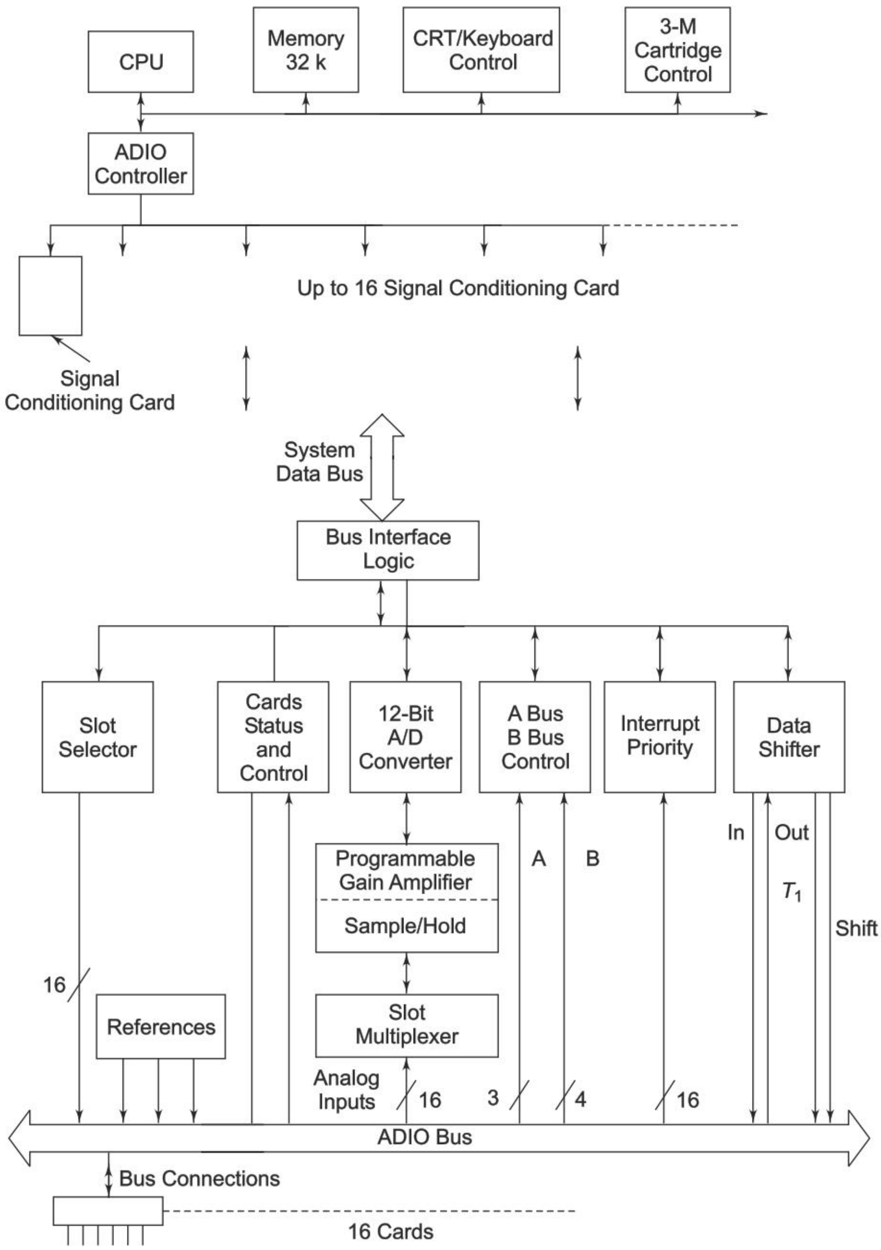


Fig. 17.30 Minicomputer based DAS

Programming is a Macbasic (a version of BASIC) specially enhanced for easy system operation. For example, the statement

$$V = AIN\ (2,1) - 1.53$$

assigns to the variable V, the value of the analog voltage on channel 1 of the analog input card in the I/O slot 2, minus 1.53. Similarly, the statement

AOT (1, 3) = 3.42 * 1.3

places a voltage of 4.44 V on Channel 3 of the analog output card in I/O slot 1.

For digital input variable DIN

I' = DIN (3,1)

takes the digital logic level from Channel 1 of the digital input card in I/O slot 3 and places it in I', where I' is an integer variable.

For digital output variable DOT

DOT (8,3) = X

turns on Channel 3 of the digital output card in I/O slot 8 if $X = 1$ and turns it off if $X = 0$.

System timing functions are eased by the availability of statements such as

WAIT 5.6

which causes the program to wait 5.6 before proceeding to the next statement.

A common requirement of many applications is the ability to perform several operations or tasks independently of one another in the same program. Examples include the monitoring of several analog signals in the laboratory or control of several process loops. To provide for this requirement, Macbasic is structured as a multitasking language and contains the necessary words for implementation.

Tasks are groups of Macbasic language statements that are defined as a task, and are executed each time the task is activated unconditionally or by satisfaction of a condition (such as an external event), or on a periodic basis. Up to 18 tasks may be defined at a given time and if more than one task is active at a given moment, the active task share resources and run simultaneously, unless priority is assigned to a particular task. If a task completes its operation, it can DISMISS itself an d return its resources to the system until the task is reactivated.

A simple Macbasic program might go as follows.

10 X = AIN (8,0)	— Acquire data on Channel 0 of analog input card I/O slot 8
20 PRINT X	— Print data
30 AOT (0,1) = 5 * X	— Scale data and output scaled value on Channel 1 of analog output card in I/O slot 0.
40 IF X > 5 DOT (1,5) = 1	— Activate alarm connected to Channel 5 of digital output card in I/O slot 1 if $X > 5$
50 WAIT .5	— Wait for 0.5 second
60 GO TO 10	— Acquire the next value

While this program is valid, certain capabilities are added if we convert it to the multitasking mode. These include the possibility of examining and modifying variables and/or listing the program while it is executing, which is a useful interactive feature.

```
10 TASK 1,40                — define task 1 which starts on line 40
20 ACTIVATE 1               — Active task
30 STOP                     — Stop primary task
40 K = 0.5                  — Set numerical value
50 L = 5                    — Set numerical value
60 X = AIN (8,0)            — Input analog data of Channel 0
70 AOT (0,1) = K* X         — Output scaled data on Channel 1
80 IF X > L DOT (1,5) = 1   — If X > 5 sound alarm
90 WAIT .5                  — Wait for 0.5 second
100 GO TO 60                — Acquire next value
```

All programs contain a primary task, such as lines 10, 20 in the above example, which define and activate the original program task. Once the primary task is completed, control is returned to the keyboard (line 30) and the user can interact with the program variables and tasks and/or list the program. Variable *L*, for example, could be modified while the program is running simply by typing, say $L = 5$.

An example with three separate tasks provides further details on multi tasking.

Task 1: Channels 0 through 9 of an analog input card in slot 0 are to be examined every 5 seconds.

Task 2: If any input is greater than 5.5 V, an alarm connected to Channel 5 of the digital output card in slot 1 is turned ON.

Task 3: The alarm remains ON until an operator activates a switch connected to a process interrupt card Channel 2 in slot 3. (This is defined as an event.) When the switch is activated, the alarm turns OFF and the current value of each analog input channel is printed.

```
10 TASK 1,100               — Define task 1, starting on line 100
20 TASK 2,200               — Define task 2, starting on line 200
30 TASK 3,300               — Define task 3  starting on line 300
40 ACTIVATE 1 PERIOD 5      — Activate task 1 every 5 s
50 ACTIVATE 3 ON EVENT (3, 2, 1) — Activate task 3 when Channel 2 of
                              Interrupt card in slot 3 goes to 1
60 STOP                     — Control is returned to keyboard for user
                              interaction with program variable tasks, etc.
100 K = 5.5                 — Set alarm limit value
110 FOR I = 0 TO 9          — Start standard basic loop
120 IF AIN (0, 1) >
K ACTIVATE 2                — Check for alarm condition
130 NEXT I                  — End standard basic loop which scans channel
                              0 to 9
140 DIMISS                  — Place task 1 on alert status, pending next
                              activation in 5 s
150 GO TO 110               — Restart here-go to repeat scan
200 DOT (1,5) = 1           — Turn on alarm
```

210 DIMISS — Task 2 complete
220 GO TO 200 — When reactivated, repeat
300 DOT (1,5) = 0 — Turn Off alarm
310 FOR I3 = 0 to 9 — Start standard basic loop scanning Channel 0 to 9
320 PRINT "CHANNEL"; I3, "="; AIN (0,I3) — Print each channel value
330 NEXT I3 — End standard basic loop for printing channels 0 to 9 when alarm has sounded and been turned Off
340 DISMISS — Task 3 complete
350 GO TO 300 — Repeat when reactivated

Of course, the above examples cannot explore all the programming possibilities of the Macbasic language, but they should make clear the ease with which numerical and logical operations needed to implement a desired measure ment/control system can be achieved with a high level language and computer hardware specially designed for such applications.

The standard system includes a key board and built in a CRT display, giving a stand alone programming capability. Optional features include IEEE-488 and RS 232 C communications, floppy disk hardware/software, interactive graphs, printers and remote interactive CRT terminals.

ELECTROMECHANICAL A/D CONVERTER 17.10

Another area of application in which A/D conversion is very important, involves the translation of the angular position of a shaft into digital information. (A very common application of this type of conversion is found in large radar installations, where the azimuth and elevation information are determined directly from the shaft position. There are many other examples in aircraft and aerospace fields.)

The method is not necessarily limited to rotational information, since rectilinear information can be translated into rotational information by means of a gearing arrangement.

In any case, the job involves changing position information (which can usually be considered as analog information, since it is continuous) into equivalent digital information. This is most generally obtained by the use of a code wheel, as shown in Fig. 17.31.

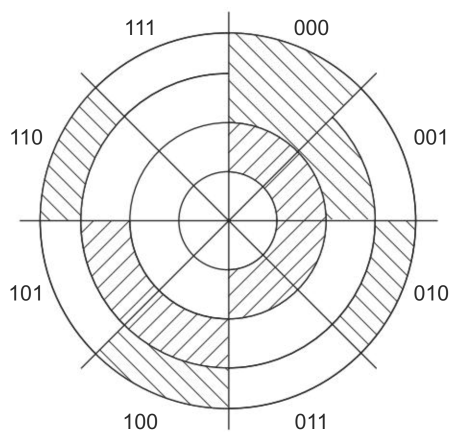


Fig. 17.31 3 binary code wheel

This particular wheel is coded in a straight binary fashion and represents 3 bits. The wheel is divided into three concentric bands, each representing one bit. The innermost band is divided into two equal segments and represents the MSB. The middle band is divided into four equal segments and represents the second MSB. The outer band has eight equal segments and is the LSB.

If the light areas on the wheel are transparent and the dark areas are opaque, the digital information can be obtained by placing light sources and photo sensors on opposite sides of the disk, as shown in Fig. 17.32. The output of the sensor is high if light is sensed and low if no light is sensed. These outputs then represent 1s and 0s and can be amplified, passed through logic circuits and used to set F/F. Such system is called an optical encoder.

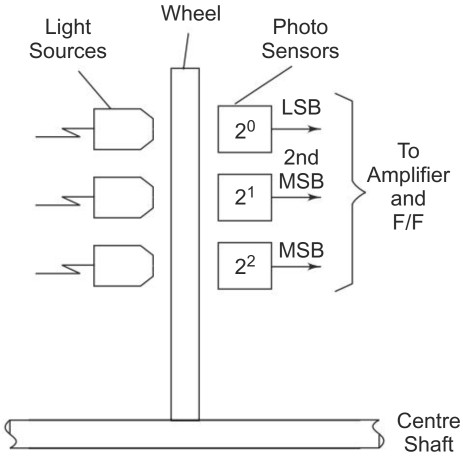


Fig. 17.32 3 binary code wheel showing photo sensor

The sensing could also be accomplished by placing a brush on each band making the light areas conducting and the dark areas insulating.

Now consider the wheel positioned under the sensors, such that the output reads 011. Further, let us assume that the wheel is very near the dividing line between 100 and 011. If there is an ambiguity in reading the LSB (that is, the sensor cannot decide whether to read 1 or 0, since it is right on the dividing line), the output may vary between 011 and 010. This is not too bad, since the output is jumping only from one adjacent position to the next. However if the ambiguity is in the MSB, the results could be disastrous, since the output jumps from 011 to 111, which is a 180° error.

It might be quite difficult to resolve the problem of sensing when the wheel stops on the dividing line, and it may be better to use a different code.

A code which changes only one bit at a time is used when going from any one position to the next. This ensures an ambiguity in only one position and eliminates the 180° ambiguity. Such a code used is called a gray code.

A code wheel constructed using a gray code is shown in Fig. 17.33. An examination of the wheel shows that the greatest error caused by a reading ambiguity is one segment of rotation.

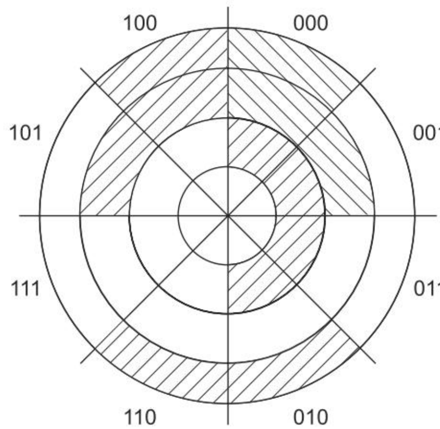


Fig. 17.33 3-bit gray code wheel

The 3-bit wheel, has the ability to digitise the shaft position into an equivalent 3-bit binary number. This implies eight positions around the wheel, each shaft position representing a 45° segment.

To obtain a closer reading, it is only necessary to add extra bands to the code wheel and hence extra bits to the digit number. In general, the degree of resolution obtained is given by $360°/2^n$, where n is the number of bits in the binary number.

DIGITAL TRANSDUCER 17.11

By the use of a digital code, it is possible to identify the position of a movable test piece in terms of a binary number. The position is converted into a train of pulses. This is achieved by a digital transducer and is also termed as encoder.

Since the binary system uses only two states, 0 or 1, it can be easily represented by two different types of systems, optical or electrical. Digital transducers using optical methods are called Optical encoders, while those using electrical methods are called resistive electrical encoders.

17.11.1 Optical Encoders

A sector may be designed as shown in Fig. 17.34, with a pattern of opaque and translucent areas. A photo sensor and a light source is placed on the two sides of the sector. The displacement is applied to the sector and therefore changes the amount of light falling on the photo electric sensor. The pattern of the illuminated sensor then carries the information to the location of the sector.

Figure 17.34 shows a possible pattern on sector of opaque and translucent areas. The number of levels in the encoder determines the accuracy with which the device operates.

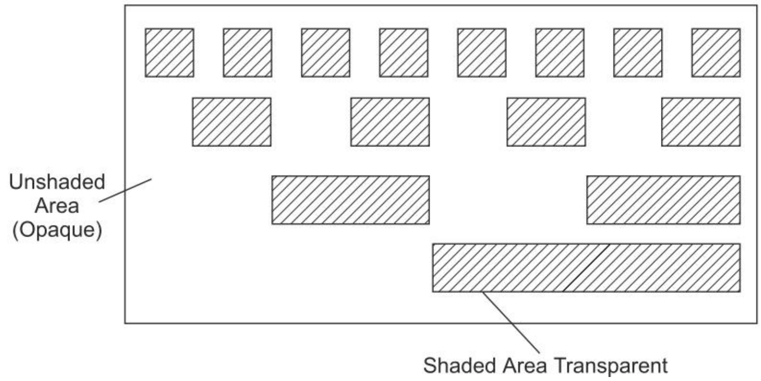


Fig. 17.34 Rectangular optical encoder

Advantages

1. They give a true digital readout
2. No mechanical contact is involved and therefore problems of wear and tear and alignment are not present

Disadvantages

1. Light sources burn out. (However, the life of the light is about 50,000 hours.)

17.11.2 Resistive Digital Encoders

Another method in which a pattern may be used is the resistive electric encoder. The shaded areas are made of conducting material and the unshaded areas of insulating material. Sliding contacts are used for making the contacts. Circuits of the sliding contacts which come in contact with the conducting areas are completed, while those which make contact with insulated areas are not completed. The encoder gives a digital readout which is an indication of the position of the device, and hence determines the displacement.

Advantages

1. It is relatively inexpensive.
2. It can be made to any degree of accuracy desired, provided the sector is made large enough to accumulate the required number of rows for binary numbers. The sectors are quite adequate for a slowly moving system.

Disadvantages

1. Wear and tear of the contacts causes error.
2. There is often an ambiguity of 1 digit in LSB.

17.11.3 Shaft (Spatial) Encoder

A spatial encoder is a mechanical converter that translates the angular position of a shaft into a digital number. It is therefore an A/D converter.

An increasing number of measuring instruments are being used to communicate with digital computers for measurement and control applications. There are two ways of generating digital signals. The first converts the analog variable to a shaft rotation (or translation in linear measurements) and then uses many types of shaft angle encoders to generate digital voltage signals.

The other form converts the analog variable into an electrical analog signal and then converts this into digital form. These two forms are very close to a true digital transducers.

To understand the operation of a shaft encoder, let us consider a translational encoder (a linear displacement transducer) shown in Fig. 17.34. The encoder shown has four tracks (bits) and is divided into conducting and insulating positions, with a smallest increment of 0.01 mm. As the scale moves under the brushes, the lamp circuits are made or broken, so that the number shown on the readout lamps is at every instant equal to one hundredth mm.

For angular displacements, the pattern given in Fig. 17.34 is changed or modified, so that the length of the scale becomes the circumference of a circle on a flat disc. The brushes are then placed along a radial line on the disc, as shown in Fig. 17.35.

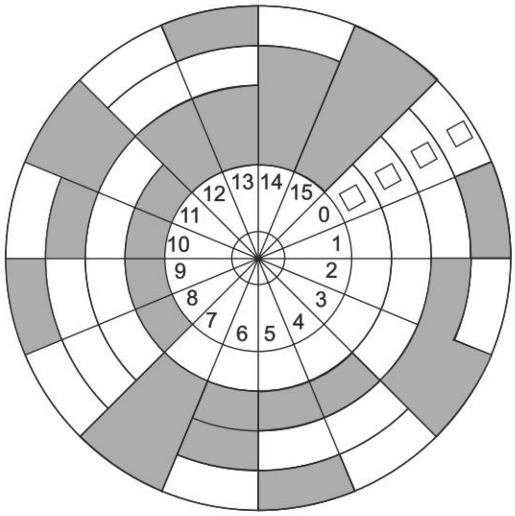


Fig. 17.35 Shaft encoder

The disc is divided into concentric circular tracks, each of which is then divided into segments in a manner depending upon the code being used.

For pure binary code, the inner most track is halved, the next quartered, the next divided into eight parts, and so on. Each track has twice as many segments as the adjacent one near the centre. The detection method determines the treatment of the disc. Alternate segments on each track are made transparent and opaque, if transmitted light and photo cells are used. If the segments are made reflecting and non-reflecting, reflected lights and photo cells are used. Electrical methods are used for detection in case the segments are made alternately conducting and non-conducting.

The accuracy depends upon on the number of tracks used and if there are n tracks, the accuracy obtained is $360°/2^n$.

Such discs are manufactured with diameters ranging from 5–25 cm and give unique codes for 100–50,000 positions per 360°. The sequence and order of indications represent the position of the shaft in the coded form.

Review Questions

1. What is a DAS?
2. State the importance factors that decide the configuration and subsystem of DAS.
3. Explain with a block diagram a generalized DAS.
4. State the various configurations of a DAS.
5. State the objective of a DAS.
6. What is signal conditioning? Why is it necessary in a DAS?
7. State the different parameters that are required by a signal conditioner.
8. What do you understand by ratiometric?

9. Explain with a block diagram the operation of a ratiometric conversion.
10. Explain logarithm conversion. How is it achieved?
11. Describe with the help of a block diagram the operation of a basic single channel DAS. Give examples.
12. State the drawbacks of a single channel DAS.
13. Explain in brief the working of an analog to digital converter.
14. Why is it necessary to use a preamplifier and filtering before data processing?
15. Explain with a diagram the operation of a DAS with preamplification.
16. What do you understand by multichannel DAS?
17. State the different ways in which multichannel DAS are used.
18. Explain with a block diagram the operation of a multichannel analog multiplexed DAS.
19. What do you understand by multiplexer? Where it is used.
20. Explain with a block diagram the operation of a multichannel DAS with multiplexing the output of Sample-Hold (S/H) circuits.
21. What do you understand by S/H circuit? Why is it necessary?
22. Explain with a block diagram the operation of a multichannel DAS using Digital multiplexing.
23. Explain with a block diagram how low level multiplexing is achieved.
24. What are the factors to be considered to achieve low-level multiplexing?
25. Describe briefly computer based DAS. State the advantages of using a computer based DAS.
26. How does the computer based DAS aid the operator?
27. What do you understand by analog to digital converter (A/D) and digital to analog (D/A) converter?
28. Explain with a diagram the operation of a variable resistor network D/A converter.
29. State the drawbacks of a resistive network.
30. Describe how Millman's theorem is used to find the output voltage of a resistive ladder.
31. State the basic principle of a ladder type D/A converter.
32. Explain with diagram the (R-2R) ladder D/A converter.
33. Explain with a circuit diagram the operation of a binary ladder. How is digital input converted to its equivalent analog value.
34. Explain with a circuit diagram, D/A converter using OPAMP.
35. State the advantages of using an OPAMP in a D/A converter.
36. What do you understand by weighted converters?
37. How is a binary equivalent weight determined?
38. Explain weighted converters using transistor switches with a diagram. State the advantages and disadvantages of Weighted converters.
39. What do you understand by monotonicity?
40. Explain with a block diagram the operation of a practical D/A converter.
41. Explain with a circuit diagram the working of a 4-bit D/A converter.
42. What are data loggers? What are the functions of a data Logger?
43. State the different elements of a data logger.
44. Explain with block diagram the operation of a data logger. State the functions of each block.
45. State the different input signals that can be fed to the input scanner of the data logger.
46. What do you understand by a input scanner.
47. Explain with diagram the operation of a input scanner.
48. What is a scanner? What do you understand by interlaced scanning.

49. State the desired characteristics that must be considered in the design of contacts and their operations.
50. State the various switching elements available commercially.
51. Explain with diagram the working of a selector switch or matrix. How does it select the various input signal.
52. Explain in brief the type of signal conditioning required in a data logger.
53. What is input conditioning. How is it performed?
54. State the different types of recorders used in a data Logger.
55. State the functions of a programmer.
56. List the various sequential operation performed by a programmer.
57. What is the basic assembly of a data logger? What are the optional modules provided?
58. Why is a D/A converter usually considered as a decoder?
59. Why is a A/D converter usually considered as a encoder.
60. Describe with block diagram the working of a compact data logger. State the functions of each block.
61. State the various characteristics of a modern data logger.
62. Explain in brief a sensor based computer DAS.
63. Explain with a block diagram the operation of a micro computer based DAS.
64. How is system expansion done by clustering?
65. Explain with block diagram the operation of a minicomputer based DAS.
66. What do you understand by an electromechanical A/D converter?
67. State the applications of an electromechanical A/D converter.
68. Explain the construction and operation of a 23 binary code wheel.
69. Describe how a Gray code eliminates larger ambiguity errors in reading optical encoders.
70. Explain why you would (or would not) make the innermost ring of the code wheel for an optical encoder the LSB.
71. What are digital transducers?
72. What are optical encoders and electrical encoders?
73. Explain in brief with a diagram the operation of optical encoder. State the advantages and disadvantages of an optical encoder.
74. Explain in brief the working of a resistive digital encoders. State the advantages and disadvantages of resistive digital encoders.
75. What is a shaft encoder?
76. Explain the construction and operation of shaft encoders. State the advantages and disadvantages of using a shaft encoder.

Multiple Choice Questions

1. With a high level signal (say 10 V) in direct digital conversion, when affected by noise of the order of 40 mV, an 8 bit converter can produce
(a) 1 bit ambiguity
(b) 2 bit ambiguity
(c) 3 bit ambiguity
(d) none of the above

2. If in pre-amplification circuit, signal levels below a tenth of an mV is needed then the following is used:
(a) amplifier
(b) attenuator
(c) differential amplifier
(d) summing amplifier.

3. In the pre-amplification circuit, when differential output from a bridge circuit is to be handled, then the following is used:
(a) summing amplifier
(b) OPAMP
(c) instrumentation amplifier
(d) differential amplifier.
4. Digital to analog converter can be considered as a
(a) decoding device
(b) encoding device
(c) multiplexer
(d) summing amplifier.
5. Analog to digital converter can be considered as a
(a) decoding device
(b) encoding device
(c) multiplexer
(d) summing amplifier
6. A binary ladder is constructed by using resistors having
(a) only two values
(b) only one value
(c) individual values
(d) none of the above
7. For a monotonic D/A converter, the error allowed is less than
(a) 1 LSB (b) + ½ LSB
(c) – ½ LSB (d) 2 LSB
8. If the staircase in weighted converters has 255 steps then the resistance tolerances must be better than
(a) +0.04% (b) +0.01%
(c) +0.03% (d) +0.40%
9. A digital to analog converter uses a
(a) resistive ladder
(b) capacitive ladder
(c) inductive ladder
(d) none of the above
10. A data logger processes the readings to render immediately as recognizable
(a) voltage (b) current
(c) scientific units (d) power
11. Modern scanners of a data logger have input scanners which can scan at the rate of
(a) 100 inputs/s (b) 50 inputs/s
(c) 10 inputs/s (d) 150 inputs/s
12. Interlaced scanning in data loggers have
(a) higher scan rate
(b) lower scan rate
(c) mixed scan rate
(d) same scan rate
13. The outer band of a 3 binary code wheel has
(a) 8 segments (b) 4 segments
(c) 2 segments (d) 6 segments
14. For a pure binary code in a shaft encoder, the innermost track is
(a) quartered
(b) halved
(c) divided into 8 parts
(d) divided into 6 parts

Practice Problems

1. What is the binary equivalent weight of each bit in a 4-bit resistive ladder?
2. Draw the schematic of a 4-bit resistive ladder?
3. For a 4-bit resistive divider, determine the following:
(i) The change in output voltage due to change in the LSB, 2nd LSB
(ii) The output voltage for a digital input of 1100 and 1110
(Assuming 0 = 0 V and 1 = + 10 V)
4. What is the binary equivalent weight of each bit in a 5-bit resistive ladder?
5. Draw the schematic of a 5-bit resistive ladder?
6. Assuming the divider in Problem 5 has +10 V full scale output, then find the following

(i) The change in the output voltage due to change in the LSB
(ii) The output voltage for an input of 11001

7. Assuming the divider in Problem 5 has +5 V full scale output, then find the following:
   (i) The change in the output voltage due to change in the LSB
   (ii) The output voltage for an input of 10011
8. A 6-bit resistive divider is constructed such that the current through the LSB is 100 μA. Determine the maximum current that will flow through the MSB resistor.
9. A 4-bit resistive divider is constructed such that the current through the LSB is 50 μA. Determine the maximum current that will flow through the MSB resistor.
10. For a 4-bit binary ladder, if the input levels are 0 = 0 V and 1 = +10 V. What are the output voltages for each bit?
11. If V_{in} = 5 V, R_{in} = 5 kΩ R_F = 1 kΩ, calculate the output voltage V_{out} referring to Fig 17.18(a).
12. If V_{ref} = 5 V, R = 2 kΩ R_F = 1 kΩ, calculate the analog output current for various binary input referring to Fig 17.18 (c).
13. Find the output voltage of a 6-bit binary ladder with the following inputs
    (i) 110001 (ii) 010101
    (iii) 110011 (iv) 110110
    (v) 111111
14. What is the full-scale output voltage of a 5-bit binary ladder if the full-scale output is +10 V?
15. What is the full-scale output voltage of an 8-bit binary ladder if the full-scale output is +10 V?
16. What is the resolution of a 10-bit D/A converter which uses a binary ladder? If the full scale output is + 10 V, what is the resolution?
17. How many bits are required in a binary ladder to achieve a resolution of 1 mV, if the full scale is +5 V?
18. Calculate the degree of resolution that can be obtained using a 10-bit optical encoder.

Further Reading

1. Ernest O. Doebelin, *Measurement Systems: Applications and Design*, 4th Edition, McGraw-Hill, 1990.
2. Rangan Sharma and Mani, *Instrumentation Devices and Systems*, Tata McGraw-Hill, 1985.
3. R.K. Jain, *Mechanical and Industrial Measurements, 5th Edition, Khanna Publishers, 1983.*

Data Transmission

Chapter 18

INTRODUCTION 18.1

A data transmission system can be described simply in terms of three components, the transmitter (also called the source), the transmission path (usually referred to as the channel, but sometimes as the line), and the receiver (usually called the sink).

It is easier to think of a data transmission system between points *A* and *B* (Fig. 18.1) in terms of the universal seven part data circuit, which consists of the following:

1. The data terminal equipment/transducer (DTE) at point *A*
2. The interface between the DTE and the data circuit terminating equipment (DEC) at point *A*
3. The DCE at point *A*
4. The transmission channel between points *A* and *B*
5. The DCE at point *B*
6. The DCE-DTE interface at point *B*
7. The DTE at point *B*

The DTE is the source, the sink or both in the system. It transmits and/or receives data by utilising the DCE and data transmission channel. The DTE could be a CRT or teletype terminal, transducer, PC, printer, processor for large main frame computers, or any other device that can transmit and receive data.

The whole purpose of data transmission system is to transmit useful information between points *A* and *B*, the information may be used directly by the DTE or the DTE may process or display the information for use by a human operator.

The DCE and the transmission channel perform the function of moving the data from points *A* to *B*. In general, they do not know the content of the information transmitted.

The data transmission system is concerned only with the correct transmission of the information given to it, and system does not operate on the content of the information at all. The information received is identical to the information transmitted.

Protocol is the name given to the hardware and software rules and procedures for making sure that transmission errors are detected. The basic elements of a communications protocol are a set of symbols, called a character set, a set of rules for the sequence and timing of messages constructed from the character set, the procedures for determining when an error has occurred in the transmission and how to correct it. The character set consists of a subset which is the signal information (usually called printing characters), and another subset which conveys control information (usually called control characters). There is a correspondence between each character and a group of symbols on the transmission channel.

The set of rules to be followed by the sender and receiver gives the meaning, permissible sequence, and time relationships of the control characters and messages formed from the symbols. The error detection and correction procedure allows for the detection of, and orderly recovery from, errors caused by factors outside the control of the terminal at either end.

The protocol may be as simple as transmitting an extra bit of information in each character to detect errors, or some complex system.

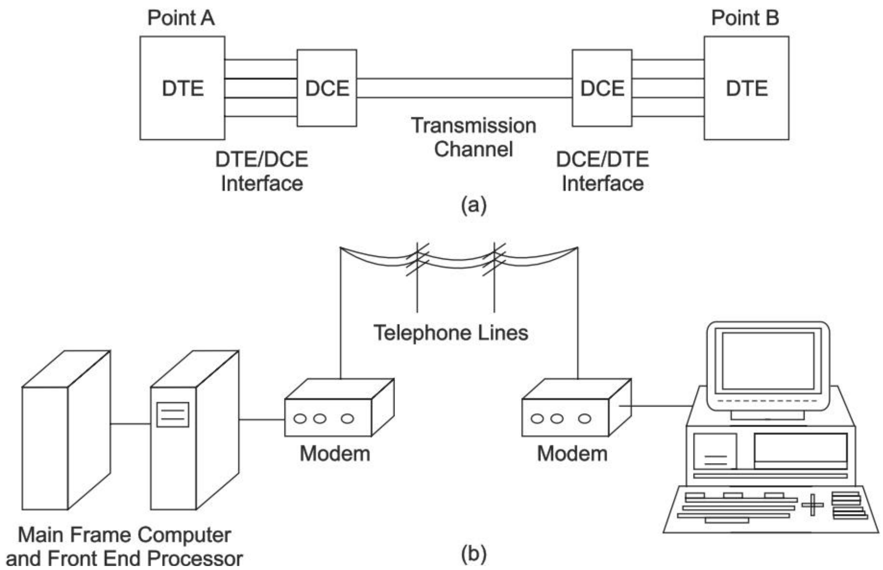


Fig. 18.1 Data transmission system (basic) (a) block diagram (b) pictorial

The basic concept of data transmission is the transfer of information from one location to another. In contrast to analog signals which have an infinite number of voltage levels, digital signal information is binary (having two levels). A binary level is called a Bit. A selective arrangement of seven bits provide 2^7 (128) distinct character combinations, or 128 bytes. The ASCII code is an excellent example of such an arrangement. It is used for the transmission of numbers, alphanumeric characters and miscellaneous information.

For uniform flow of data, the following important points to be agreed upon by the sender and the receiver.

1. The nominal rate of transmission (or how many bits are to be emitted per second by the sender).

2. The specified information code, providing a one to one mapping ratio of the information to the bit pattern and vice versa.
3. A particular scheme by which each bit can be positioned properly, within a byte, by the receiver of the data (in the case of bit-serial transmission).
4. A protocol (handshaking) sequence necessary to ensure an orderly flow of information.

DATA TRANSMISSION SYSTEMS 18.2

A typical data transmission system contains the information source from which data has to be transmitted through a suitable medium to the destination, termed the information sink. The information source can be a computer terminal, a digitized transducer output, or any other device generating a stream of bits at the rate of one bit every t_B seconds. The information rate of the system is therefore $1/t_B$ bits/s. As shown in Fig. 18.2, the information source feeds to a source encoder which performs logic operations on the data, on the associated clock, and perhaps on past data bits as well. Hence the source encoder produces a stream of data controlling the line drivers. The line driver interfaces the internal logic levels of the sources (TTL, MOS, etc.) with the transmission line. The transmission line in turn carries the signal produced by the line driver to the line receiver. The line receiver makes the decision on the signal logic state by comparing the received signal to a decision threshold level, and the sink decoder performs the logic operations on the binary bit stream recovered by the line receiver. The recovered binary data passes to the information sink, which is the destination for the information-source data.

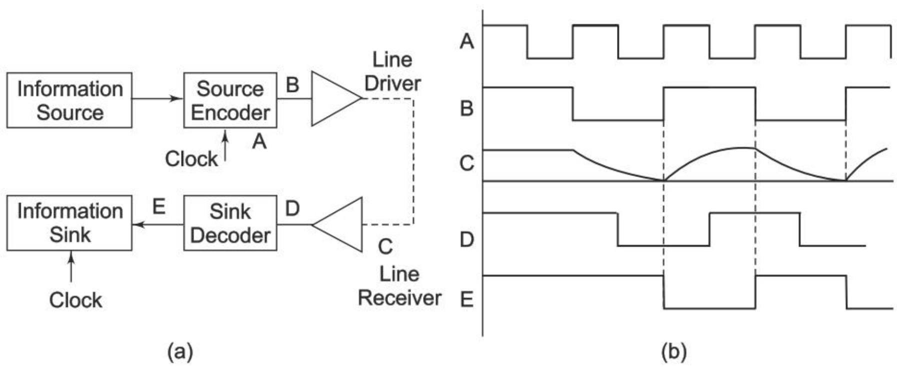


Fig. 18.2 (a) Block diagram of a data transmission system
(b) Shows the waveform at different points

ADVANTAGES AND DISADVANTAGES OF DIGITAL TRANSMISSION OVER ANALOG 18.3

Digital communication has several distinct advantages over analog transmission systems, but there are definite disadvantages to it as well. Because of the heavy investment in analog facilities, it will be many years before digital becomes the major technique in communication system. In fact, if it were not for the many analog earth stations already in existence, satellite communications would probably have been almost wholly digital by now. The following lists provide a good summary of the advantages and disadvantages of digital communication.

Advantages

1. Transmission quality is almost independent of the distance between terminals. The error rate is virtually unaffected by distance due to the regeneration, or cleaning up, of the signal at the regenerators.
2. A mixture of traffic, ranging from telephony and telegraphy to data and video information can be easily carried.
3. The capacity of certain existing transmission systems can be increased. For instance, the capacity of a single channel over a cable pair can be increased by employing digital multiplexing equipment, whereby several signals are combined into a higher rate signal without any significant impairment. Similarly, the capacity of a satellite communication system can be increased by going to TDM access.
4. Digital cable systems are more economical for distances of 15–40 km, particularly if cable capacity is nearly exhausted. It is more economical to convert to digital on the old cable rather than by laying new cables or provide new ductways.
5. Digital communication lends itself to such novel facilities as cryptography, storage and other forms of digital processing.
6. Digital communication is more suitable for the newer types of transmission media, such as light beams in optical fibers and circular or helical wave guides operating around 30 to 100 GHz.
7. Digital signal characteristics are convenient for electronic switching, in which groups of digits are selected to be switched in turn onto various highways. This is called packed switching.

Disadvantages

1. One of the major disadvantages of a digital system is its large band width requirement. An audio channel, which may normally require an analog band width of 4 kHz, will require 64 kHz when transmitted by a PCM (digital) system. This can be reduced by signal processing and the employment of special modulation techniques, but at the expense of increased costs and complexity.
2. Time division digital transmission is not compatible with the frequency division analog transmission in current use. Both cannot be carried simultaneously, as can be done, for instance, with mono and stereo FM.

Because of the basic incompatibility, however, it is very difficult to gradually change over from analog to digital transmission. Unless digital systems become highly economical or demand becomes high, it will be some time before an international satellite digital system becomes viable.

TIME DIVISION MULTIPLEXING (TDM) 18.4

Any pulse modulation scheme involves translating the audio, or modulating signal into a series of encoded pulses, sending these pulses over a transmission medium and reverting the pulses back to an analog signal. Regardless of the encoding used, a PAM signal is always obtained initially.

The audio signal is first sampled at a sufficiently high rate in order to minimise distortion. By keeping the samples short in comparison to the time intervals between them, many different channels can be interleaved in one sampling period. This process, as performed by a commutator, is illustrated in Fig. 18.3. It is commonly called Time Division Multiplexing (TDM).

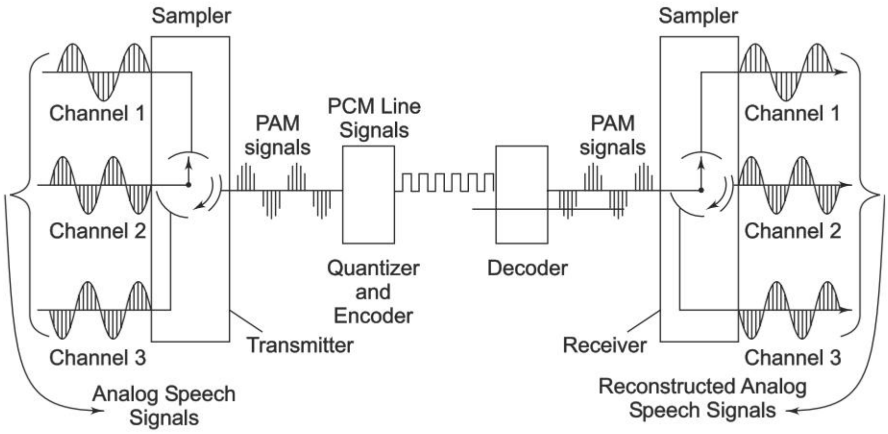


Fig. 18.3 Simplified diagram of TDM system

TDM allows each channel the full band width of the transmission medium whenever its signal is transmitted, although each channel is not continuously on the system. This is in contrast to Frequency Division Multiplexing (FDM), in which each channel is continuously on the system but allowed only a limited portion of the system band width. This is clearly illustrated in Fig. 18.4.

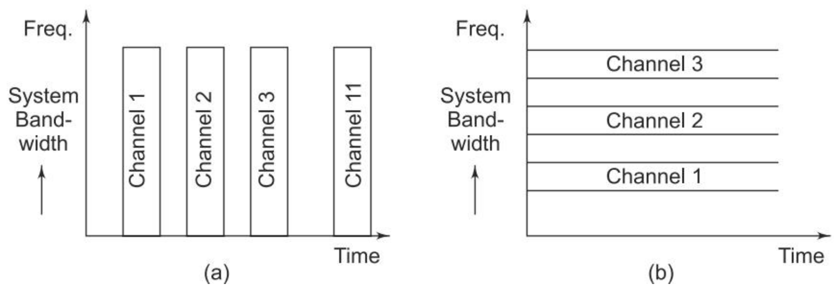


Fig. 18.4 Bandwidth utilisation of TDM and FDM system (a) TDM (b) FDM

After a PAM signal is obtained, it can be transmitted directly or quantized and encoded to provide improved noise immunity. The latter is more commonly done.

An example of a 4 channel electronic sampling system is shown in Fig. 18.5. This system employs 2 F/Fs and MOS p-channel enhancement transistors as switches.

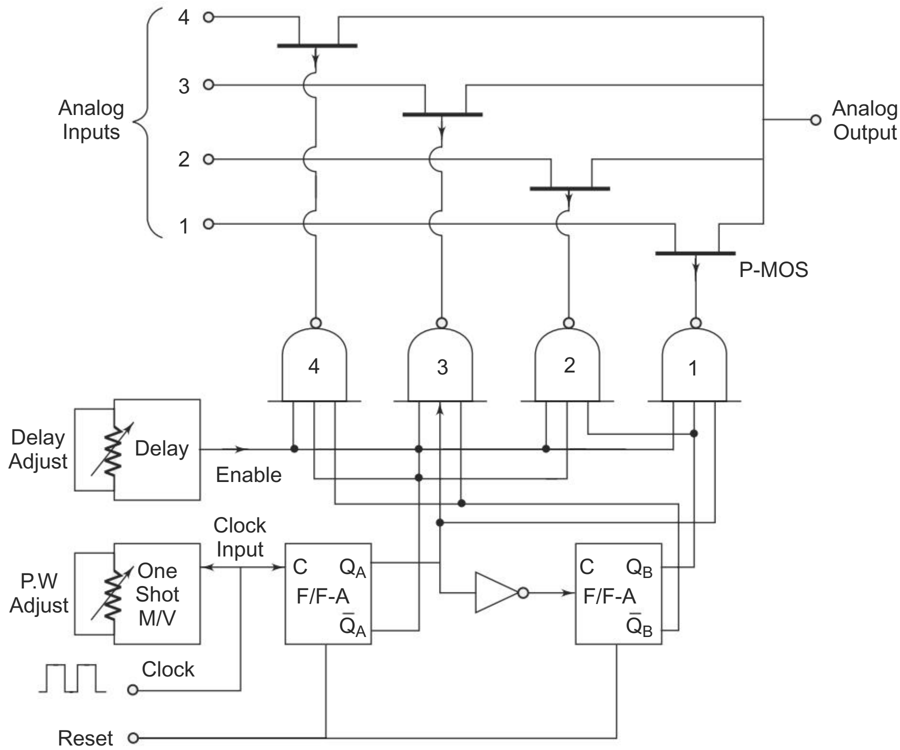


Fig. 18.5 4 channel electronic sampling system

When the gate to the MOS p-transistor becomes negative, the p-channels act as a low impedance circuit, with a resistance of about 200 Ω.

F/Fs A and B convert the serial clock input to binary parallel output. The four NAND gates decode this to sequentially drive the switches. When all the inputs to a NAND gate go high, the output goes low, thus causing the attached transistor to act as a low resistance path.

In order to establish a variable sampling aperture, the clock pulse is fed to a monostable multivibrator with pulse width adjusting control. This establishes the aperture time. A delay circuit then moves the sampling pulses such that the sampling occurs some time away from the gates switching edge.

Table 18.1 gives the F/F conditions for the channel selector.

Table 18.1

Q_A	Q_B	*Channel Selected*
Low	Low	4
Low	High	2
High	Low	3
High	High	1

PULSE MODULATION 18.5

18.5.1 Introduction

Pulse modulation may be used to transmit analog information, such as continuous speech or data. It is a system in which continuous waveforms are sampled at regular intervals.

Information regarding the signal is transmitted only at the sampling times, together with any synchronising pulses that may be required. At the receiving end, the original waveforms may be reconstructed from this information regarding the samples. The resulting output has negligible distortion.

Pulse modulation may be broadly subdivided into two categories, analog and digital. In the former, the indication of the sample amplitude may be infinitely variable, while in the latter a code which indicates the sample amplitude to the nearest predetermined level is sent.

Pulse amplitude and pulse time modulation are analog, while both pulse code modulation (PCM) and delta modulation are digital.

All modulation systems have sampling in common, but they differ from each other in the manner of indicating the sample amplitude.

In the PAM, the baseband signal modulates the amplitude of a pulse train which is spaced at regular intervals and has fixed time slots.

Rather than varying the pulse amplitude, alternative modulation schemes vary either the pulse intervals, called Pulse Position Modulation (PPM), or the duration of the time slots, called Pulse duration (Width) Modulation (PWM).

Figure 18.6 shows various analog pulse modulation waveforms for the same message signal.

In PDM, the pulse width is proportional to the amplitude of the modulating signal.

In PPM, the pulse delay from some reference point is proportional to the amplitude of the modulating signal.

In both PDM and PPM information is conveyed by a time parameter, or the location of the pulse edges. Thus, these modulation types are referred to as pulse time. In PAM and PDM, the sample value equals zero, usually represented by a non-zero amplitude or duration in order to prevent missing pulses and to preserve a constant pulse rate. This is important for synchronization purposes when time division multiplexing is used.

18.5.2 Pulse Amplitude Modulation (PAM)

Pulse Amplitude Modualtion (PAM) is the simplest form of Pulse Modulation. It is shown in Fig. 18.7.

PAM is a pulse modulation system in which the signal is sampled at regular intervals, with each sample proportional to the amplitude of the signal at the instant of sampling. The pulses are then sent either by wire or cable, or else used to modulate a carrier.

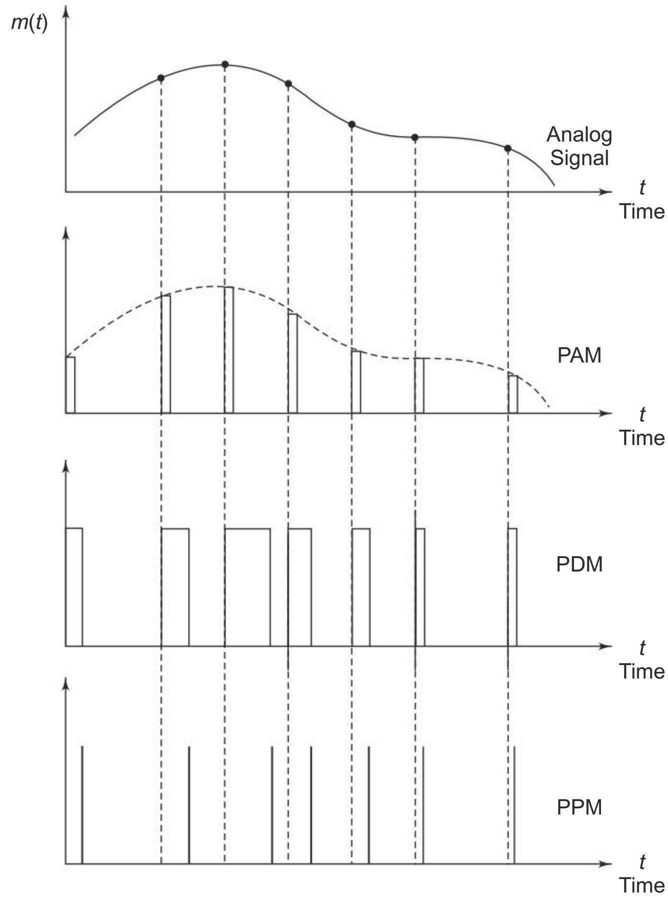


Fig. 18.6 Analog pulse modulation schemes

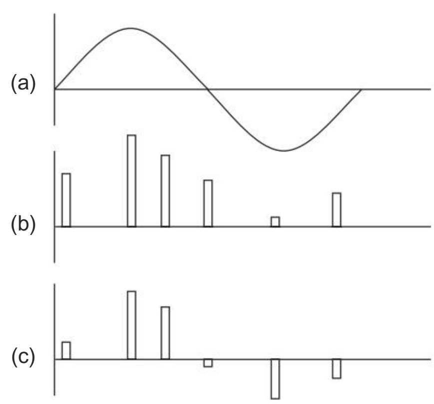


Fig. 18.7 Pulse amplitude modualtion (a) Signal (b) Single polarity PAM (c) Double polarity PAM

As shown in Fig. 18.7 the two types used are single polarity PAM and double polarity PAM.

In single polarity PAM, a fixed dc level is added to the signal to ensure that the pulses are always positive.

The ability to use constant amplitude pulses is a major advantage of pulse modulation. Since PAM does not utilise constant amplitude pulses, it is less frequently used. If it is used, the pulse frequency modulates the carrier.

It is very easy to generate and demodulate a PAM. In a generator, the signal to be converted to PAM is fed to one input of an AND gate. Pulses at the sampling frequency are applied to the other input of the AND gate, to open it during the wanted time intervals. The output of the gate then consists of pulses at the sampling rate, equal in amplitude to the signal voltage at each instant. The pulses are then passed through a pulse shaping network, which gives them a flat top.

Since FM is used in the receiver, the pulses are first recovered with a standard FM demodulator. They are then fed to an ordinary detector, which is followed by a low pass filter. If the cutoff frequencies of this filter are high enough to pass the highest signal frequency but low enough to remove the sample frequency ripple, an undistorted replica of the original signal is reproduced.

18.5.3 Pulse Time Modulation (PTM)

In PTM the signal is sampled in the same way as in PAM, but the pulses indicating instantaneous sample amplitudes have a constant amplitude themselves. However, one of their timing characteristics is varied, being directly proportional to the sampled signal amplitude at that instant.

The variable characteristics may be width, position or frequency of the pulses, so that three different types of PTM are possible.

Pulse frequency modulation has no significant practical applications, and is hence omitted.

The other two forms are discussed below.

All forms of PTM have an advantage over PAMs in that the pulse amplitude remains constant, so that amplitude limiters can be used to provide a good degree of noise immunity.

Sampling Theorem The sampling theorem states the following.

If the sampling rate in any pulse modulation system exceeds twice the maximum signal frequency, the original can be reconstructed in the receiver with minimal distortion.

This theorem is used in practice to determine the minimum sampling speeds, e.g. pulse modulation used for speech modulation. Transmission is generally over standard telephone channels, so that the audio frequency range is 300 to 3400 Hz. For this application, a sampling rate of 8000 samples/s is almost a worldwide standard. This pulse rate is, as can be seen, more than twice the highest audio frequency. The sampling theorem is satisfied, and the resulting system is free from sampling error.

Pulse position or pulse width measurements are usually based on the leading or trailing edge of the pulse. Usually the uncertainty of the pulse position is stated in terms of the rise and fall time of the pulse. An analysis of typical low pass filter shows that the rise time t_r is inversely proportional to the bandwidth. As a result, the band widths required for PDM and PPM are much greater than for PCM, where only the presence or absence of a pulse is of interest.

18.5.4 Pulse Width Modulation (PWM)

Pulse width modulation of PTM is also often called Pulse Duration Modulation (PDM). In this system, shown in Fig. 18.8, we have a fixed amplitude and starting time of each pulse, but the width of each pulse is made proportional to the amplitude of the signal at that instant.

In Fig. 18.8, there may be a sequence of signal sample amplitudes of (say) 0.9, 0.5, 0 and –0.4 V. These may be represented by pulse widths of 1.9, 1.5, 1.0 and 0.6 μs. In this system the width corresponding to zero amplitude can be selected as 1.0μs, assuming that the signal amplitude varies at this point between + 1 V (width = 2 μs) and – 1 V (width = 0 μs). Zero amplitude is thus the averaging signal level, which corresponds to the average pulse width of 1 μs.

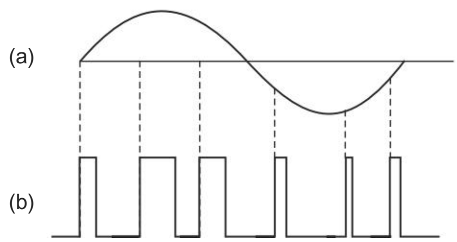


Fig. 18.8 Pulse width modulation (a) Signal (b) PWM

In this contrast, a negative pulse width is not possible, because it would make the pulse end before it began. If the pulse in a practical system has a recurrence rate of 8000 pulses/s the time of starting the adjoining pulses is $10^6/8000 = 125$ μs. This is enough to not only accommodate varying widths but also to permit time division multiplexing (10^6 = 1 MHz frequency clock).

Generation and Demodulation of PWM Pulse width modulation may be generated by applying trigger pulses (at the sampling rate) to control the starting time of pulses from a monostable miltivibrator, and feeding in the signal to be sampled to control the duration of the pulses. The circuit diagram for such an arrangement is shown in Fig. 18.9.

The emitter-coupled monostable multivibrator shown in Fig. 18.9 makes an excellent voltage to time converter, since its gate width is dependent on the voltage to which the capacitor *C* is charged. If this voltage is varied in accordance with a signal voltage, a series of rectangular pulses are obtained,with widths varying as required. The circuit performs two functions; it samples, and converts this sample in PWM.

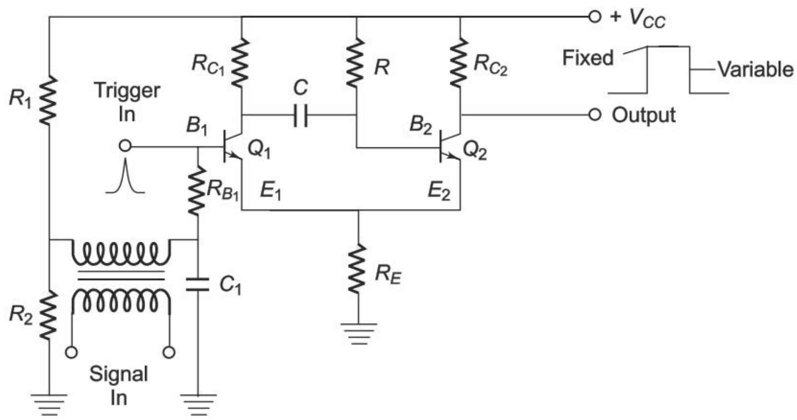


Fig. 18.9 Monostable multivibrator generating PWM

The operation of the circuit is as follows.

The multivibrator is in the stable state when Q_1 – OFF and Q_2 – ON. The applied trigger pulse switches Q_1 – ON, and thereby the voltage at C falls as Q_1 now begins to draw collector current. The voltage at B_2 falls and Q_2 is switched OFF by regenerative action. As soon as this happens, the capacitor C begins to charge up to the collector supply potential through R. After a time, determined by the supply voltage and the RC time constant of the charging network, B_2 becomes sufficiently positive to switch Q_2 ON. Q_1 is simultaneously switched OFF by regenerative action and stays OFF until the arrival of the next trigger pulse.

The voltage that the base of Q_2 must reach to allow Q_2 to turn ON, is slightly more positive than the voltage across the common emitter resistance R_E. This voltage depends on the current flowing through the circuit, which at the time is the collector current of Q_1 (which is then ON).

The collector current, in turn, depends on the base bias, which is governed by the instantaneous changes in the applied signal voltage. In order words, the applied modulation voltage controls the voltage to which B_2 must rise to switch Q_2 ON. Since the voltage rise is linear, the modulation voltage is seen to control the period of time during which Q_2 is OFF, that is, the pulse duration. This pulse duration is very short compared to even the highest signal frequencies, so that no real distortion arises through changes in signal amplitude while Q_2 is OFF.

To produce PDM, which is often called pulse width modulation, a circuit employing a 555 timer can also be used, as illustrated in Fig. 18.10(a).

When Pin 2 is high, comparator 2 sets the output Q_1 of the F/F high. This, in turn, turns Q_T ON and moves the buffer output to low. When a negative going clock pulse is applied to Pin 2, the output of F/F resets low. Q_T is cut off and C begins to charge linearly, as illustrated in Fig. 18.10(b). The buffer output remains high until the capacitor voltage, which appears at Pin 6, the input to comparator 1, reaches the value of the signal voltage which is applied to Pin

5 (input of comparator 1). This trips the output of comparator 1 and sets the F/F output high, the buffer output goes low and the capacitor rapidly discharges through Q_T. This cycle repeats upon application at the next pulse.

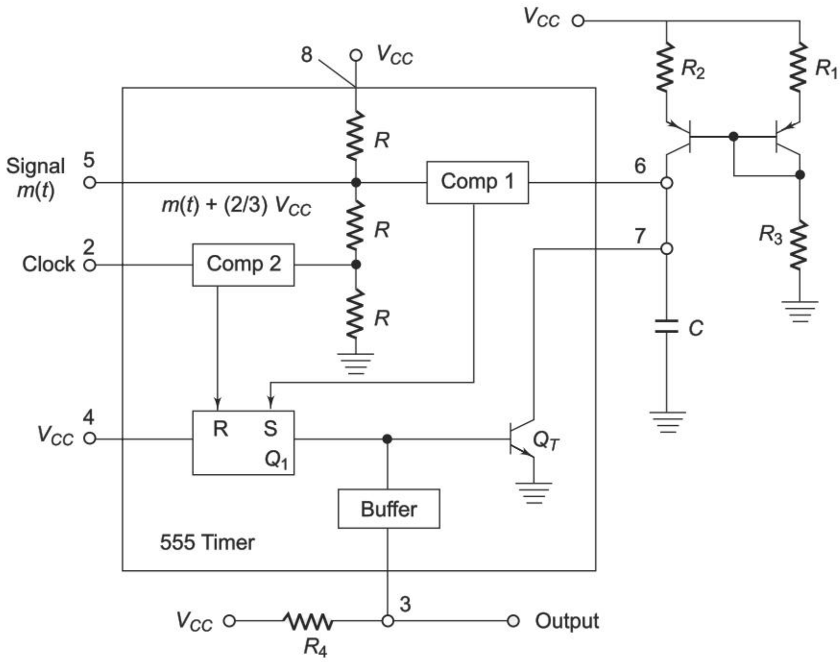


Fig. 18.10 (a) 555 timer generating PWM

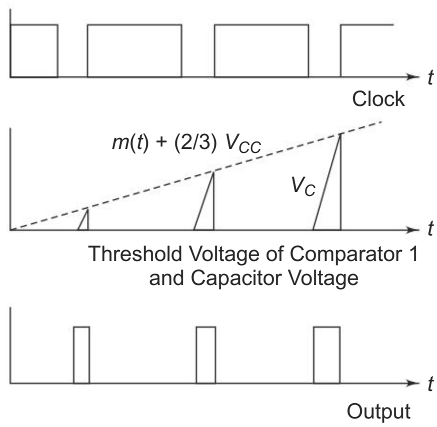


Fig. 18.10 (b) Waveform at different points of 555 timer

One method of transmitting PDMs more efficiently is to transmit only the pulse signals. Figure 18.11 shows the arrangement and the waveforms. At the receiving end, the appropriate pulse can be stretched to recover the original pulse.

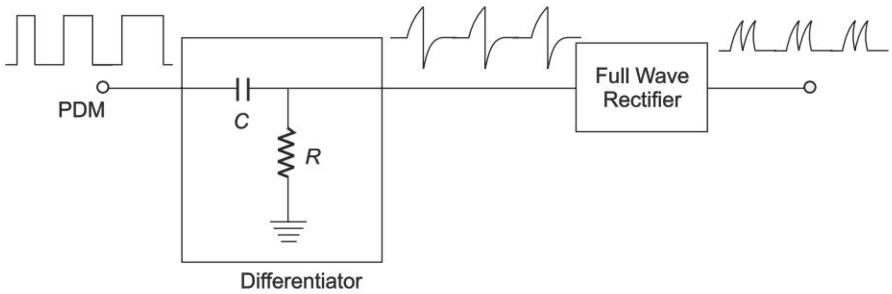


Fig. 18.11 Transmission of PDM using pulse edge only

PWM demodulation is a simple process. The result is fed to an integrating circuit from which a signal output is obtained, whose amplitude at any time is proportional to the pulse width at that time.

The circuit of Fig. 18.12 shows a PDM recovery circuit. It is assumed that the maximum duration of PDM is less than one third the signal period T_s. The one shot will set the F/F to zero only when two successive PDM pulses pass through with separations greater than $T_s/3$. This assumes that pulse 2 triggers the F/F to zero and that the next pulse that triggers the F/F corresponds to the leading edge of the transmitted pulse.

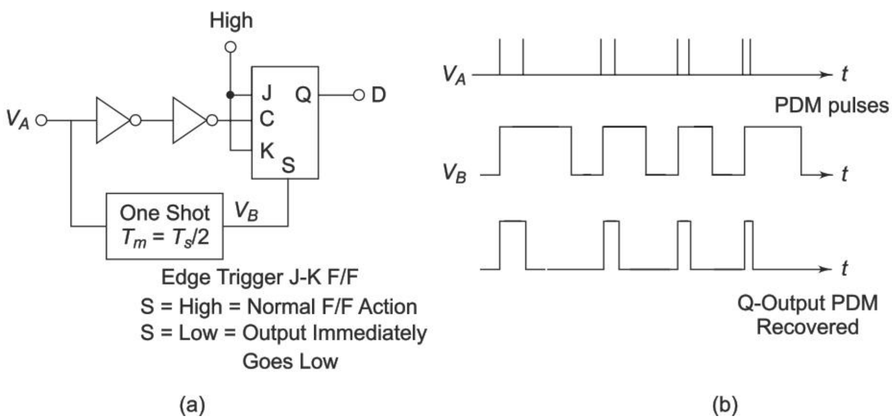


Fig. 18.12 (a) PDM recovery circuit (b) Waveform

In demodulating PDM, the best results are obtained when it is converted to PAM and low pass filtered. A general block diagram of a PDM or PAM is shown in Fig. 18.13.

The constant current generator assures a linear voltage rise on the capacitor. Note that the discharging pulses have to be located beyond the widest pulse width. The easiest way to obtain such pulses is to produce narrow pulses by triggering a multivibrator and delaying it with another monostable.

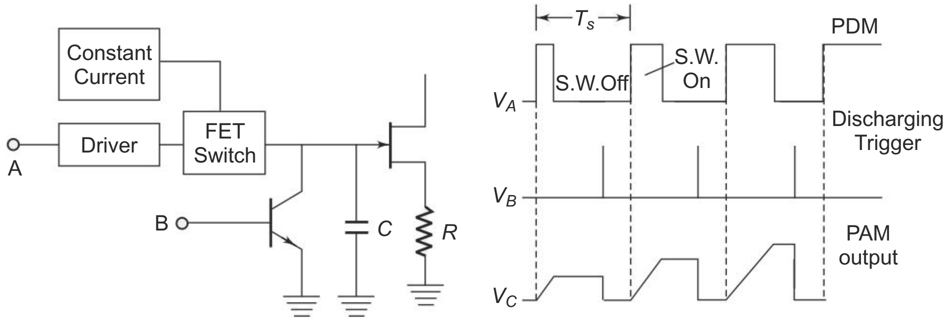


Fig. 18.13 Conversion of PDM to PAM

PPM is very closely allied with PDM. In fact, it can be derived directly from PDM, as shown in Fig. 18.14.

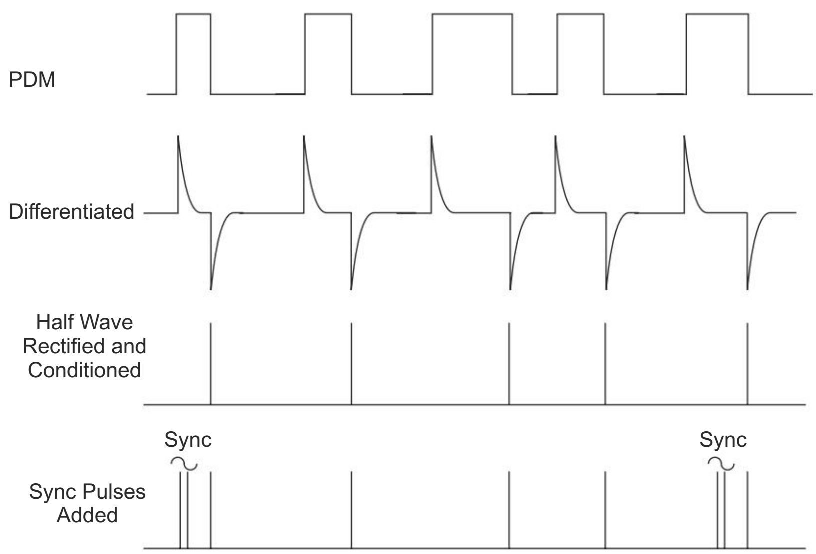


Fig. 18.14 Generation of PPM from PDM (Note: Sync pulses are double for recognition)

The PDM is differentiated, and then rectified and shaped. Remember that in PDM, it is the position of the pulse edges that carry the information and not the pulse itself. Therefore, PPM carries exactly the same information as long as the position of the clock pulses (leading edge) is well defined in the received signal.

It is easy to see that PPM is superior to PDM for message transmission, since the wide pulses of PDM require more energy than PPM when transmitted. Specifically, PPM is suited for communication in the presence of noise. Very high peak narrow pulses can be transmitted and the pulse position can be determined even when the noise level is high. Note however that transmitting very narrow pulses requires a large band width.

When light is used as the media for transmitting analog signals, PPM or PCM are the most suitable types of modulation because the maximum power output in the modulated light source, such as LED or LASER is achieved when it is pulsed at a very low duty cycle.

In transmitting PPM, it is necessary to transmit a series of sync pulses at a much lower repetition rate than the sampling pulses, so that they do not interfere with the original signal and/or minimise the number of pulses transmitted in order to conserve transmission power. In PPM, a series of double pulses are used for synchronisation, since the sync pulses must be narrow and at the same time distinguishable from the signal pulse.

18.5.5 Pulse Position Modulation (PPM)

In this system, the amplitude and width of pulses is kept constant, while the position of each pulse, in relation to the position of a recurrent reference pulse, is varied by each instantaneous sampled value of the modulating wave.

(PPM has the advantage of requiring constant transmitted power output, but has the disadvantage of depending on transmitter-receiver synchronism.)

PPM may be obtained very simply from PWM, as shown in Fig. 18.15.

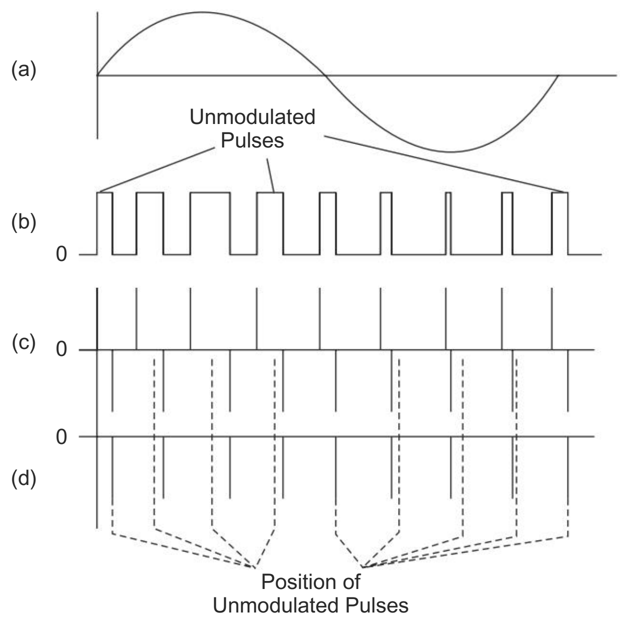


Fig. 18.15 Generation of pulse position modulation (a) Signal (b) PWM (c) Differentiated (d) Clipped (PPM)

Considering PWM and its generation again, it is seen that each pulse has a leading edge and a trailing edge. However, in PWM the locations of the leading edges are fixed while those of the trailing edges are not. Their position depends on pulse width, which is determined by the signal amplitude at that instant.

Hence, it can be said, that the trailing edges of PWM pulses are, in fact, position modulated.

Figure 18.15 shows PWM corresponding to a given signal. If the train of pulse thus obtained is differentiated, another pulse train results. This has positive narrow pulses corresponding to the leading edges and negative going pulses corresponding to trailing edges. If the position corresponding to the trailing edge of an unmodulated pulse (an unmodulated PWM pulse is one that is obtained when the instantaneous signal value is zero) is counted as zero displacement, then the other trailing edges arrive earlier or later. They therefore have a time displacement other than 0. This time displacement is proportional to the instantaneous value of the signal voltage. The different pulses corresponding to the leading edges are removed with a diode clipper or rectifier and the remaining pulses shown in Fig. 18.15(d) are position modulated.

When the PPM is demodulated in the receiver, it is again first converted into PWM with a F/F or bistable multivibrator. (One input of the multivibrator receives trigger pulses from a local generator which is synchronised by the trigger pulses received from the transmitter, and these triggers are used to switch OFF one of the stages of F/F.) The PPM pulses are fed to the other base of the F/F, and switch that stage ON (by switching the other one OFF). The period of time during which this particular stage is OFF depends on the time difference between the two triggers, so that the resulting pulse has a width that depends on the time displacement of each individual PPM pulse. The resulting PWM pulse train is then demodulated.

PWM has a disadvantage, when compared to pulse position modulation, in that its pulses are of varying power width and therefore of varying power content. This means that the transmitter must be powerful enough to handle pulses of maximum width (although the average power transmitted is only half the peak power).

DIGITAL MODULATION 18.6

The most common types of systems in which the encoded signal consists of binary digits are pulse code modulation, and delta modulation.

18.6.1 Pulse Code Modulation (PCM)

Pulse Code Modulation (PCM) is different from the other forms of pulse modulation studied so far, PCM also uses sampling techniques, but it differs from the others in that it is a digital process. Instead of sending a pulse train capable of continuously varying one of the parameters, the PCM generator produces a series of numbers or digits (hence the name digital process). Each one of these digits, almost always in a binary code, represents the approximate amplitude of the signal sample at that instant. The approximation can be made as close as desired.

Principle of PCM In PCM, the total amplitude range which the signal may occupy is divided into a number of standard levels, as shown in Fig. 18.16. Since these levels are transmitted in binary code, the actual number of levels is a power of 2, 16 levels are shown here for simplicity, but practical systems use as many as 128. By a process called quantitizing, the level actually sent by any sampling time is the nearest standard (as Quantum) level.

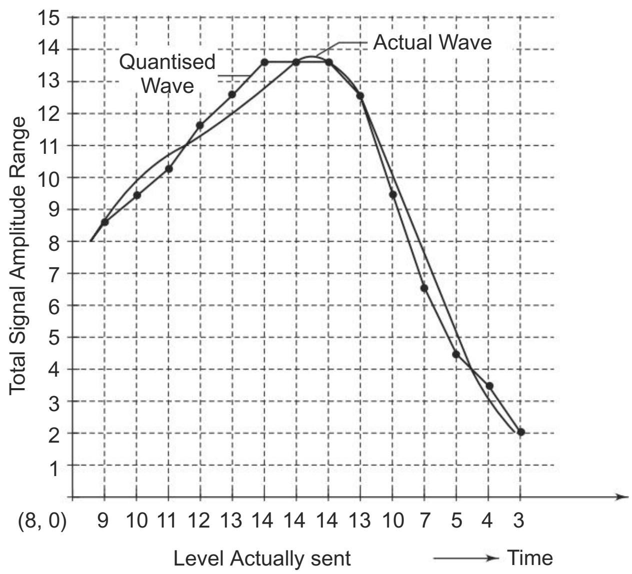


Fig. 18.16 Quantization of signal for PCM

As shown in Fig. 18.16, should the signal amplitude be 6.8 V at any time, it is not sent as a 6.8 V pulse, as it might have been in PAM, nor as a 6.8 μs wide pulse as in PWM, but simply as the digit 7, because 7 V is the standard amplitude nearest to 6.8 V. Furthermore, the digit 7 is sent at that instant of time as a series of pulses corresponding to the number 7. Since there are 16 (2^4) levels, 4 binary places are required; the number becomes 0111 and can be sent as 0PPP, where 0 = no pulse and P = pulse.

Actually, it is often sent as a binary number back-to-front, that is, as 1110 or PPP0, to make demodulation easier.

As shown in Fig. 18.16, the signal is continuously sampled, quantized, coded and sent, as each sample amplitude is converted to the nearest standard amplitude and into the corresponding back-to-front binary number. If sufficient quantizing levels are used, the result cannot be distinguished from that of analog transmission.

A signaling bit is generally added to each code group representing a quantized sample. Hence, each group of pulses denoting a sample here is called a word, expressed by means of $(n + 1)$ bits, where 2^n is the selected number of standard levels.

If the actual signal is compared with the signal in Fig. 18.16, it is seen immediately that the quantizing process introduces some distortion. This is quantizing noise, also called noise, because the errors are random in character. The randomness occurs simply because the difference between the digit sent and the actual signal at that instant is completely unpredictable, that is, random.

The largest error that can occur is equal to half the size of the sampling interval.

An obvious method of reducing quantizing noise is to increase the number of standard levels until the noise level becomes acceptable. However, more levels require more bits to sent them. The bandwidth required is proportional to the number of bits sent per second. In practical systems, 128 levels for speech are considered quite adequate.

Generation of PCM The generation of PCM is a complex process. Essentially, the signal is sampled and converted to PAM, the PAM is quantized, encoded and supervisory signals added. Two important terms need to be considered before understanding the generation of PCM.

1. Companding
2. PCM encoding

The quantized staircase waveform (Fig. 18.16) is an approximation to the original waveform. The difference between the two waveforms amounts to noise added to the signal by the quantizing circuit. The mean square quantization noise voltage has a value of

$$E_{nq}^2 = \frac{s^2}{n} \tag{18.1}$$

where s is the voltage of each step, or the subrange voltage span. As a result, the number of quantization levels must be kept high, in order to keep the quantization noise below some acceptable limit. For a sinusoidal signal which occupies the full range, the mean square signal voltage is

$$E_s^2 = \frac{1}{2} \times E^2 \text{ peak} = \frac{1}{2}\left[\frac{M \times s}{2}\right]^2 = \frac{(Ms)^2}{8} \tag{18.2}$$

where M is the number of steps and s is the step height voltage. The signal to noise ratio (S/N) for $n = 12$ bit is now given by

$$\frac{\text{Signal}}{\text{Noise}} = \frac{E_s^2}{E_{nq}^2} = \frac{(Ms)^2}{8} \times \frac{12}{s^2} = \frac{3}{2} \times M^2$$

The number of levels M is related to the number of bits per level (n) by $M = 2^n$. Therefore

$$S/N = \frac{3}{2} \times 2^{2n} \tag{18.3}$$

1. *Companding* The S/N ratio derived above was related to a maximum amplitude signal with a peak to peak voltage equal to the full signal range. In practice, the signal may be many times smaller than this (as

much as 30 db less). Since the noise level is dependent on the step size, which is a constant, with small signals a much lower S/N ratio will result. The process of companding is used to overcome degradation of the S/N ratio.

Companding is a compound process of volume compression before transmission, combined with volume expansion after transmission. By itself it does not produce any distortion in the recovered signal, but it does reduce the level of quantization noise during low signal level periods. The compressor amplifies low level signals more than it does high level signals, thus compressing the input voltage range into a smaller span. The steps transmitted have equal amplitude but are equivalent to smaller steps being used in low level signals and larger steps used in high level signals. The result is lower amplitude quantization noise during low level periods.

2. *PCM Encoding* The encoding process generates a binary number corresponding to the quantization level number to be transmitted for each sampling interval. Any one of the codes, such as ASCII or Bi-quinary may be used, as long as it provides sufficient number of symbols to represent all the levels to be transmitted.

   Ordinary binary coding, in which the binary number corresponding to the decimal number of the level in question is transmitted and most often used. This binary number contains a train of 1 and 0 pulses with a total of $\log_2 N$ pulses in each number (N is the number of levels in the full range). This system is very economical to realise, because it corresponds exactly to the process of A/D conversion.

PCM System The block diagram of a PCM system is shown in Fig. 18.17.

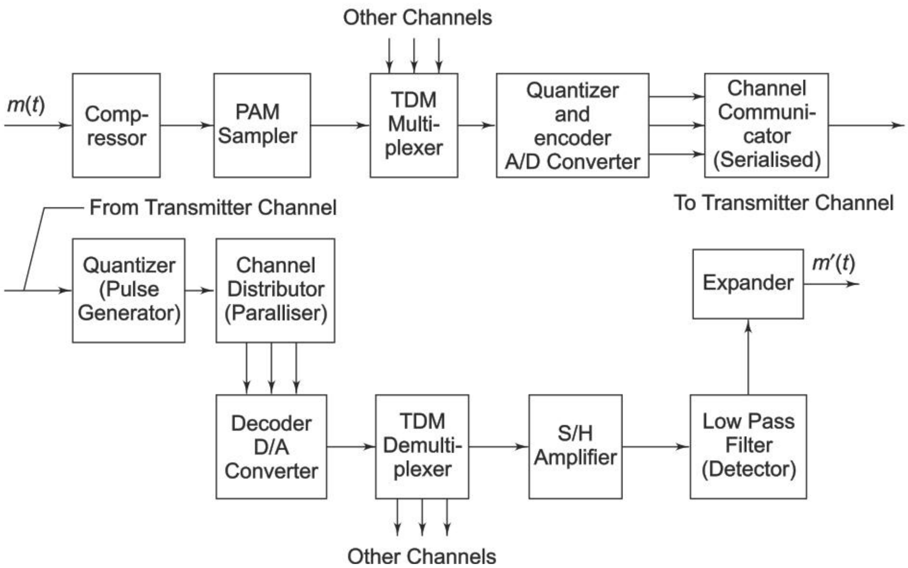


Fig. 18.17 PCM transmission system

The modulating signal $m(t)$ is applied to the input of the volume compressor unit. A sampling circuit generates a PAM signal from the compressed signals, which is then multiplexed with signals from other input channels. An A/D converter performs the two functions of quantization and encoding, producing a binary coded number for each channel sampling period. A commutator circuit transmits the code bits in serial form.

At the receiver, a two level quantizing circuit reshapes the incoming pulses and eliminates most of the transmission noise. A distributor circuit decommutates the pulses and passes the bits in parallel groups to a D/A converter for decoding. Another distributor demultiplexes the several PAM signals and routes them to the proper output channels. Each channel has an S/H amplifier which maintains the pulse level for the duration of the sampling period, recreating the staircase waveform approximation of the compressed signal. A low pass filter may be used to reduce the quantization noise and an expander circuit removes the amplitude distortion which was intentionally introduced in the compression of the signal to obtain the output signal $m'(t)$.

The Bell T1PCM system is an example of such an operational system. In this system, 24 voice channels are multiplexed, the sampling frequency being 8 kHz.

The channeling scheme is shown in Fig. 18.18. The frame period is the time interval between successive samples of a given channel, and as seen from Fig. 18.18, the frame period is equal to the sampling period, that is, $t_s = 1/f_s = 1/8\text{ kHz} = 125\ \mu\text{s}$.

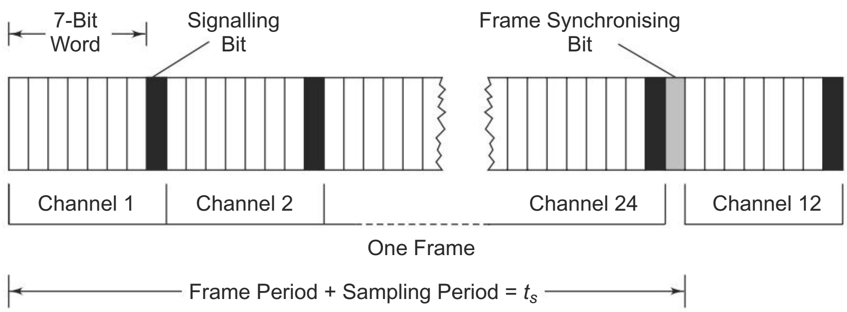


Fig. 18.18 Channelling scheme for bell T1 PCM 24 channel system

Each channel is coded into 8 bits, which includes signaling information. Therefore, in one frame period the 24 channel requires $24 \times 8 = 192$ bits.

One frame synchronizing bit is also required in each frame period. Therefore, the total number of bits in a frame is 193. The signalling rate is

$$r = \frac{193}{125\ \mu\text{s}} = 1.544\text{ Mbits/s}$$

The most common form of digital transmission is PCM. PCM performs three functions of (i) sampling the analog signal, (ii) Quantizing the sampled amplitudes, and (iii) encoding the quantized sample into a digital system.

Applications The principle area of PCM application has been in linking relatively close terminals or nearby local exchanges by wire or cable. In due course, PCMs may be employed in very large networks spanning great distances. Figure 18.19 shows a basic block diagram of a PCM speech terminal.

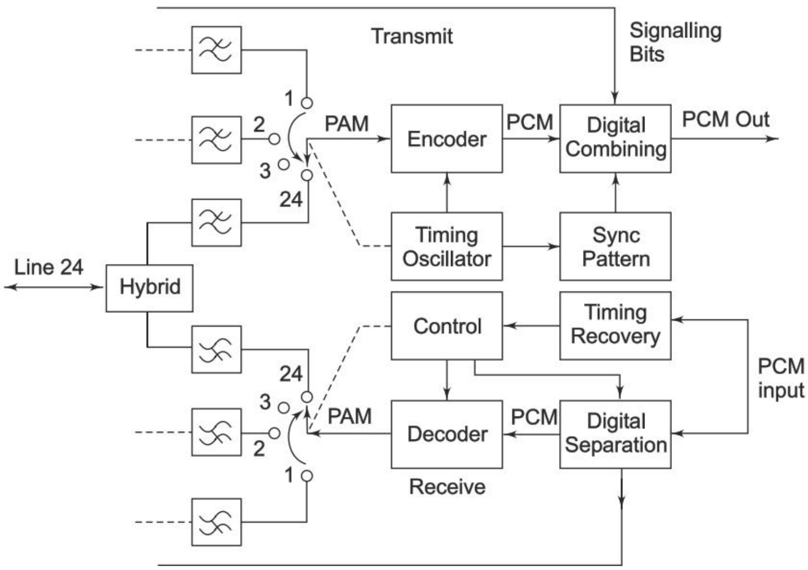


Fig. 18.19 Block diagram of a PCM speech terminal

When transmitting, the audio signal is first band limited to minimise aliasing or foldover distortion and then sampled to form a PAM signal. At this point, the signals from all channels go to a common equipment where coding and addition of sync is done. Since each individual channel requires very little equipment, and the more exotic switching and coding is done by common circuitry, PCM terminals can be inexpensive for a large number of channels.

After sampling, the voltage amplitude is quantized, or forced to take on certain discrete voltages. This new signal forms an approximation to the old one. Because the two forms are not identical, an inherent error occurs. The difference between the old and new signal is called the quantizing error, and is referred to as quantizing noise when heard and measured. The error signal has frequency components extending from above the voice band. Because of aliasing, these higher frequency components fold back into the voice band and result in an error signal at received filter output. In PCM transmission, this quantizing noise is one of the dominant sources of impairment and it remains significant even when other noise contribution are minimised.

Advantages of PCM In addition to PCMs immunity to noise, which gives transmission characteristics almost independent of distance, the following factors also favour this mode of modulation.

1. It can easily carry a mixture of traffic, such as telephony, telegraphy, data and encoded video information, provided the medium has sufficient capacity.
2. It can increase the capacity of single telephone channels over cable pairs by multiplexing.
3. It can lend itself to such novel facilities as cryptography, storage and other forms of digital processing.
4. It is more suitable for the newer types of transmission media, such as light beams in optical fibers and multiple access satellites.
5. Its signal characteristics allow easy access to electronic switching in which groups of digits are selected to be switched in turn onto various highways.

18.6.2 Delta Modulation (DM)

Delta modulation (DM) is process of modulation in which train of fixed width pulses is transmitted. Their polarity indicates whether the demodulator output should rise or fall at each pulse. The input is caused to rise or fall by a fixed step height at each pulse. Figure 18.20 shows the block diagram of a DM system.

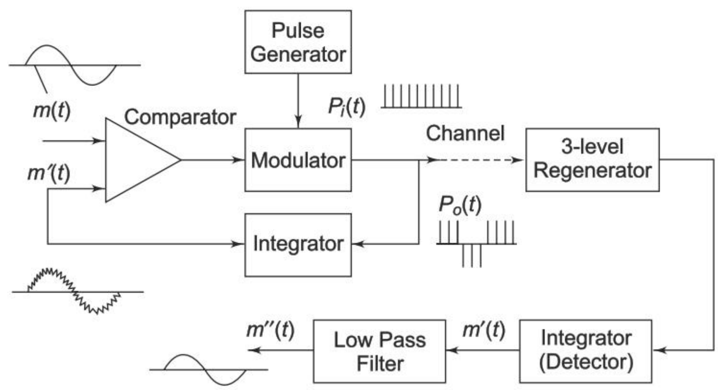


Fig. 18.20 Delta modulation (DM) transmission system

In the DM system, the modulating signal $m(t)$ is applied to the non inverting input of a high gain differential comparator. A reconstructed signal version of this signal, $m'(t)$, is applied to the inverting input. The difference comparator is in saturation, either positive or negative, depending on the polarity of the difference voltage between the input signals. Hence, the output is $\pm$ 1.

The modulator receives a train of unipolar pulses [$P_i(t)$] recurring at the desired sampling rate and either transmits directly for (a + 1) inputs or inverts their polarity for (a – 1) inputs. This signal is transmitted as the output signal [$P_o(t)$] and is also passed to a local integrating circuit. This integrator causes $m'(t)$ to rise or fall by a fixed step height for each positive and negative pulse applied to its input. The height of the pulse can be adjusted by changing the gain factor of the integrator. An increase causes the step waveform to have a higher maximum rate of rise (R_{max}).

At the receiver, a regenerator reshapes the received signal and removes most of the noise. The signal is then fed to another integrator which reconstructs $m'(t)$, the step waveform. This is then passed through a low pass filter to remove the quantization noise, leaving a replica $m''(t)$ of the original signal.

Delta modulation is less frequently used than PWM and PCM. Currently, it is being applied to satellite communications, TV and subscriber telephone loops. It is less complex and therefore less costly than PCM, more tolerant to transmission errors and does not need the synchronization requirements of PCM. On the other hand, it is sensitive to slope overload and is unsuitable for timesharing as an encoder/decoder among multiple channels.

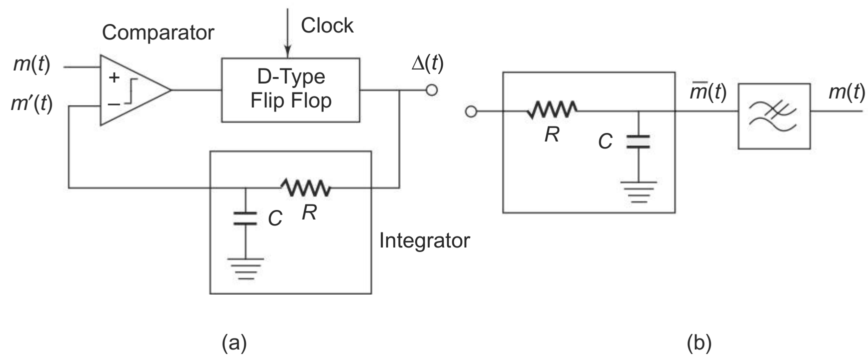


Fig. 18.21 Delta modulator (a) Encoder (b) Decoder

The delta modulator transmits binary output pulses whose polarity depends upon the difference between the modulating signal and the feedback signal corresponding to the history of the signals previously sent. In the simple DM of Fig. 18.21, the integrated feedback $m'(t)$ is compared to the input signal $m(t)$. If $m(t) > m'(t)$, a positive pulse forms the output Δt, and if $m(t) < m'(t)$, a negative pulse forms the transmitted signal. This output pulse train also forms the approximation $m'(t)$, after being integrated.

The following waveform sketches will clarify the operation of this modulator. As illustrated in Fig. 18.22, a series of triangular waveform results for $m'(t)$ when there is no signal applied to the delta modulator. This is because integration of a constant gives a ramp. In this idling condition, the transmitted signal consists of alternate positive and negative pulses. The difference between the original signal, $m(t)$ and the reconstructed signal $m'(t)$, result in an error signal, which is often called granular or quantization noise. This noise can be decreased by either decreasing the magnitude of the step size a or by increasing the sampling frequency.

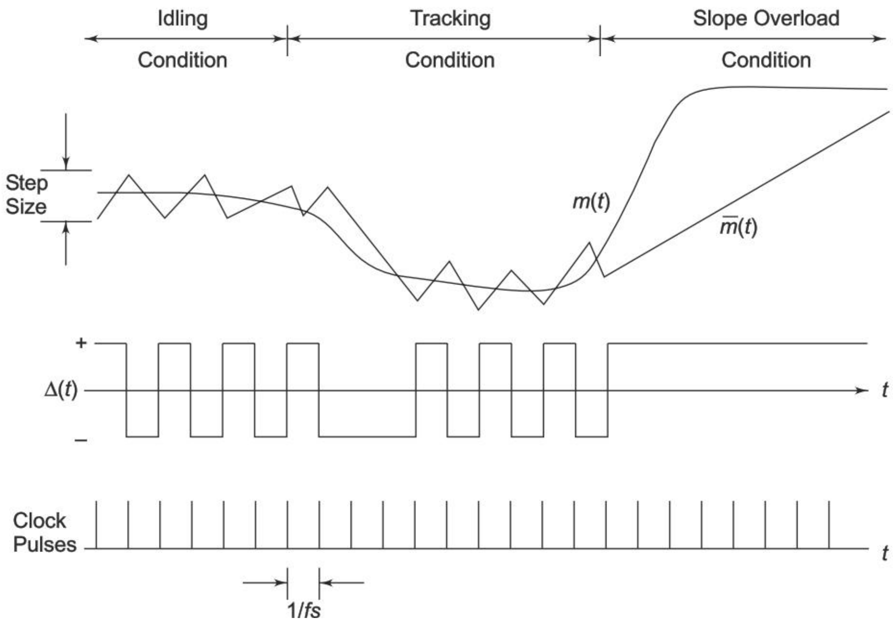


Fig. 18.22 DM waveforms illustrating idling, tracking and slope overload

If $m(t)$ exceeds the feedback signal $m'(t)$, the transmitted wave train remains positive. When $m'(t)$ begins to overshoot $m(t)$, the encoder output begins transmitting negative pulses. If the modulating signal changes more rapidly than the encoder can follow, slope overload occurs. This slope overload, shown towards the right of Fig. 18.22, causes waveform distortion and is one of the severe limitations of DM. The integrator is unable to closely track large amplitude, high frequency signals.

The difference between $m'(t)$ and $m(t)$ in slope overload is called the slope overload noise. This noise greatly exceeds quantization noise. The S/N ratio, as a function of input signal power, initially increases with signal power until slope overload is reached as illustrated in Fig. 18.23. This increase is due to the signal power term increasing while the ω-term, chiefly quantization noise, remains fixed. Upon a further increase in signal power, slope overload occurs and the S/N ratio suffers.

Companded Delta Modulation Studying the S/N response of the simple DM scheme shown in Fig. 18.23, shows that the S/N ratio varies considerably with signal power. At low signal power, the S/N ratio is much worse than at high signal power, as long as the signal remains within the quantization noise region.

In order to maintain a more constant S/N ratio, the step size is made adaptable to the input signal amplitude. For small inputs $m(t)$, the step size is kept small, and increased as the input signal is increased. This makes the noise power vary with signal power, maintaining a constant S/N ratio over a fairly large range of input power levels. In addition, slope overload is less likely to occur since the step size are the largest when nearing the region of constant S/N ratio.

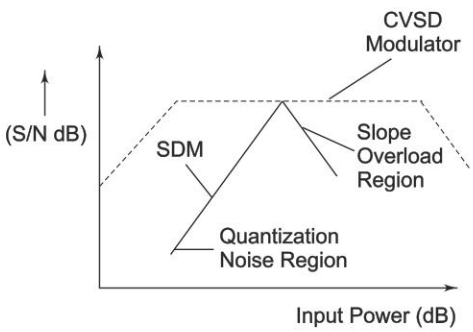


Fig. 18.23 S/N for simple and adaptive DM

For telephony, the type of adaptive delta modulation used is a continuous variable slope DM (CVSD). A block diagram of such a modulator is shown in Fig. 18.24. The squaring circuit causes the gain of the feedback amplifier to increase with increase in signal amplitude, regardless of the polarity of the voltage. In addition, by squaring the quantization noise, power is made to vary linearly with input signal variations.

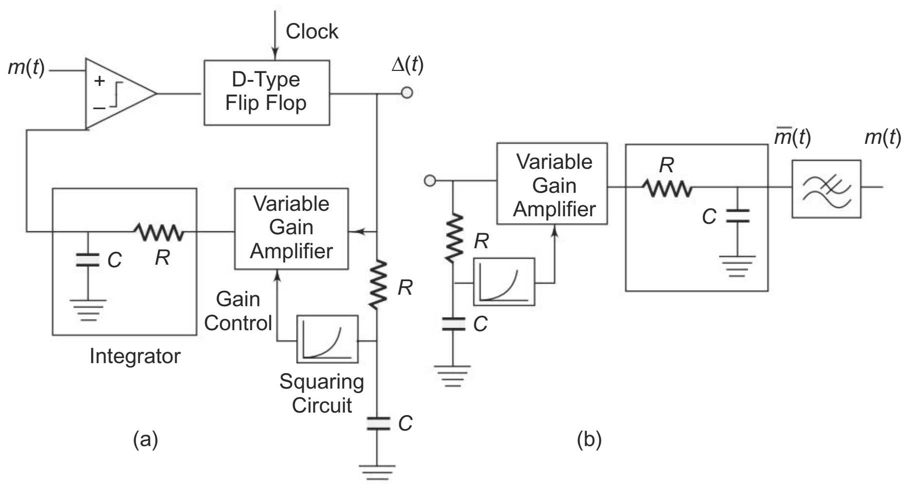


Fig. 18.24 Variable slope delta modulator (a) Encoder (b) Decoder

PULSE CODE FORMAT 18.7

There are many different code formats of pulse waveforms. The classification is based on three criteria, namely, form of information transmission, relation to zero level, and direction.

Based on the form of information transmission, the format used can be any one of the following

1. Full binary transmission, where both 0 and 1 bits are part of the format.
2. Half binary transmission, where only the 1 are transmitted. The 0 is recognised by the absence of a pulse at the time of clock transition.
3. Multiple binary transmission, where ternary and quadratic codes are used for each transmitted pulse.

Based on its relationship to the zero level, the transmission format can be either Return to Zero (RZ) in which there is a return to zero level after the transmission of each bit of information, or Non return to Zero (NRZ), where there is no voltage level change if consecutive bits are transmitted, although there is a level change when there is an information variation from 0 to 1 or 1 to 0. In the case of the third criteria, that is, direction, the code format used can be either unipolar, where the pulses are in a single direction, or bipolar, where the pulses are in both directions. Some of these code formats are shown in Fig. 18.25.

18.7.1 Full Binary Transmission

The full binary bipolar return to zero (RZ) format shown in Fig. 18.25 (a) is one of the most reliable pulse code formats employed for slow speed transmission. The speed is typically up to 600 bits/s. Using frequency shift keying (FSK), opposite polarity pulses are used to transmit "1" and "0" bits. No power is transmitted between the pulses, resulting in a space between each pulse.

In the non-return to zero (NRZ) unipolar type of transmission, shown in Fig. 18.25(b) the pulses are spread out in time so that they occupy the full time slot and permit an increasing rate of transmission. The term NRZ is directly related to the fact that the pulses do not return to zero between successive pulses of 0 and 1. This format is most popularly used in serial computer applications and for data transmission speeds of 600, 1200 and 2400 bits/s. The transmission band width is efficiently utilised, since the entire bit period contains signal information.

The RZ unipolar format, shown in Fig. 18.25 (c), has symmetrical format characterised by the absence of a dc level. The pulse format has a zero crossing for each bit period, a feature that simplifies synchronization, at the expense of increasing the total band width required. This code is also referred to as a split phase, biphase or Manchester code.

18.7.2 Half Binary Transmission

In the half binary transmission system, the binary 1s are represented by a pulse or a polarity change, but the 0s are seen as spaces. This is based on the statistical assumption that the number of 1s in a pulse train is equal to the number of 0s, resulting in a reduction of the transmission power with a possible increase of transmission speed. The RZ unipolar format in half binary transmission is shown in Fig. 18.25 (d). The pulse train is similar to that in Fig. 18.25 (a), except that the pulses corresponding to the 0 bits are absent. As a result, the frequency spectrum of the pulse train has fewer high frequency components, which in turn results in less cross-talk. However, a dc component also results, which is difficult to transmit. The coding method is not efficient, since 50% of the band width is wasted because information is contained in only half the bit periods.

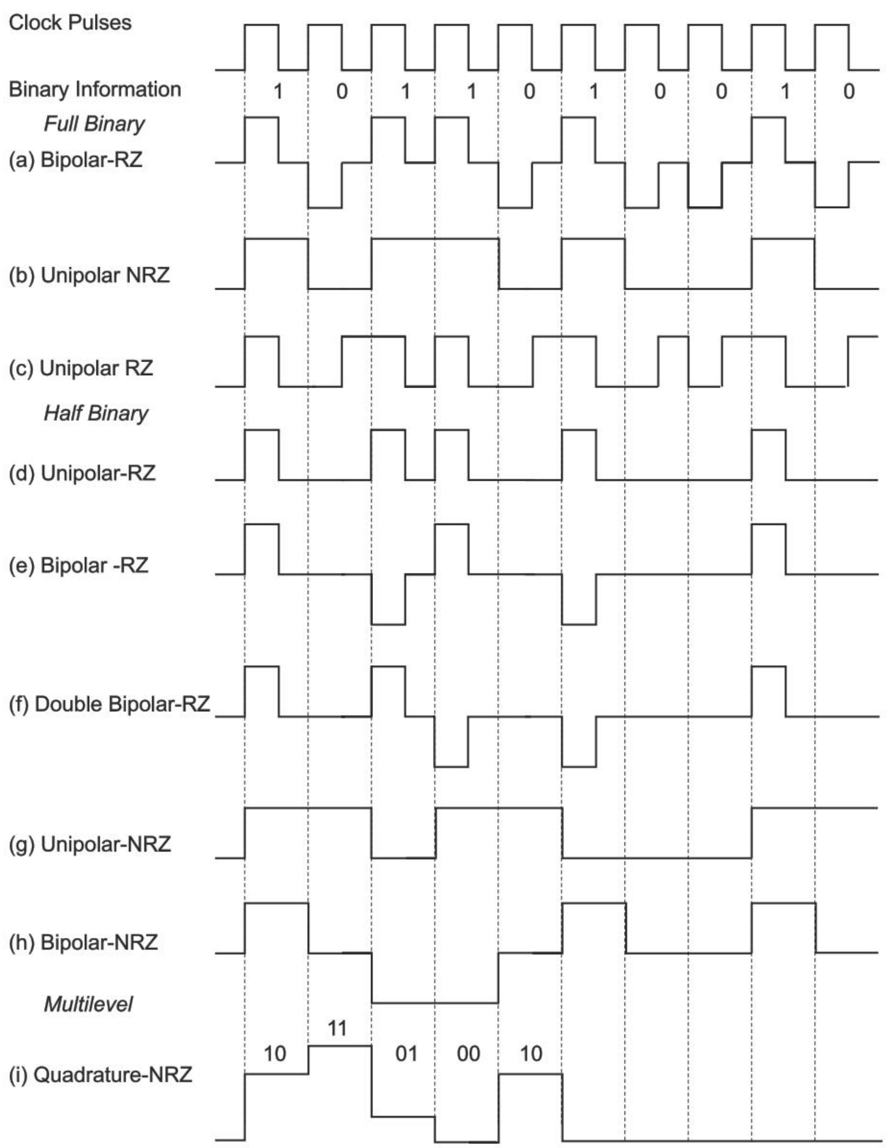


Fig. 18.25 Pulse code format used in data transmission

MODEMS 18.8

The term Modem is an acronym for modulator-demodulator. The primary modem function is to convert digital data into an analog form which is suitable for transmission on common carrier circuits (example telephone lines). Modulation is the D/A conversion in which the digital data is placed on the transmission line by modulation of a tone or carrier. Demodulation is the reverse process.

In a data communication system, transmitting and receiving modems are necessary at each end of the analog transmission line. The interfaces in the modulator and demodulator sections are usually EIA RS 232 C or current loop interfaces providing connections for standard external devices. The output transmitting circuits and receiving circuits are networks required for transmitting and receiving analog information to and from the transmission line.

Three modulation techniques in common use are amplitude, frequency and phase modulation. In a simple amplitude modulation system, the amplitude of the modulated carrier frequency corresponds to the value of the data bits. The spectrum of the modulated waveform includes the carrier frequency plus the upper and lower side bands. The side bands are displaced from the carrier by the frequency of the modulating input. The resulting band width is therefore twice that of the data rate.

In a frequency modulation system, digital signals are connected to one of the two frequencies corresponding to the 0 and 1 values of the data. Modulation of this form is known as Frequency Shift Keying (FSK). FSK is a commonly used technique for low speed transmission (typically 0 to 600 bits/s).

Modems operate with one functioning as an originate unit and the other as an answer unit. The originate modem transmits on a low frequency channel, using 1.270 kHz for a mark and 1.070 kHz for a space. It receives on a high frequency channel using 2.225 and 2.025 kHz respectively for mark and space. The answer modem transmits on the high frequency channel and receives on the lower. These frequencies are industry standards and assure compatibility with most commercially available low speed data channels, including national time sharing services. The carrier frequencies occupy discrete portions of the pass band. However, the frequency modulation generates side bands whose deviation from the carrier is directly dependent on the transmission bit rate. The total required band width therefore depends on the transmission speed.

Phase modulation is widely utilized in high speed systems. In its simplest form, two carriers which have identical frequencies but are 180° out of phase with one another are used.

Each phase is used to represent a mark or space condition. In such a system, both phase angles are referenced to a defined phase angle that is known by the transmitter and receiver. Another method transmits information contained in the relationship of the two successive phase angles, known as differentially coherent phase shift keying (PSK). PSK eliminates the need for a stable phase reference at the receiver.

The transmission timing for the digital data exchange rate can be either asynchronous or synchronous. Asynchronous timing is simpler and less expensive, but has the disadvantage of a lower data exchange efficiency. Start and stop bits are used for each character, which may account for nearly 30% of the transmission time. In operation, the start bit is recognized by the receiver and the succeeding bits are received at a specified rate. The stop bit(s) permits the receiver to reset and prepare for the next start bit. Low speed data systems employing devices such as teletypewriters often operate in this fashion.

The greater complexity and cost of synchronous modems over asynchronous units is due to the circuitry necessary to derive the timing from the incoming data and to pack more than one bit into one baud (the number of signaling elements per unit time).

Synchronous modems typically consist of four sections, as shown in Fig. 18.26(a). The transmitter, receiver, terminal control and power supply. The transmitter section of a synchronous modem typically consists of timing (clock), scrambler, modulator, digital to analog converter and equalizer circuits. The expanded block diagram is shown in Fig. 18.26(b). The timing circuit provides the basic clocking information for both the modem and the data terminal equipment (DTE) that is providing the data to be transmitted. The internal timing is usually controlled by a crystal oscillator to within about 0.05% of the nominal value.

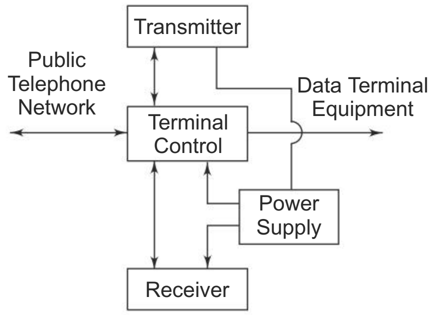


Fig. 18.26 (a) Synchronous modem block diagram

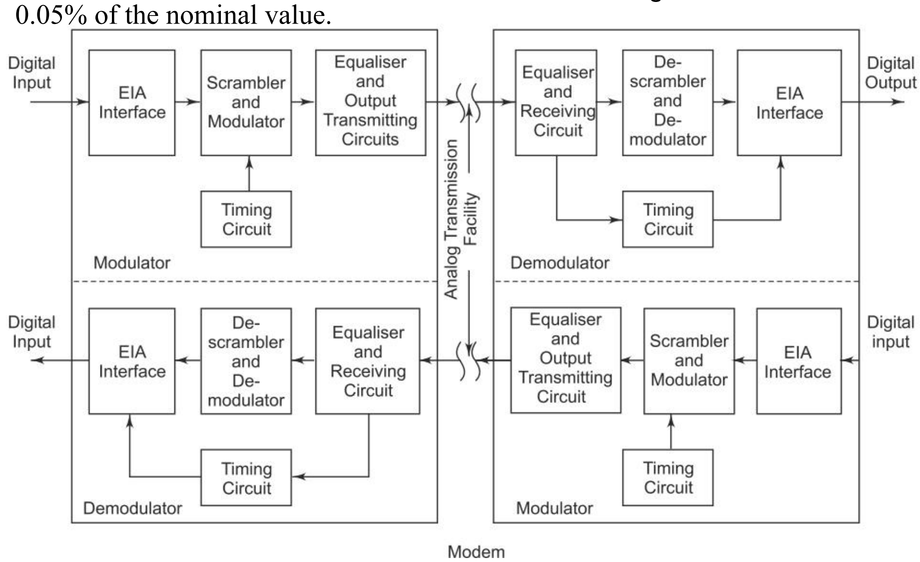


Fig. 18.26 (b) Expanded block diagram of a data communication system employing modems

Since the receiver clock is derived from the received data, those data must contain enough changes from 0 to 1 (and vice versa) to assure that the timing recovery circuit stays in synchronization. In principle, the data stream provided by the associated terminal or business machine can consist of any arbitrary bit pattern. If the pattern contains long strings of the same value, the data will not provide the receiver with enough transitions for synchronization. The transmitter must prevent this condition by changing the input bit stream in a controlled way. The part of the transmitter circuitry that does this is called the scrambler. Scrambler are usually implemented as feedback shift registers which may be cascaded or connected in series. They are designed to ensure that each possible value of phase angle is equally likely to occur, to provide the receiver demodulator with enough phase shifts to recover the clocking signal. Although scrambling is necessary for the reasons cited above, it increases the error rate, since an error in one bit is likely to cause an error in subsequent bits. To counteract this, some modems encode the input to the scrambler in gray code, so that the most likely error in demodulation will cause only a 1 bit error when decoded at the receiver.

The modulator section of the transmitter converts the bit patterns produced by the scrambling process into an analog signal representing the desired phase and amplitude of the carrier signal. The carrier frequency, baud rate, and number of bits represented by each baud is different for modems of different rates. The modulator collects the correct number of bits and translates it into a number giving the amplitude of the electrical signal that is correct for the carrier frequency and phase of the carrier at that instant in time. Modulation differs for modems of different speeds, and those made by, from different manufacturers.

The equalizer section of the transmitter is relatively simple, since it can compensate only for the average of expected errors on the output channel. The receiver equalizer, however, must compensate for the actual errors introduced in the transmission path. This is done by using adaptive equalizers which measure errors observed in the received signals and adjusts some parameter of the circuit (usually the receiver clock frequency) to track slowly varying changes in the condition of the transmission line.

At the receiver, the incoming signal from the line is modulated or frequency translated using an internal clock. The resulting intermediate frequency is processed to produce a clock signal at the rate at which the data is actually being received. This signal is applied as the reference to a phase locked loop oscillator. The output of this oscillator is a stable signal locked to the incoming line frequency in both phase and frequency.

The descrambler section of the receiver performs an operation that is the inverse of the scrambler.

Synchronous timing is used in higher speed data systems. The transmitter and receiver are timed from synchronized clocks. A synchronized preamble code is transmitted first. The code is followed by the information to be transmitted, without start or stop bits. The receiving modem maintains synchronism by sampling the received bits. Elimination of the start/stop bits substantially increases the throughput of the synchronous system.

Modems operate in both half and full duplex modes. Since full duplex operation allows simultaneous transmission and reception, two transmission lines are required.

The public telephone network is the most commonly used transmission system. Dial-up lines having bandwidths of 3 kHz may be used for transmission rates of up to 4800 bits per second, whereas lines used for high speed transmission must be leased.

Review Questions

1. What is data transmission?
2. Describe with a block diagram a basic data-transmission system.
3. What do you mean by protocol?
4. State the elements of a protocol.
5. What do you men by a character set?
6. Explain the basic concept of data transmission.
7. State the important points to be agreed upon by the sender and the receiver.
8. Describe the basic block diagram and draw waveforms at each points of a data transmission system.
9. State the advantages and disadvantages of digital transmission over analog transmission.
10. What do you mean by Time Division Multiplexing (TDM).
11. Explain with a simplified diagram the operation of a TDM system.
12. Explain with diagram how a 4-channel electronic sampling system operates.
13. What is pulse modulation?
14. State the various types of pulse modulation.
15. Draw different waveforms of analog pulse modulation for the same message.
16. Describe the principle of PAM.
17. Explain with diagram the working of a pulse amplitude modulation. How is it achieved?
18. Describe the principle of operation of pulse time modulation.
19. State the sampling theorem of pulse modulation.
20. State with diagram the working principle of a Pulse Width Modulation (PWM).
21. How is generation and demodulation of PWM obtained?
22. Explain with a diagram the operation of a monostable multivibrator generating PWM.
23. Explain with a diagram the operation of a 555 Timer generating PWM.
24. Explain with a diagram a method of transmitting PDM's.
25. Explain with diagram the working of a PDM recovery circuit. Draw the necessary waveforms.
26. Describe with a diagram and waveform, how PDM is converted to PAM.
27. Describe with a waveform, how PPM can be derive from PDM.
28. State the principle of Pulse Position Modulation (PPM).
29. Explain with a diagram and waveforms, how a PPM is generated.
30. What is PPM and how is it different from PWM?
31. Compare analog modulation and digital modulation.
32. How is digital modulation achieved?
33. What is a PCM?
34. State the principle of PCM. How is it generated?
35. What do you mean by companding and PCM encoding?

36. Explain with a block diagram the operation of a PCM system.
37. Describe with a diagram the channeling scheme of a PXM system.
38. State the applications of PCM.
39. Describe with block diagram the operation of a PCM speech terminal.
40. State the advantages of PCM.
41. State the principle of delta modulation.
42. What is delta modulation?
43. Explain with a diagram the operation of a delta modulation system.
44. Explain in brief with a diagram the working of a delta modulation encoder and decoder.
45. Explain with waveforms the operation of a delta modulation.
46. What do you mean by slope overload and how is it overcome?
47. What do you mean by companded delta modulation.
48. Explain with a diagram the working of a variable slope delta modulation encoder and decoder.
49. What is a pulse code format?
50. Explain in brief the different types of pulse code format used in data transmission.
51. Explain with waveforms the operation of a full binary transmission.
52. Explain with waveforms the operation of a half binary transmission.
53. Compare full binary transmission and half binary transmission.
54. What are modems? Where are they used?
55. Explain with diagram the working of a basic synchronous modem.
56. Explain the expanded block diagram of a data communication system using Modems.

Multiple Choice Questions

1. The basic elements of a communication protocol are a set of symbols called
   (a) hardware set
   (b) character set
   (c) software set
   (d) none of the above
2. TDM allows each channel
   (a) full BW transmission
   (b) half BW transmission
   (c) a limited portion of transmission.
   (d) none of the above.
3. FDM allows each channel
   (a) full BW transmission
   (b) half BW transmission
   (c) a limited portion of transmission.
   (d) none of the above
4. In PAM, the following is modulated:
   (a) amplitude (b) duration
   (c) position (d) code.
5. In PDM, the following is modulated:
   (a) amplitude (b) duration
   (c) position (d) code
6. In PPM, the following is modulated:
   (a) amplitude (b) duration
   (c) position (d) code
7. In demodulating PDM, the following filter is used:
   (a) LPF (b) HPF
   (c) BPF (d) BSF
8. Companding is the process of
   (a) volume compression
   (b) frequency compression
   (c) current compression
   (d) data compression.
9. Delta modulation is the process of modulation in which a train of
   (a) fixed width pulses are transmitted
   (b) variable width pulses are transmitted
   (c) fixed amplitude pulses are transmitted
   (d) none of the above.

Further Reading

1. Radio Shack, *Understanding Data Communication*, Howard W. Sons, 1988.
2. D. Roody and J. Collen, *Electronic Communications, 3rd Edition*, Prentice-Hall of India, 1984.
3. William Sinnema, *Digital Analog and Data Communications*, Reston Publishing Co.
4. Rangan, Sharma and Mani, *Instrumentation Devices and Systems*, Tata McGraw-Hill, 1985.
5. George Kennedy, *Electronic Communication Systems*, McGraw-Hill, 1987.
6. John L. Hilburn and Paul M. Julich, *Microcomputer/Microprocessor, Hardware, Software and Applications*, Prentice-Hall of India, 1976.

Chapter

Frequency Standards 19

INTRODUCTION 19.1

The fundamental standard of frequencies is the period of rotation of the earth. It can be measured with great accuracy by astronomical methods and is taken as a standard frequency of 1 cycle/day. The error in the measurement is approximately 0.000164 s (164 μs). All standards of frequency must ultimately be referred to this fundamental source for calibration purposes.

Practical standards can be classified as primary or secondary standards.

A primary standard of frequency is an oscillator which generates a frequency that is constant over a long period and is checked directly with the earth's rotation.

Secondary standards of frequency are very stable oscillators which have their frequencies checked periodically against the primary standards.

Recent developments in microwave spectroscopy have introduced the possibility of using spectrum lines as frequency standards. For example, ammonia gas has an absorption line at a frequency of 23,870.1 MHz. This line is taken as a reference line for controlling the frequency of an oscillator by means of an automatic frequency control system that operates to keep the frequency of the oscillator close to the frequency corresponding to the spectrum line.

PRIMARY STANDARDS 19.2

One arrangement is a carefully designed quartz crystal oscillator operating in the range 50 to 100 kHz, having a low temperature coefficient, constant amplitude output and voltage regulated power supply. The crystal oscillator has a long time (several months) frequency stability of a few parts in a 100 million, without readjustment.

A typical arrangement for comparing the frequency of a primary standard with the period of rotation of earth is given in Fig. 19.1. Here the frequency of the crystal oscillator is 100 kHz. Frequency dividers are used to reduce the frequency until an output frequency of 1 kHz is obtained, to drive the electronic clock. The clock keeps correct time when supplied with exactly 1 kHz frequency.

Time indicated by the clock is checked periodically with the observatory time as broadcast on radio channels, i.e. the radio time signal.

The primary standard must be provided with very accurate means for making comparisons, since 0.01s in one day represents more than 1 part in 10,000,000.

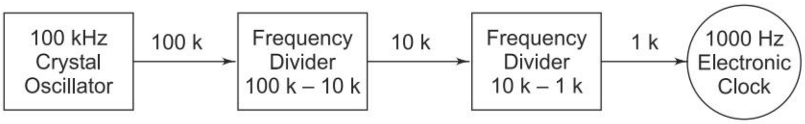


Fig. 19.1 Arrangement for comparing primary standards

SECONDARY STANDARDS OF FREQUENCY 19.3

Secondary frequency standards are ordinarily based on carefully designed crystal oscillators. When the utmost in frequency precision is not required of the secondary standard, it is possible to relax the design in certain respects, such as less precise temperature control to the crystal, use of crystals at frequencies lying outside the optimum range of 50 – 100 kHz, etc.

The crystal oscillator frequency used for a secondary standard frequency can be maintained constant for long periods without readjustment, to within a few parts in a million.

PRACTICAL FREQUENCY STANDARDS 19.4

Frequencies are categorised as either primary or secondary standards.

A primary standard is one whose frequency output is compared directly with the fundamental standard of frequency which is the rate of earth's rotation.

A secondary standard is one whose frequency output is calibrated against a primary standard.

The unit of time universally accepted is the mean solar second, which is defined as 1/86400 of the average time required for one complete rotation of the earth around the sun. However, this rotation is not constant. There is a continuous reduction of angular velocity amounting to an increase in length of the day of 0.00164 s per century. Secondly, there are cyclic seasonal variations in the length of the day.

The primary standards of frequency consist of eight precision frequency standards.

The 100 kHz outputs of these standards are automatically compared with each other and the time determined by the US Naval Observatory. The National standards of frequency are made available by continuous broadcast from radio stations on a total of eight radio carrier frequencies of 2.5, 5, 10, 20, 25, 30 and 35 MHz.

Two standard audio frequencies, 440 Hz and 4000 Hz, are broadcast on these radio carrier frequencies. The accuracy of each of the radio carrier frequencies as transmitted, is better than one part in 50 million.

In each case, however, if the accuracies of the received carrier frequencies of this order are required, it is necessary to make repeated measurements over a long interval in order to take into account the fluctuations in the transmitting time of the received signal. These fluctuations are a direct result of any changes in the propagation medium, i.e. changes in ionospheric density and level cause the frequency of the received signal to exhibit small variations about the transmitted signal frequency.

These variations, known as the doppler effect, are related to geographic location, season, sunspot cycle, and other factors, and necessitate time averaging measurements over a long interval, or application of some corrective technique.

RADIO SIGNALS AS FREQUENCY STANDARDS 19.5

The US National Bureau of Standards conducts regular transmission that includes continuous transmissions on several carrier frequencies. The carrier frequencies, modulated frequencies and the time interval associated with these signals are derived from a primary standard that has an accuracy that is better than 1 part in 50 million.

A shift in the height of the ionosphere leads to the Doppler effect, which causes the received frequency to differ from the transmitted frequency, causing an error. However, the maximum error introduced is very slight, if the received frequency is averaged over an extended time interval, the doppler effect will cancel out.

Signals from broadcast stations are particularly good secondary standards, because they are required to maintain their assigned frequency to within 20 Hz, and commonly maintain it to within a few parts/million.

PRECISION FREQUENCY STANDARDS 19.6

Most high precision frequency standards in use utilise a geometric transverse (GT) cut quartz crystal as the frequency controlling resonant elements having changes in frequency of less than 1 part in 10 million per day. This GT quartz has the outstanding characteristic of exhibiting little variation in frequency over a temperature range of 0 – 100°C.

The bridge stabilised Meacham oscillator given in Fig. 19.2 is a constant frequency oscillator of very high stability.

It consists of a Wheatstone's bridge and an amplifier section. One of the four bridge arms is a thermally controlled resistance element, two of the arms are fixed resistances, and the fourth is a frequency controlling sensitive element, i.e. crystal.

Because of the presence of a thermally sensitive element having a large positive temperature coefficient of resistance (a small tungsten filament lamp of low wattage rating), the output of the bridge is amplitude stabilised and relatively insensitive to small changes in circuit parameters and fluctuations in power supply.

To compare the frequency of the primary standard to that of the earth's rotation, the output of the primary standard may be reduced in frequency by a succession of frequency dividers, to the point where it can be used to operate an electric clock. This arrangement is known as a quartz crystal clock which can be compared frequently with time signals from the Naval Observatory.

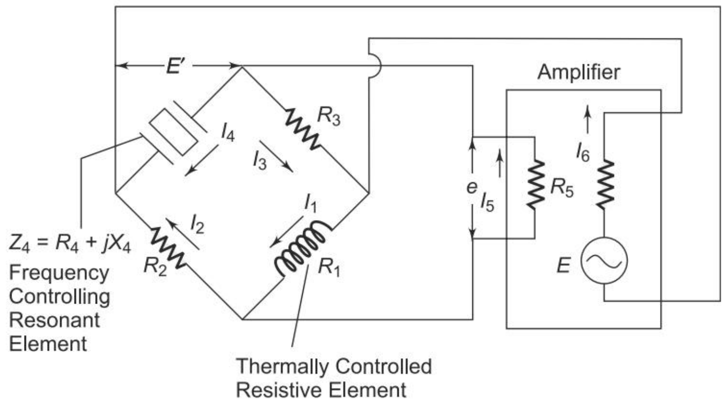


Fig. 19.2 Bridge stabilised meacham oscillator

THE ATOMIC CLOCK 19.7

The primary time and frequency standards are based on astronomical determination of the earth's rotation. A secondary primary standard, an atomic clock developed at the National Bureau of Standards, is based on control of a quartz crystal oscillator by the constant natural frequency associated with the vibration of the atoms in the ammonia molecule. The molecular spectral line of ammonia gas is taken as a stabilizing control.

An auxiliary oscillator, known as search oscillator, is swept through a small frequency range of 6 MHz centered on the frequency of gas resonance. The strongest absorption line observed for ammonia gas is at a frequency of 23,870.1 MHz. A waveguide cell containing ammonia gas at low pressure exhibits a sharp resonance at this frequency. It is coupled to the gas cell and there is a decrease in the transmitted power when the frequency passes through the gas resonance.

The detected signal therefore consists of a series of negative pulses. The search oscillator is mixed with the stabilised oscillator, and when the difference in their frequency is the same as that of the tuned amplifier, a pulse of energy is transmitted through the amplifier. A second series of pulses is thus obtained. The two sets of pulses are mixed in a discriminator, which produces an error voltage proportional to the phase difference of the pulses. The controlled oscillator is thus held at a frequency which differs from that of the gas resonance by the frequency of the tuned amplifier.

Review Questions

1. What is a frequency standard?
2. What do you mean by a fundamental standard?
3. What are the standards for measurement of frequency?
4. Describe a primary standard.
5. Describe a secondary standard.
6. State the difference between a primary standard and secondary standard.
7. What are practical frequency standards?
8. How can radio signals be used as frequency standards. Explain.
9. What are precision frequency standards?
10. Explain with a diagram the working of a bridge stabilised Meacham oscillator.
11. What is atomic time?
12. What are the disadvantages of transmitting time and frequency standards by HF (3–30 MHz) radio signals.
13. What are some of the methods used to improve the dissemination of time and frequency standards.

Multiple Choice Questions

1. The crystal oscillator are used to generate frequency of
   (a) 1 kHz (b) 10 kHz
   (c) 100 kHz (d) 1000 kHz
2. The crystal oscillator has a frequency stability of a few parts in
   (a) 1 million (b) 10 million
   (c) 100 million (d) 1000 million
3. The two standard audio frequencies used are
   (a) 400 Hz and 1000 Hz
   (b) 50 Hz and 20 kHz
   (c) 440 Hz and 4000 Hz
   (d) 1000 Hz and 4000 Hz
4. The accuracy of primary standard used is better than 1 part in
   (a) 1 million (b) 10 million
   (c) 100 million (d) 50 million
5. The quartz used in precision frequency standard has a stability over a temperature range of
   (a) 0–200°C (b) 100–200°C
   (c) 0–100°C (d) 0–50°C

Further Reading

1. Terman and Petit, *Electronic Measurements*, McGraw-Hill, 1952.
2. *Handbook of Electronic Measurements, Volumes I and II*, Polytechnic Institute of Brooklyn, 1956 (Microwave Research Institute).

Measurement of Power

Chapter 20

INTRODUCTION 20.1

At AF and RF it is easier to measure voltage, current and impedance, than to measure power. Direct measurement of power is not done in this range. Power is calculated from the equation $P = E^2/R = I^2 R$.

On the other hand, at microwave frequencies, voltage, current and impedance are difficult to determine. Their values may vary at different points in the circuit, and they are affected appreciably by small changes in geometry. Therefore, in this range, direct measurement of power is more accurate and the actual load is replaced by a dummy load.

REQUIREMENTS OF A DUMMY LOAD 20.2

The following are the requirements of a dummy load.

1. Ability to dissipate the required amount of power
2. Low reactances
3. Low skin effect
4. A suitable value equal to that of the actual load

Dummy loads may take any of the following forms.

1. Non-inductive wire wound resistor, cooled by air or water, for frequencies up to 25 MHz and power up to 50 W.
2. Water resistance, for large powers up to 50 W.
3. $\lambda/4$ Marconi grounded antenna suitably matched to the line, such an antenna presents a load resistance of 36.7 Ω.
4. Bolometer element for very small powers.

BOLOMETER 20.3

Bolometric measurements are based on the dissipation of the RF power in a small temperature sensitive resistive element, called a Bolometer. This bolometer may be a short ultra thin wire having a positive temperature coefficient (PTC) of resistance, called Baretter, or a bead of semi-conductor having a negative temperature coefficient (NTC) called Thermistor.

Both Baretters and Thermistors can measure small powers, of the order of a fraction of micro-watts. They can also indicate or monitor large amounts of power by inserting a directional coupler.

The RF power to be measured heats the bolometer and causes a change in its electrical resistance, which serves as an indication of the magnitude of power.

The bolometer is generally used in a bridge network so that small changes in resistance can be easily detected and hence power can be measured.

BOLOMETER METHOD OF POWER MEASUREMENT 20.4

Bolometric measurements are based on the dissipation of RF power in a small, temperature sensitive resistive element, known as a bolometer. This may be a short ultra thin wire, a bead of semiconductor material or a thin conducting film of small dimensions.

The bolometer is generally incorporated into a bridge network, so that small changes in resistance can be readily detected and the power measurement performed by substitution of low frequency power.

In one procedure, the bridge is initially balanced with low frequency (or bias) power. (Power for the purpose of controlling or biasing bolometer resistance and bridge operating conditions is termed bias power.)

When RF power is applied to bolometer, low frequency power is withdrawn until the bridge is balanced again.

In any case, whatever the bridge technique used, the underlying assumption of the measurement is that equal amount of low frequency and RF power produce the same heating effect and resistance changes in the bolometer.

Bolometric measurement are well suited to the measurement of low and medium power, ranging from a few μ watts to a fraction of a μ watt.

BOLOMETER ELEMENT 20.5

In order that a bolometer element be well matched to the RF line, be equally responsive to low frequency and RF power, and be highly sensitive, it must be small in size.

In most bolometer elements, cross-sectional diameters are usually of the order of the skin depth of penetration of RF current at the highest frequency of operation. The dc and RF resistivities are essentially the same and the reactive component of the bolometer impedance is held to a minimum. The maximum cross-sectional area that can be tolerated is inversely proportional to the conductivity of the bolometer material and the highest frequency of operation. This explains why ultra thin metallic wires are required at micro wave frequency.

Metal wire bolometers are referred to as barreters. They have a positive temperature coefficient of resistance and are generally operated at bias powers which heat to 100 – 200°C. (They consist of short lengths of Wollaston wire whose external silver sheath has been removed by etching or deplating, so as to expose the extremely thin noble metal core, usually consisting of platinum or platinum alloy. Wire diameters range from approximately 1 – 3 microns and wire lengths range from as short as feasible to about 2.5 mm. The resistance of the deplated portion is selected so that the bolometer presents a good match to the RF line when properly biased with low power.)

Since the dc and RF resistances are nearly the same, the resistance value is generally close to the characteristic impedance.

Thermistor elements for RF measurements are usually minute beads of a ceramic like semiconductor mixture of metallic oxides having a large negative temperature coefficient of resistivity.

Two fine platinum or platinum alloy wires are embedded in the bead, which is then sintered and coated with a film of glass. The beads are sometimes enclosed in a glass envelope.

BOLOMETER MOUNT 20.6

The bolometer support consists of a thin mica disc on one side of which silver electrodes are sprayed. Silver painted holes through the mica connect the (lower) outer electrode to a circular electrode on the opposite side. Between the centre and outer electrodes are mounted two short (1 mm) lengths of deplated Wollaston wires of 1 micron diameter, each measuring 100Ω at dc with normal bias power. The mica disc is clamped in the holder, as shown in Fig. 20.1, so that the upper electrode makes contact with the metal case, while the other two electrodes are insulated from a co-axial line by a thin mica sheet which provides the bypass capacitor necessary to complete the RF circuit, and to place two wires in RF parallel across the line.

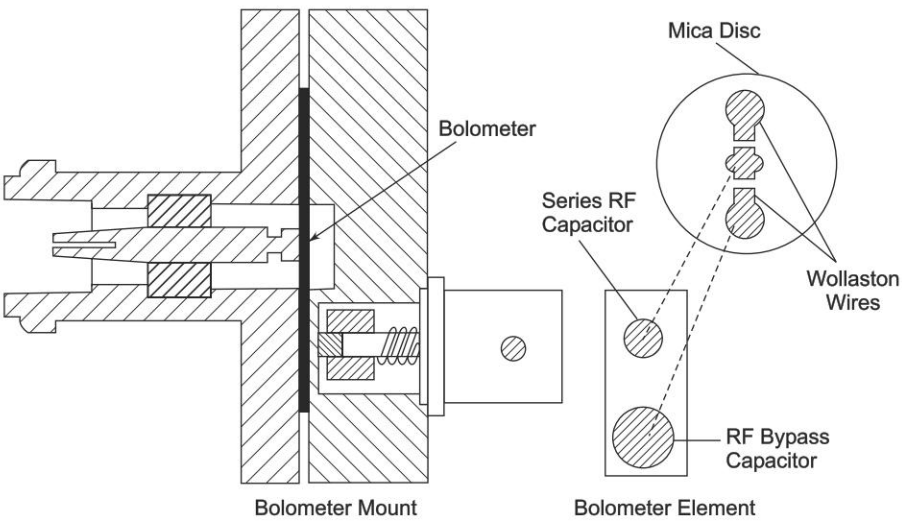


Fig. 20.1 Schematic view of bolometer mount and bolometer element

MEASUREMENT OF POWER BY MEANS OF A BOLOMETER BRIDGE 20.7

To measure unknown RF power, a small known AF power, indicated by the voltmeter V_1, is superimposed on the RF test power. The direct current from voltage E is next adjusted by varying R, which heats the bolometer element until its resistance equals R_1, which is the value required to balance the bridge. The test RF power is then turned off, which unbalances the bridge. Balance is now restored by increasing the AF voltage to V_2. Hence RF power equals.

$$\text{RF power} = \frac{V_2^2 - V_1^2}{4R_1}$$

Since power delivered to the bolometer element is 1/4th power given to the bridge.

In a co-axial and wave guide transmission system, the bolometer mount must provide the necessary impedance transformation (matching). In a co-axial cable, a 50Ω transmission line impedance has to be transformed to a 125 – 200Ω resistance value of a bolometer. This is usually done by means of a tapered section, as shown in Fig. 20.2.

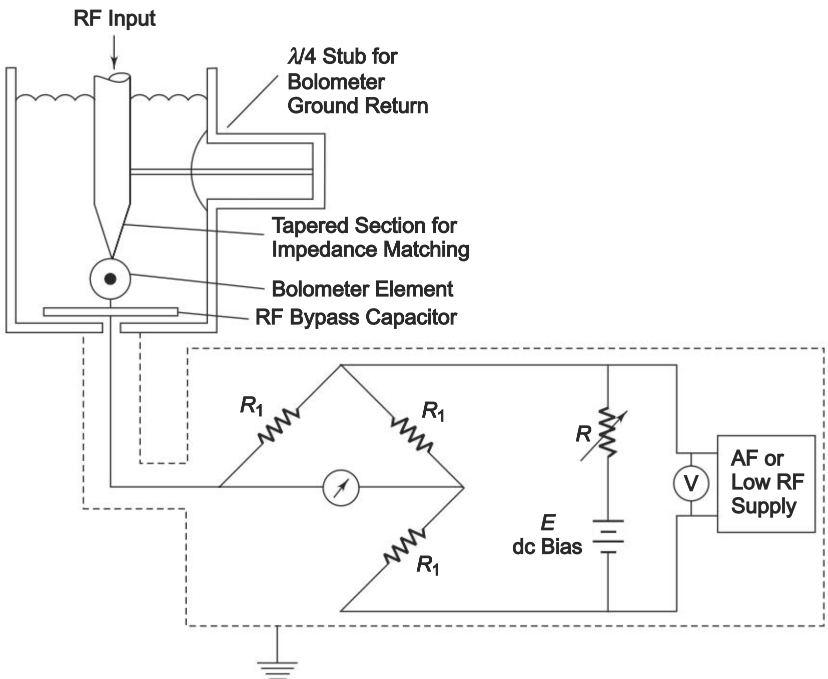


Fig. 20.2 Bolometer bridge

Example 20.1 *A small AF voltage of 20 V is superimposed on the RF test power and balance is achieved. If the RF test power is now turned off, 30 V AF is required to balance the bridge. If the bridge arms has a resistance of 100 Ω. Calculate the RF test power.*

Solution Given $V_1 = 20$ V, $V_2 = 30$V, $R_1 = 100\ \Omega$

RF test power is given by

$$= \frac{V_2^2 - V_1^2}{4R_1} = \frac{(30)^2 - (20)^2}{4 \times 100}$$

$$= \frac{900 - 400}{400} = \frac{500}{400}$$

$\therefore$ RF test power = 1.25 W

UNBALANCED BOLOMETER BRIDGE 20.8

An unbalanced bolometer bridge is shown in Fig. 20.3. In the absence of RF power, the bridge is brought into balance by adjusting the exciting source voltage. The detector now indicates the balance condition of zero. The test RF is now applied to the bolometer element. The resistance of the bolometer changes, and the bridge is unbalanced by an amount indicated by the detector, which gives the magnitude of RF power directly.

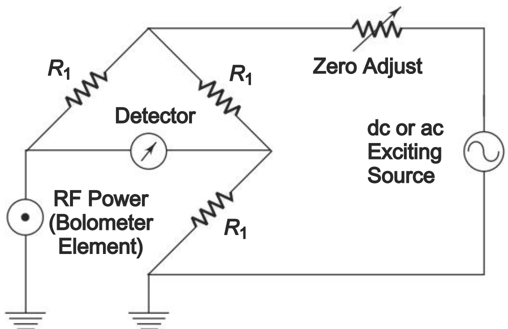


Fig. 20.3 Unbalanced bolometer bridge

The unbalanced bridge method is probably the simplest means of realising a direct reading bridge and is particularly suited for low power measurement.

The disadvantage is that the impedance of the bolometer changes (because the resistance of the bolometer is a function of RF power level), thereby upsetting the impedance match to the RF system. The advantage is that the bridge gives a direct reading.

One method of calibration is to note the detector deflections corresponding to a series of different RF power inputs, each of which is separately determined by rebalancing the bridge and measuring the equivalent bias power (that must be retracted).

Another method, which is not accurate, is to calibrate only the largest deflection that is used and rely on the linear performance for intermediate positions.

SELF BALANCING BOLOMETER BRIDGE 20.9

The term self balancing is used to describe bridges which are automatically rebalanced when unknown RF power is applied to the bolometer. A typical circuit is illustrated in Fig. 20.4.

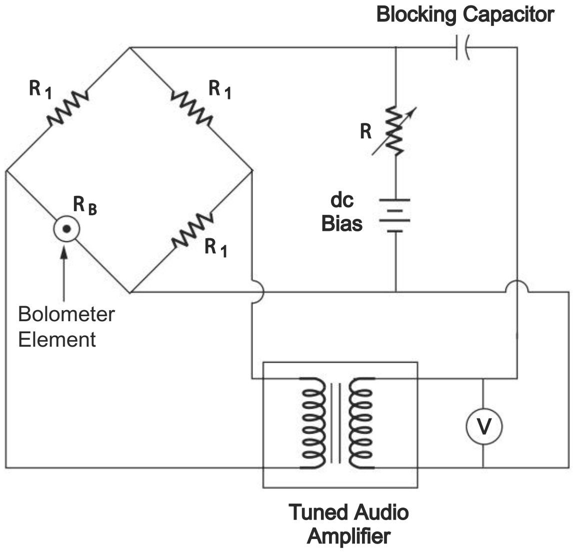


Fig. 20.4 Self balancing bolometer bridge

The bolometer bridge is used as the coupling network between the output and input of a high gain frequency selective audio amplifier. The feedback is in proper phase to produce sustained AF oscillations of such amplitude as well maintain the resistance of the bolometer at the fixed value which nearly balances the bridge.

When the supply is switched ON, the bridge is unbalanced. The gain of the amplifier is large, so that oscillations are allowed to build up until the bridge is almost balanced. The higher the gain of the amplifier, the more closely the bridge balances.

The test RF is now dissipated into the bolometer element, which causes an imbalance in the bridge circuit. The AF output voltage automatically adjusts itself to restore the bolometer resistance to its original value. The amount by which the AF power level in the bolometer is reduced equals the applied RF power. The voltmeter reads the AF voltage and can be calibrated to read the magnitude of RF power directly.

A typical bridge circuit offers seven power ranges, from 0.1 – 100 mW full scale, for use with bolometers having a resistance within + 10% of five selected values from 50 – 250 Ω.

MEASUREMENT OF LARGE AMOUNT OF RF POWER (CALORIMETRIC METHOD) 20.10

Principle RF power may be directly converted into heat. Water acts as the load, and RF power may be used to heat an element such as a long transmission line, water being used as a coolant. The RF power may be absorbed directly in the calorimeter fluid.

The system should have no RF leakage either by radiation or through lossy joints.

Important Features

1. All the RF power should be converted into heat and none allowed to leak through radiation.
2. No heat should be allowed to radiate from the fluid

The power radiated is given by the relation.

$$P = 4.18 \times \text{Mass} \times \text{Sp. Heat} \times \Delta t$$

where m = mass, flow gm/s
sp = specific heat, calories per gm °C
Δt = temperature rise
P = power, in W.

Alternatively, the relationship between temperature rise and power dissipated may be established by dissipating a known amount of ac (50 Hz) or dc power in the calorimeter system.

Figure 20.5 is a diagram for measurement of power in a lossy cable using the calorimetric method. In this figure, thermometers are used for measuring the difference in temperature.

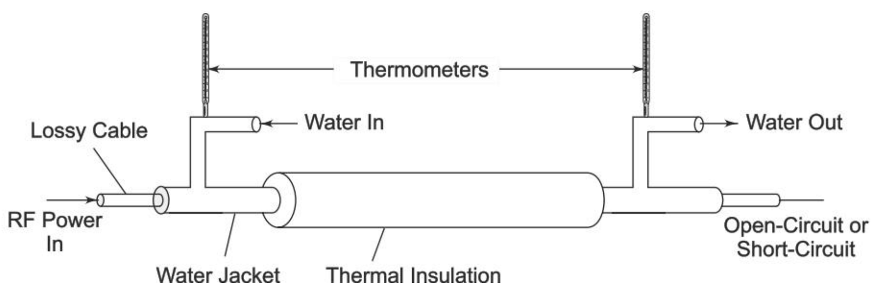


Fig. 20.5 Calorimeter using thermometer

As in the bolometer method of power measurement, the RF portion of the calorimeter system must be designed to provide a proper transformation of impedance. This is usually done by a tapered section.

The length of the cable should be sufficient to provide an attenuation of 10 db or more at the operating frequency, so that the reflected wave will be down by 20 dbs, corresponding to 100 W of power. Thus, the load virtually absorbs almost all the RF power. The cable is available with a standard impedance of 50 Ω, and hence there is no problem of impedance matching. The water acts as the load and directly absorbs the power from the source.

Another arrangement used is the co-axial system, illustrated in Fig. 20.6.

Here, the RF power is absorbed in a water filled section of line. The water serves as a lossy dielectric that absorbs the power because of its high power factor and at the same time functions as the calorimeter fluid.

An impedance match between the section of the co-axial line having an air dielectric and the portion having a water dielectric is provided by the tapered section of titanium dioxide. The material has a dielectric constant very close to that of water (approximately 80) and is ideally suited to match air to water.

The temperature difference between the incoming and outgoing water is measured with a thermopile instead of thermometers, as in Fig. 20.6.

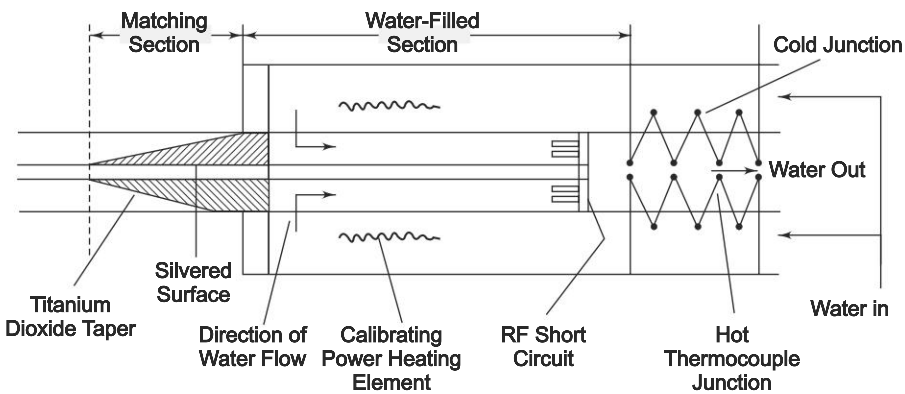


Fig. 20.6 Co-axial water load with titanium dioxide taper calorimeter

The thermopiles consist of a succession of thermocouple junctions connected in series, as shown in Fig. 20.6, with alternate junctions placed in the incoming (cold) and outgoing (hot) fluid.

A thermopile mounted in this way responds to the difference in temperature between the hot and cold fluids and has a greater sensitivity than does a single hot and cold junction.

Calibration is carried out using dc or 50 Hz power. This consists of an auxiliary heater directly immersed in the flow system. Calibration is carried out by observing the thermopile response produced by the dissipation of a known amount of 50 Hz or dc power in the heater.

Example 20.2 *Calculate the power radiated by a transmission line having the following parameters.*

$$\text{Mass} = m = 200 \text{ gm}$$
$$\text{Sp. heat of water} = sp = 1 \text{ cal/gm°C}$$
$$\text{Initial temperature} = T_1 = 30°C$$
$$\text{Final Temperature} = T_2 = 40°C$$

Solution The power radiated is given by

$$P = 4.18 \times \text{Mass} \times \text{Sp. heat} \times \Delta t$$

$$P = 4.18 \times \text{Mass} \times \text{Sp. heat} \times (T_2 - T_1)$$

$$\therefore \quad P = 4.18 \times 200 \times 1 \times (40 - 30) = 8.3 \text{ kW}$$

MEASUREMENT OF POWER ON A TRANSMISSION LINE 20.11

When power is applied to transmission line, the source to load impedance must be matched in its characteristics impedance (Z_o). The impedance all along the line is equal to the characteristic impedance. Therefore, the transmission line is said to be correctly terminated. Hence no reflection or standing wave is produced, but only an incident or travelling wave is obtained.

The voltage distribution along the line is uniform (constant). With a loss-free line, the characteristics impedance is purely resistive and can be indicated by R_o. The power obtained in such a non-resonant transmission line equals to I^2R_o or E^2/R_o, where I is the current flowing along the line and E is the voltage across the characteristic impedance R_o.

Power measurements done in this way, will be accurate, provided the load impedance of the transmission line closely approximates the characteristics impedance. The instrument used to measure voltages or current should not produce an appreciable reflected wave.

Thermocouple connected in series with the transmission line, can be used, to measure the current flowing in a transmission line, up to the highest frequency in the RF range. The accuracy obtainable is quite high provided the heater wire is small enough to have negligible skin effect.

The voltage across a transmission line may be determined satisfactorily by means of a diode VTVM, provided the frequency is low. The input capacitance of the voltmeter (or the voltmeter multiplier) that is shunted across the transmission line at this low frequency has a reactance that is much higher than the characteristic impedance Z_o. The requirement is not satisfied when the frequency is high. The voltmeter then introduces a large reflected wave at this high frequency.

An arrangement employing a diode voltmeter is as shown in Fig. 20.7. When the voltmeter is connected to the line, an additional capacitance is added by the diode and L/C ratio is changed, the distributed L of the meter section is increased by making the diameter larger.

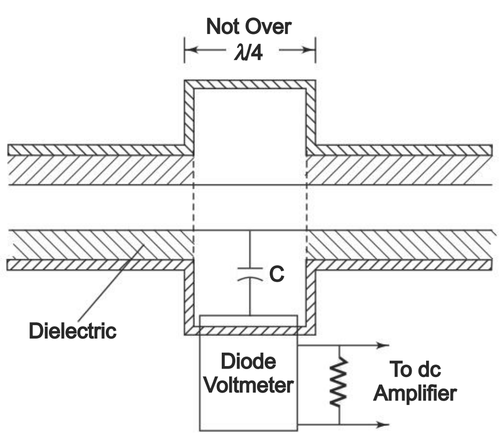


Fig. 20.7 Arrangement using a diode voltmeter

Hence, the characteristic impedance of the meter section equals R_o and the power flow is undisturbed by the introduction of the meter.

STANDING WAVE RATIO MEASUREMENTS 20.12

Standing waves are created along the length of a transmission line due to the mismatch between the characteristic impedance Z_o and the terminating impedance of the transmission line.

The actual voltage E on the line at any point is the sum $(E_i + E_r)$ of the voltage of the incident and reflected waves at the point. This results in a voltage distribution on the line, such as illustrated in Fig. 20.8, called a standing wave pattern.

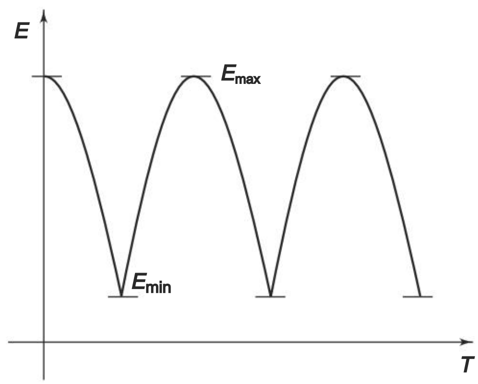


Fig. 20.8 Standing wave pattern

If on a certain line there is both an incident and a reflected wave, a voltage maxima occurs at points where the two waves are in the same phase, and a minima at points where the two waves are in phase opposition.

When the difference between maxima and minima is more pronounced, the larger the reflection coefficient. In particular, when the reflection coefficient of the load is unity, the minima are very deep, while when the reflection coefficient of the load is zero, there is no standing wave pattern.

The distance between adjacent minima (or maxima) is exactly half the line wavelength.

The ratio of the maximum to the minimum value of voltages (or currents) in the standing wave pattern is termed the standing wave ratio.

Voltage standing wave ratio

$$S = \frac{E_{max}}{E_{min}} = \frac{|E_i| + |E_r|}{|E_i| - |E_r|}$$

where $E_{max} = |E_i| + |E_r|$
$E_{min} = |E_i| - |E_r|$

If the line has no attenuation, the standing wave ratio, S, is the same everywhere. (If the line has losses, S decreases with increasing distance from the load.)

If directional couplers are used for measurement of the standing wave ratio, then V_2 indicates $|E_i|$, V_1 indicates $|E_r|$ and the standing wave ratio is

$$S = \frac{(V_1 + V_2)}{(V_1 - V_2)}$$

Example 20.3 *If the maximum value of voltage is 8 V and the minimum value of voltage is 2 V in a standing wave pattern, determine the standing wave ratio.*

Solution

$$\text{Standing wave ratio} = \frac{V_1 + V_2}{V_1 - V_2}$$

$$V_1 = 8\text{V} - \text{Maxima}$$

$$V_2 = 2\text{V} - \text{Minima}$$

$$\therefore \quad \text{Standing wave ratio} = \frac{8+2}{8-2} = \frac{10}{6} = \frac{5}{3} = 1.66$$

Hence the standing wave ratio is 1.66.

MEASUREMENT OF STANDING WAVE RATIO USING DIRECTIONAL COUPLERS 20.13

A directional coupler is a device, which when coupled to a transmission line or waveguide, responds only to the wave travelling in a particular direction on the primary transmission system, while being unaffected by a wave travelling in the opposite direction on the primary line.

Directional couplers serve as stable, accurate and relatively broad band coupling devices, which can be inserted into a transmission line to measure either incident or reflected power.

Referring to Fig. 20.9(a), the arrangement is so designed that a wave travelling towards the right in the co-axial line will induce a wave travelling towards the left in part A of the coupled co-axial cable, while a wave travelling towards the left in the co-axial line will produce no effect in part A of the secondary line.

The usefulness of a directional coupler in power measurement arises from the fact that in RF transmission systems each wave on the system can be considered as transmitting power in the direction in which the wave travels.

The net power transmitted towards the load is the difference between the power of the incident wave and the power associated with the reflected wave, i.e. the load power is the difference between the incident and reflected power at the load point.

The power represented by the induced wave in the directional coupler is a definite function of the power associated with the corresponding wave on the primary system. The ratio of the induced power to total power is termed as coupling, and is commonly expressed in decibels.

A single directional coupler determines the power flowing in only one direction on the primary line. By using two directional couplers on the same primary system, with one directional coupler responsive only to the wave travelling towards the load and the other responsive only to the reflected wave, one can separately measure the powers of the two primary waves, and by subtraction obtain the net load power.

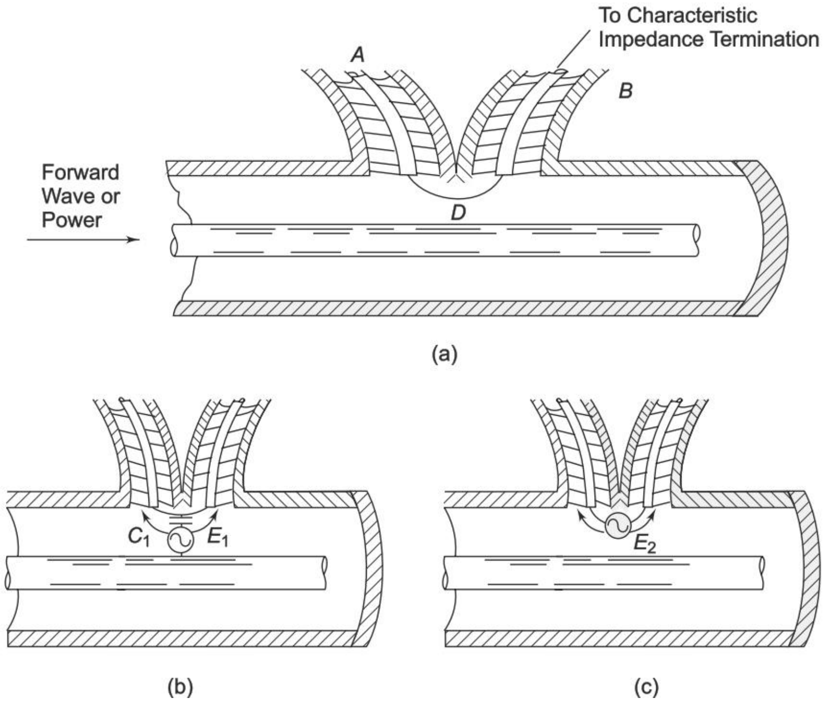


Fig. 20.9 (a) Directional coupler (b) Electric coupling (c) Magnetic coupling

In Fig. 20.9(a) the primary system is a co-axial line, and the secondary system consists of two co-axial lines *A* and *B*, interconnected by a loop *D* that projects into the primary line and is subjected to the simultaneous influence of the electric and magnetic fields that exist in it.

Consider now the case of a wave travelling towards the right on the primary system. The electric field of this wave induces a charge on the loop *D* that produces waves in both parts *A* and *B* of the secondary systems.

The equivalent circuit that describes this action is illustrated in Fig. 20.9(b) and consists of a voltage E_1 that is applied to co-axial systems *A* and *B* in parallel through series capacitance C_1, producing currents as indicated by the arrows.

The loop *D* links with the magnetic flux from the wave in the primary line, and therefore has a voltage E_2 induced in series with it, as illustrated in Fig. 20.9(c).

This series voltage gives rise to waves in parts *A* and *B* of the secondary system, which are characterised by currents flowing in the direction indicated by the arrows.

The two waves in section *A*, produced by magnetic and electrostatic coupling respectively, are of the same polarity and hence add, while the two waves produced in section *B* are of opposite polarity, and so tend to cancel each other.

If the electric and magnetic couplings are so proportioned that the waves induced by the magnetic effect have the same amplitude as the waves induced by electric coupling, complete cancellation takes place in section B.

With a wave travelling on the right in the primary line, the net effect is to produce only one resultant wave in the secondary system, which travels in the direction of A. By terminating section A of the secondary system in its characteristic impedance, this induced wave is absorbed. (If a voltmeter is connected across it, it will read voltage V_1.)

The relative magnitude of electric and magnetic coupling in Fig. 20.9(b) can be readily controlled by the design of the coupling loop D. The electric coupling depends on the amount of electric field that terminates on the loop, and is hence determined by the length of the loop and the diameter of the conductor.

Similarly, the magnetic coupling is determined by the amount of magnetic flux that links with the loop, which is determined by the area enclosed between the loop and the outer conductor of the line, and the orientation of the loop with respect to the axis of the line.

Assuming that the coupling arrangement in Fig. 20.9(a) has been designed so that a wave travelling to the right on the primary system produces no induced wave in section B, consider the effect of a wave travelling to the left in the primary system.

The component waves induced in A and B by the electric and magnetic fields in the primary co-axial line will again be equal to each other, since their magnitudes are not affected by the direction of travel on the primary line.

However, the polarity of the wave produced by the magnetic coupling is now reversed with respect to the polarity of the induced wave resulting from electric coupling.

Accordingly, the two waves induced in A now cancel each other, while the two waves produced in B add. Hence, a wave travelling to the left in the primary line produces no effect in section A, but does produce an induced wave in section B. By terminating B of the secondary system in its characteristics impedance, this induced wave is absorbed. (Also, if a voltmeter is connected across it, voltage V_2 is obtained.)

The final result is then that any wave that appears in section A is determined only by the wave travelling to the right in the primary system, and is independent of the presence or absence of a wave travelling to the left in the primary system. Hence, one has achieved a directional coupling system.

Directional couplers are extensively used for monitoring and measuring power in microwave systems.

In particular, radar transmitters have built in directional couplers for monitoring power. They are also used in the determination of reflection coefficients.

Two similar directional couplers can be used to measure the net power delivered to a load if arrangements are made such that one coupler samples the incident power and the other coupler extracts the same fraction of the reflected power. The difference in the power outputs of the two couplers then gives a measure of the load power. The reflection coefficient can be determined as

$$|\rho| = \frac{S-1}{S+1}$$

where S is the standing wave ratio given by

$$S = \frac{|E_i| + |E_r|}{|E_i| - |E_r|}$$

By terminating both A and B in well matched thermocouple elements, so that their opposed output is read on a micro ammeter, a wide band type of wattmeter can be realized for measuring the power in an arbitrary load.

Review Questions

1. How is power at AF, RF and microwave frequency measured?
2. What are the requirements of a dummy load?
3. What is significance of using a dummy load?
4. What is a bolometer?
5. State the basic principle of working of a bolometer.
6. How is power measured using a bolometer?
7. Why is a bolometer used in bridge configurations?
8. List different types of bolometer element.
9. Describe in brief the construction of a bolometer element and bolometer mount.
10. Explain with diagram the operation of a bolometer bridge for measurement of power.
11. Describe the procedure of measuring power using a bolometer in a bridge circuit.
12. What is an unbalanced bridge? State its advantages and disadvantages.
13. Explain with diagram the operation of an unbalanced bridge.
14. How is calibration done in the unbalanced bridge?
15. What do you understand by self-balancing?
16. How is a large amount of power measured in the RF range?
17. State the basic principle of the calorimetric method.
18. Explain with a diagram, measurement of power in a lossy cable using calorimetric method.
19. Explain with a diagram, measurement of power in a coaxial cable using calorimetric method
20. Why is the titanium dioxide section tapered?
21. How is calibration done in the calorimeter?
22. How is power measured in a transmission line?
23. Explain measurement of power in a transmission line with a diagram.
24. What are standing waves?
25. How is standing ratio measured?
26. Derive the expression for a standing wave ratio.
27. What are directional couplers? How do they function?
28. How is standing wave ratio measured using directional couplers?
29. State the advantages and disadvantages of using directional couplers?
30. Describe with a diagram the measurement of power using directional couplers.

Multiple Choice Questions

1. Bolometers are used to measure power in
   (a) AF range (b) RF range
   (c) ac mains range
   (d) microwave range
2. Power at RF is measured using a
   (a) $P = E^2/R$ or $P = I^2R$
   (b) voltmeter
   (c) directly (d) wattmeter
3. Bolometers are used to measure
   (a) small power
   (b) medium power
   (c) large power
   (d) very large power
4. Power at microwave frequencies are measured using
   (a) $P = E^2/R$ (b) $P = I^2R$
   (c) directly (d) wattmeter
5. A metal wire bolometer is referred to as
   (a) thermistor (b) barreter
   (c) resistance (d) thermocouple
6. In the calorimetric method the load used is
   (a) resistance (b) impedance
   (c) water (d) inductance
7. A thermopile is
   (a) inductance in series of
   (b) Resistance in series
   (c) thermistor in series
   (d) thermocouple in series.
8. Directional couplers are used to measure
   (a) voltage (b) current
   (c) impedance (d) power.

Practice Problems

1. A small AF voltage of 10 V superimposed on the RF test power and balance is achieved. If the RF test power is now turned off, 20 V AF is required to balance the bridge. If the bridge arms has a resistance of 50 Ω, calculate the RF test power.
2. Calculate the power radiated by a transmission line having the following parameters. Mass = m = 250 gm, Sp heat of water = sp = 1 cal/g°C, Initial temperature = 25°C and final temperature = 40°C.
3. Determine the initial temperature for the following parameters of a transmission line. The radiated power is 7.56 kW , Mass = m = 300 g, Sp heat of water = sp = 1 cal/g°C and final temperature = 30°C.
4. If the maximum value of voltage is 10 V and the minimum value of voltage is 3 V in a standing wave pattern. Determine the standing wave ratio.
5. Determine the standing wave ratio. If the maximum value of voltage is 12 V and the minimum value of voltage is 8 V in a standing wave pattern

Further Reading

1. Terman and Petit, *Electronic Measurements*, McGraw-Hill, 1952.
2. *Handbook of Electronic Measurements*, Polytechnic Institute of Brooklyn, Vols. I and II (Microwave Research Institute), 1956.

Chapter 21

Control Systems

Multivariable systems are used more commonly in this age of Automatic control. Control parameters are becoming more complex by the day and the range of applications wider. In such complex systems, there are more than one input along with noise signals which are also present and there are more than one output. The physical variables such as flow, pressure, temperature, humidity, pH, etc. are to be controlled very closely. These variables may be interde-pendent. In general, the process control systems may be broadly classified as follows:

1. Open loop control systems.
2. Closed loop or feedback control system.
3. Feed forward control system.
4. Cascade control system.
5. Ratio control system.

In this section only closed loop control systems are discussed.

BASIC CONTROL ACTION 21.1

An automatic controller in a process plant works on the principle of a closed loop control system. It compares the actual value of the plant output with the desired or set-point value, a deviation (or error) is obtained which in turn produces a control signal that reduces the error to zero or to an acceptable minimum value. Figure 21.1 shows the diagram of a basic controller.

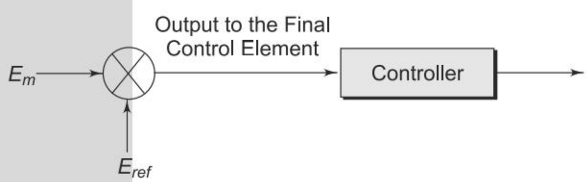


Fig. 21.1 Basic (typical) controller

The error signal is given by

$$e = E_m - E_{\text{ref}}$$

where e = Error signal to be fed to controller

E_m = measured value of the plant variable

E_{ref} = reference or set point value of the variable

The manner in which the control signal (P) produced from the automatic controller, is fed to the final control element is called the *control action.*

DEFINITION (TERMINOLOGY) 21.2

21.2.1 Error

$e = E_m - E_{ref}$ is the expression for error. But, error is most often expressed as a percentage of full scale variable range. It is given by the expression:

$$E_p = \frac{E_m - E_{min}}{E_{max} - E_{min}} - \frac{E_{ref} - E_{min}}{E_{max} - E_{min}} \times 100\% = \frac{E_m - E_{ref}}{E_{max} - E_{min}} \times 100\%$$

where E_p = error as percentage of range from set-point

E_m = measured value of the process variable

E_{ref} = reference, set point of the variable

E_{max} = maximum value of variable

E_{min} = minimum value of variable

$E_{max} - E_{min}$ = full scale variable range

If E_p is positive, the measured value of the process variable is above the set point and if E_p is negative, it is below the set point value.

In process plant control, the standard range for a process variable is 4–20 mA. If this standard is used, then the current can be substituted to compute the value of deviation.

Example 21.1 *Find the percentage error in measurement if the variable range is 4–20 mA and the measured value is 13 mA with a set point of 10 mA.*

Solution The percentage error is given by

$$E_p = \frac{E_m - E_{ref}}{E_{max} - E_{min}} \times 100\%$$

where given

E_m = 13 mA

E_{ref} = 10 mA

E_{max} = 20 mA

E_{min} = 4 mA

$$\therefore \quad E_p\,\% = \frac{13\text{ mA} - 10\text{ mA}}{20\text{ mA} - 4\text{ mA}} \times 100\%$$

$$= \frac{3\text{ mA}}{16\text{ mA}} \times 100 = \frac{3}{\cancel{16}_{4}} \times \cancel{100}^{25} \times 100 = \frac{75}{4}$$

$$= 18.75\%$$

Since E_p is positive, the measurement is above set point.

Example 21.2 *Find the percentage error in measurement if the variable range is 4–20 mA and the measured value is 7 mA with a set point of 10 mA.*

Solution The % error is given by

$$E_p = \frac{E_m - E_{ref}}{E_{max} - E_{min}} \times 100$$

where

$$E_m = 7 \text{ mA}$$
$$E_{ref} = 10 \text{ mA}$$
$$E_{max} = 20 \text{ mA}$$
$$E_{min} = 4 \text{ mA}$$
$$E_p = \frac{7 \text{ mA} - 10 \text{ mA}}{20 \text{ mA} - 4 \text{ mA}} \times 100$$
$$E_p = -\frac{3\,\text{mA}}{16 \text{ mA}} \times 100 = \frac{-75}{4}\% = -18.75\%$$

Since E_p is negative, the measurement is below the set point.

21.2.2 Step Function Response

The step function response consists of applying an instantaneous change in the control means and then recording graphically the result of the change on the measured variable. This graphical record obtained in response to the step excitation is called the Process Reaction curve. Two typical process reaction curves are shown in Fig. 21.2.

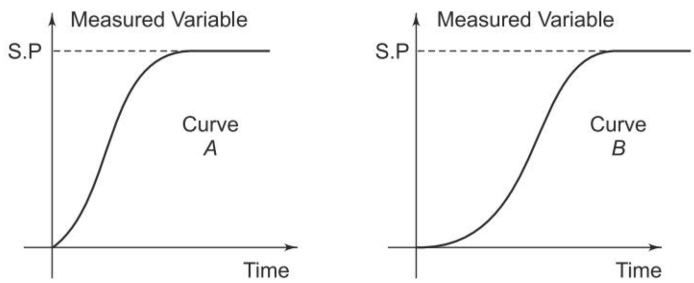


Fig. 21.2 Process reaction curves

Curve *A* shown in Fig. 21.2, describes a simple process in which the measured variable begins to change as soon as the step change occurs in the position of the final element. This means that there is minimum or no resistance to an energy change in the process. At the start, the rate of change is very rapid, but as the process continues, the rate of change slows down.

Hence, the process stores up a part of the energy change. This characteristic of a process is called its capacity. The rate of change of the process reaction curve is called the *process reaction rate.*

Curve *B* describes a more complicated process in which there is resistance, more than one capacity and another process characteristic called the *dead time.*

Dead time is defined as the time gap between the instant deviation or error occurs and the instant when the corrective action starts the first time.

Curve *B*, indicates, that at the start and for a short time thereafter, no action or nothing happens to the value of the measured variable due to the effect of the dead time. When the variable does begin to change, it does so slowly at first, then speeds up until it approaches the final value and then again it slows down. This change in reaction rate is due to the combination of capacity and resistance, into another capacity. This characteristic is called the *transfer lag.*

A process having a reaction curve similar to curve *A* can be controlled by the simplest form of controller. A process produces a reaction curve similar to curve *B* presents some difficulties. Hence, to select the appropriate controller for a process, an understanding of the control action and of their ability to handle the process characteristics is essentially required.

ON–OFF CONTROL ACTION 21.3

The ON–OFF or two position controller is the simplest, cheapest and the most used controllers. It is used in domestic heating systems, refrigeration, water tanks, etc. When the measured variable is below the set point, the controller is ON and the output signal has maximum value. When the measured variable is above the set point, the controller is OFF and output is zero.

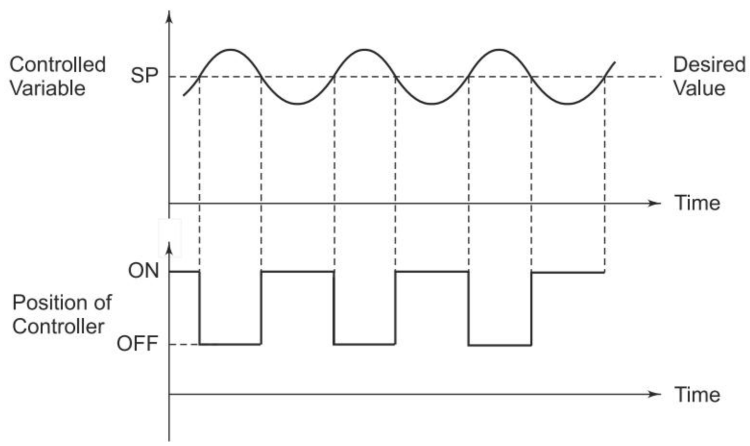


Fig. 21.3 Characteristics of an ON–OFF control action

Due to mechanical friction or arcing of electrical contacts, the controller actually goes on slightly below the set point and goes off slightly above the set point. This differential gap in the controller output may be deliberately increased to give decreased frequency of operation and reduced wear.

Dead time causes the value of the measured variable to go beyond the limit set by the differential gap, since the presence of dead time means a delay in corrective action of the controller. The greater the dead time, the greater the amplitude and period as shown in Fig. 21.4.

Transfer lag effect differs from the dead time effect, in that there is no delay in response, but rather, a slowing down of the response. Figure 21.5 shows the effect of transfer lag on a two position controller.

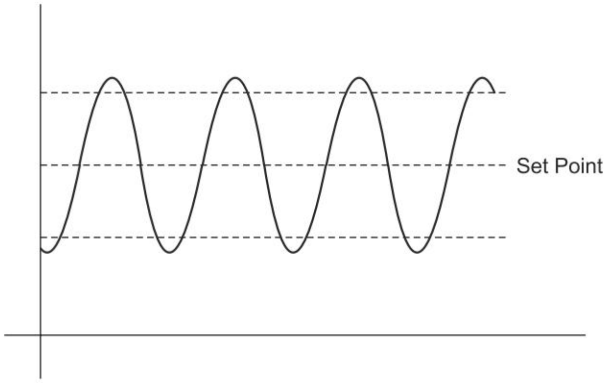


Fig. 21.4 Effect of dead time ON–OFF action

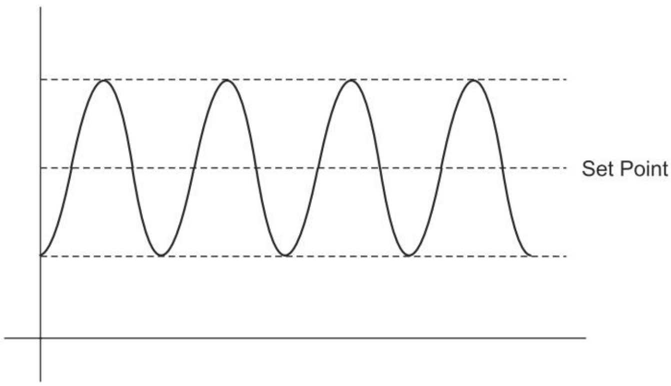


Fig. 21.5 Effect of transfer lag on ON–OFF controller

With both the dead time and the transfer lag, the measured variable exceeds the controller differential gap, as seen from Figs 21.4 and 21.5. With the transfer lag, all the changes are more gradual. Hence, the measured variable peaks due to the effect of transfer lag are rounded off on the two position controllers and are not as sharp as they are due to dead time effect.

The favourable conditions for using ON–OFF control action are as follows:

(a) The reaction rate should be slow.
(b) There should be little or no dead time.
(c) There should be little or no transfer lag.

However, in the ON–OFF control action, the process is never at the set point except momentarily on its way up and down. ON–OFF control action is however most suitable in practice due to its simplicity and low cost.

PROPORTIONAL CONTROL ACTION 21.4

In many processes, the cycling that occurs with the two position ON–OFF control is undesirable. Steady process controlled in the absence of noises, is possible when the controlled variable is a continuous function of the error.

A most widely used form of continuous control mode is the proportional control action, in which a smooth, linear relation exists between the controller output and error. Hence, proportional control action can be represented as

$$P = K_p.E_p + P_o$$

where P = controller output, including error
P_o = controller output without error (%)
E_p = error signal in proportional control action
and K_p = proportional controller gain or proportional constant between error and controller output (%)

The above equation indicates one-to-one correspondence between E_p and P. The control action can also be expressed in terms of the *Proportional Bandwidth* (PB), defined as the error required to cause a 100% change in controller output and is usually expressed as percentage of the chart width.

Hence PB = $1/K_p \times 100\%$

From the above equations for proportional control action the following can be concluded

1. If error is zero, the controller output is constant ($=P_o$).
2. In case error signal is present, the proportional controller output needs a correction of K_p% for every 1% error (E_p).
3. There is an error band about zero, having magnitude = PB, within which the output is not saturated at 0% or 100%.

The characteristics of a proportional control is as shown in Fig. 21.6 (a).

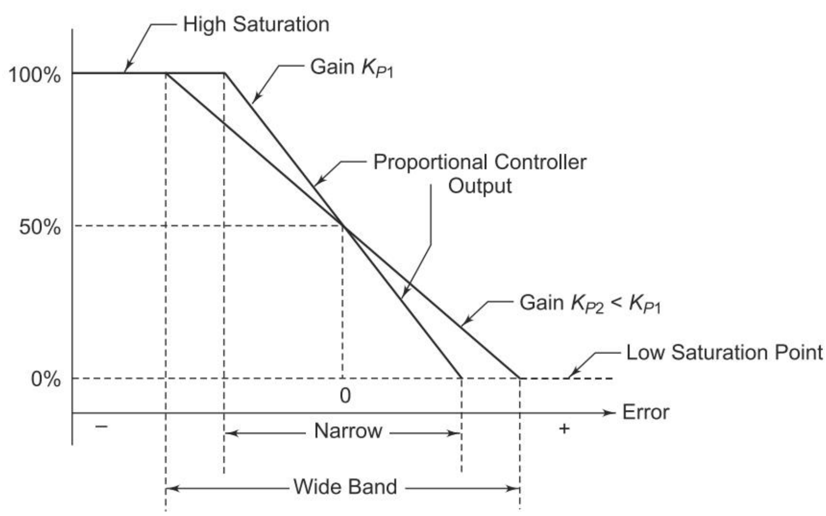


Fig. 21.6 (a) Characteristics of proportional control action

OFFSET 21.5

Proportional control action is characterized by a permanent residual error in the operating point of the controlled variable when a change in load occurs. This error is called the OFFSET.

The offset can be reduced by selecting higher value of controller gain (K_p) corresponding to narrow bandwidth.

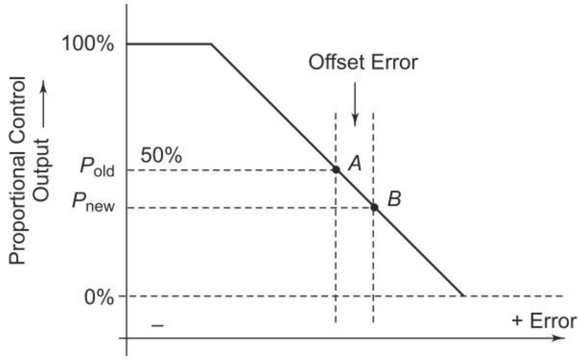


Fig. 21.6 (b) Offset error

Figure 21.6 (b), shows how offset error occurs in a proportional control action. Let the system error be zero at nominal load with controller output $P = 50\%$ corresponding.

If, however, a transient error occurs, the system tends to adjust the controller output so that the point A (corresponding to zero error) is reached again. But for this to happen there must be change in the load system. This changed controller load produces a new value (P_{new}) of the characteristic output, which gives rise to point B on the linear characteristic of the proportional controller. The permanent small difference between the percentage error values corresponding to point B and A is called the *offset error* of the f^{∞} control.

BASIC CONTROLLER CONFIGURATION 21.6

The primary requirement in most industrial processes is the control of typical process variables like temperature, pressure, flow or level, at reference value called set point. An instrument used to achieve this objective is called a controller.

It basically consists of a signal representing the value of a process variable, compares it with the set point value, and acts to minimize the difference between the two signals. A simplified block diagram of a controller configuration is shown in Fig. 21.7.

The function of a controller is:

(a) receiving a signal corresponding to the measurement to be controlled,
(b) comparing that value to a reference set point value,
(c) determining the difference or error in the measurement value, and
(d) providing a controlled output.

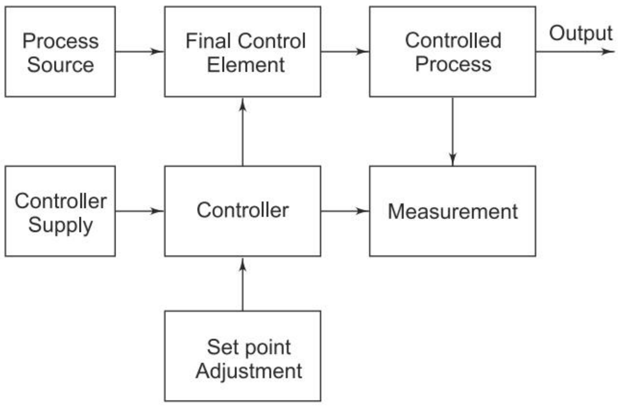


Fig. 21.7 Block diagram of a controller configuration

CLASSIFICATION OF CONTROLLERS 21.7

Controllers can be classified on the basis of signal transmission used, such as Hydraulic, Pneumatic and Electrical and/or Electronic Controller. (In this book only electronic controllers will be discussed).

ELECTRONIC CONTROLLERS (EC) 21.8

In certain industrial plants, process requires very fast response from the process control instruments. For such type of processes, electronic controllers are preferred to pneumatic or hydraulic controllers. The main component of all electronic controllers is the very high gain Op-Amp.

21.8.1 Advantages of Electronic Controllers

The main advantages of electronic controllers are as follows.

1. The main advantage of electronic controllers is their compactness.
2. They respond faster and have a decided advantage if fast response time is needed.
3. The installation cost is less for large plants.
4. They provide convenient, economic interfacing with supervisory digital computers, data processing or data acquisition systems.
5. Reliability is high.
6. There is a gradual lowering of electronic component and subsystem cost.
7. Accuracy is about ±0.25% as compared to ±0.5% for pneumatic controllers.

21.8.2 Disadvantage

The only disadvantage is that the electronic controllers are very much susceptible to fire risk. This can be avoided with proper precautions.

ANALOG ELECTRONIC PROCESS CONTROLLERS 21.9

An electronic controller must be capable of providing one or more of the three main methods of control, namely, Proportional, Integral and Derivative.

The circuits described in this section use high gain OPAMP'S with specially designed feedback circuits. They operate with their phase inverting input as a virtual earth point. If the amplifier saturates due to, say, a high value of input signal (i.e. deviation signal), the phase inverting input may no longer remain virtual earth point and circuit theory becomes invalid.

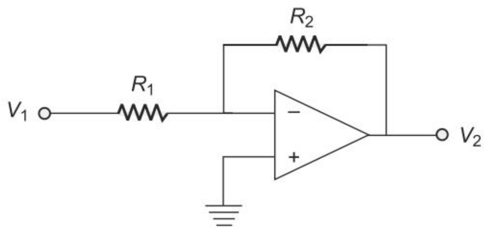


Fig. 21.8 Proportional control

The simplest form is the proportional control shown in Fig. 21.8, which has an overall voltage gain of

$$\frac{V_2}{V_1} = -\frac{R_2}{R_1} \qquad (21.1)$$

Hence, $V_2 = -V_1R_2/R_1$, so that the output voltage is proportional to the input voltage. The value of the proportional band of the controller is modified by changing the value of R_1 and R_2. The input voltage V_1 in all the circuits in Fig. 21.8 is either deviation (error signal) or it is amplified version of the deviation. The negative signal associated with Eq. 21.1 implies that the output voltage has the opposite polarity of V_1.

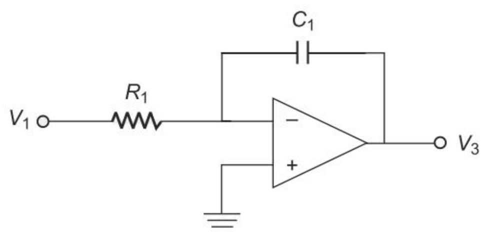


Fig. 21.9 Integral control

Pure integral action is as obtained from the circuit in Fig. 21.9. Integral circuit is often used in conjunction with proportional control (two-term control or PI control) as a means of reducing the steady state deviation or offset.

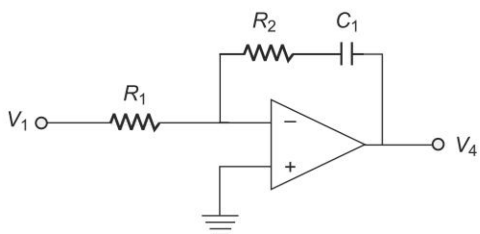


Fig. 21.10 Proportional plus integral control

Figure. 21.10 is a realistic version of proportional plus integral controller. The magnitude of the proportional gain factor is given by R_2/R_1 and the integral action time is R_1C_1 (R in Ω and C in Farads).

The integral action time of a proportional plus integral controller can best be explained in terms of the output waveform from the controller for a step change or sudden change in the input (deviation) signal shown in Fig. 21.11.

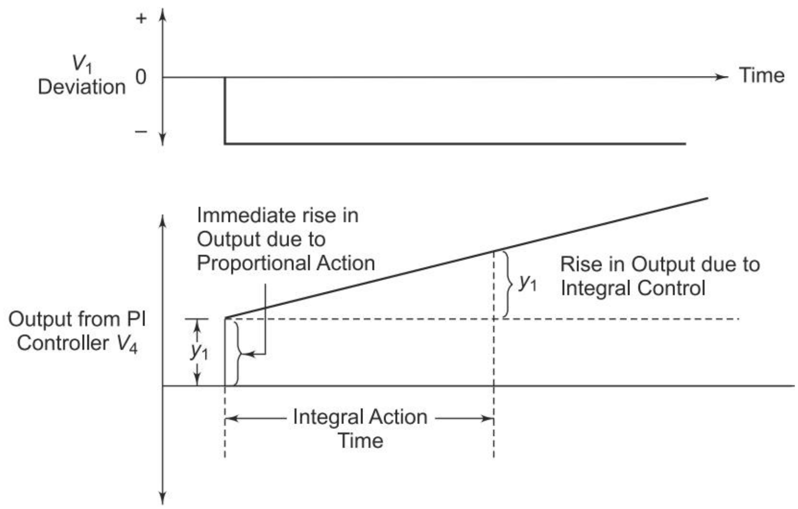


Fig. 21.11 Integral action time

Since the OPAMP in Fig. 21.10 is an inverting amplifier, a negative step change in input voltage is applied to produce a positive output voltage change. As soon as the step change is applied, the P action of the controller causes the output voltage to change suddenly by Y_1 volts. The action of the integrator capacitor then causes the controller output voltage to begin to rise at a steady rate and it takes a time known as the *integral action time* for the output voltage to rise by a further Y_1 volts.

Derivative (D) control provides a stabilizing influence on the system and is desirable in a system in which the value of the controlled variable changes rapidly. A basic proportional plus derivative controller is shown in Fig. 21.12, the P action being provided R_1 and R_2 , while the derivative action is provided by capacitor C_2 together with R_3 .

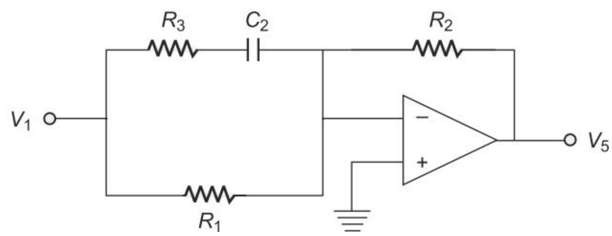


Fig. 21.12 Proportional + derivative control

When the system is in its steady state, the value of the deviation signal (V_1) is constant and capacitor C_2 is charged to a constant value equal to V_1 (the non-inverting terminal being at virtual ground). Hence no current flows through C_2, under this condition the magnitude of the voltage gain (V_5/V_1) is equal to R_2/R_1.

However, when the deviation (V_1) changes rapidly, capacitor C_2 charges rapidly and presents little impedance to the flow of current. The net effect is that R_3 can be considered as being in parallel to R_1, but since R_3 value used in a practical circuit is typically one tenth that of R_1,the combined resistance of the input circuit is very nearly equal to R_3. The amplifier gain under this conditions is approximately equal to R_2/R_3 which is 10 times greater than the gain under steady state conditions. That is, the action of the derivative control while the deviation is changing rapidly is to reduce the proportional bandwidth of the controller. The net effect is to cause a more rapid change in the controller output, thereby resulting in a more rapid change in the controlled variable. When steady state conditions are finally reached, capacitor C_2 is charged to the new steady value of V_1, and the magnitude of the gain again becomes R_2/R_1. The derivative action time of the proportional plus derivative controller is R_1C_2.

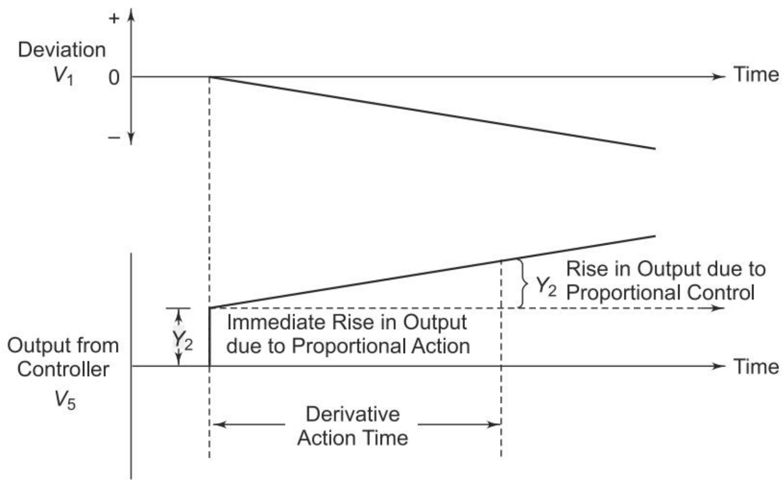


Fig. 21.13 Derivative action time

The derivative action time of a *P* plus *D* action controller can be explained in terms of the output waveform of the controller for a ramp change in input signal, in other words for a constant change in the deviation signal shown in Fig. 21.13. Since the op-amp is inverting, a negative ramp signal is applied in order to get a positive going output voltage. As soon as the amplifier change in the input signal is applied. The derivative action of capacitor C_2 causes the output voltage to suddenly change by Y_2 volts. The *P* action to the controller causes the output voltage to rise at a constant rate, and takes a time known as the *derivative action time* for the output voltage to rise by a further Y_2 volts. In this case, the derivative action time is given by R_1C_2 of the controller.

The block diagram of a 3-term controller providing P + I + D (PID) control is shown in Fig. 21.14.

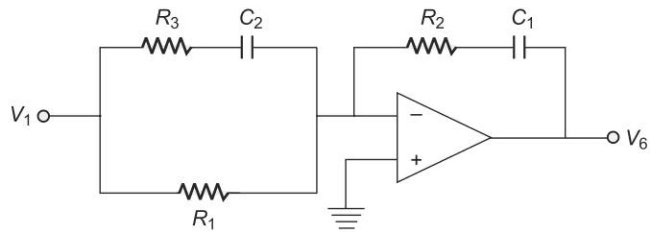


Fig. 21.14 Proportional plus integral plus derivative (PID)

Resistors R_1 and R_2 provide the basic P control, while capacitor C_1 introduces integral control and capacitor C_2 derivative control.

A pactical form of 3-term controller together with a clipping circuit or desaturation circuit which prevents the virtual earth concept being violated if the amplifier saturates is shown in Fig. 21.15.

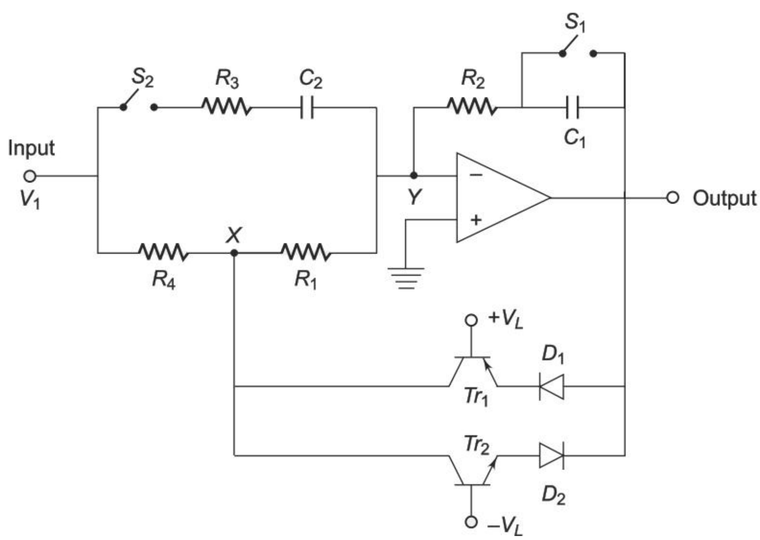


Fig. 21.15 A practical version of an electronic 3-term controller

The clipping circuit, consists of transistors Tr_1 and Tr_2 together with diodes D_1 and D_2. The diodes do not function so long as the output voltage does not exceed about $\pm (V_L + 1)$ volts, if $V_L = 7$ V then clipping circuit does not come into operation for any output voltage within the range +8 to –8 V. If the output voltage tends to exceed +8 V, due to say, sudden change in V_1, diodes D_1 and transistor Tr_1 begins to conduct, and connect point X to the output terminal. This results in the change being diverted through resistor R_4 whose value is very low when compared with that of R_1 (the value of R_4 is typically 0.5% of R_1), with the result that point Y remains virtual ground. Similar effects take place by Tr_2 and D_2 for a negative output voltage greater than about –8 V (assuming, i.e. $-V_L = -7$V).

When switch S_1 is closed and S_2 is open, the circuit provides proportional action only, the magnitude of the voltage gain being R_2/R_1, the proportional band of the controller can be changed by changing value of this ratio.

With S_1 and S_2 open, the controller becomes a proportional plus integral controller with an integral action of $T_1 = R_2C_1$.

With S_1 closed and S_2 closed, the circuit provides proportional plus deriva-tive action with a derivative action time of $T_2 = R_1C_2$.

With S_1 open and S_2 closed, 3-term control is obtained, the integral action time T_I and derivative action time T_D of this controller are

$$T_I = T_1 + T_2 \tag{21.2}$$

and

$$T_D = \frac{T_1 T_2}{T_1 + T_2} \tag{21.3}$$

where T_1 and T_2 are given above. The magnitude of the overall voltage gain of the amplifier is given by

$$\frac{R_2}{R_1} + \frac{C_2}{C_1} \tag{21.4}$$

while the output forms an electronic signal, it can be converted into air pressure by a suitable interface device.

TEMPERATURE CONTROL USING AN ANALOG ELECTRONIC CONTROLLER 21.10

A temperature control scheme using the electronic controller of Fig. 21.14 is as shown in Fig. 21.16.

The temperature setting, derived from a slide wire, is compared with the signal from a thermocouple situated in the oven being heated. The difference between these two voltages is amplified in an error amplifier whose gain can be adjusted by means of potentiometer R_V, this potentiometer provides a means of modifying the proportional bandwidth of the controller.

The output from the error amplifier is applied to the output of the 3-term controller sections (whose operation is discussed in Sec. 21.9).

The output voltage from this controller is applied to the input of a pulse generator which controls the delay angle of the pulses applied to the gate electrodes of the thyristor which regulate the current in the heating element. As the temperature of the oven rises, the gate pulses are phased back so that the oven power is reduced to a value which maintains the desired temperature.

To prevent excess current flowing in the thyristor or the heating element, electronic overcurrent protection is provided by the controller as follows. The ac current in the load is measured by means of a current transformer CT, and this current is converted into a direct voltage by means of the bridge rectifier BR and a potentiometer. Under normal operating conditions the voltage V_x at the wiper of the potentiometer is less than the breakdown voltage V_z of the zener diode, and

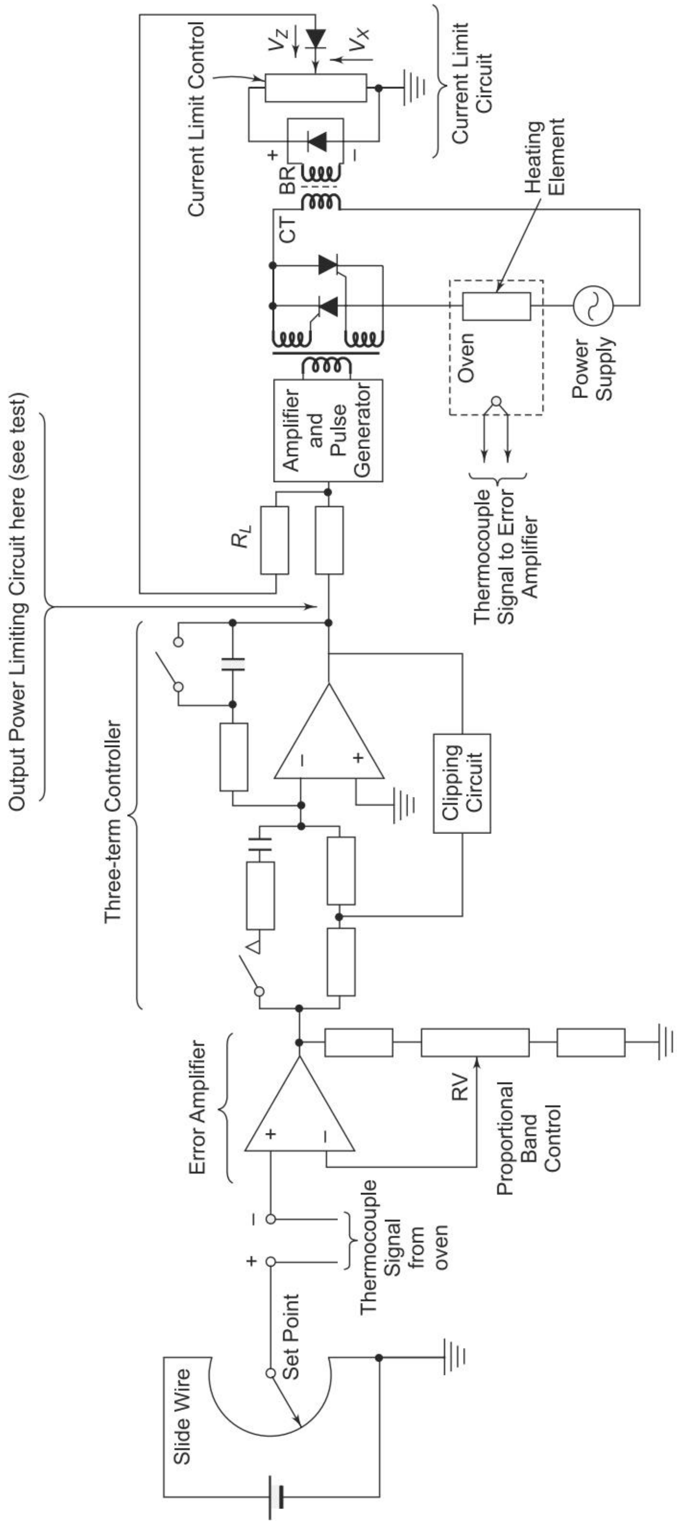


Fig. 21.16 Temperature controller scheme

no current flows through the zener diode. When excess current flows in the load, V_x exceeds V_z and current flows through the zener diode and into resistor R_L. The direction of this current is such that it causes the thyristor gate pulses to be phase backed to reduce the current flowing in the load to a safe value. The current limit setting is controlled by means of potentiometer wiper.

CHOICE OF ELECTRONIC TRANSMISSION SIGNAL 21.11

The range of air pressure used for transmission of signals in pneumatic system is limited by the dynamic operating range of the pneumatic devices used. But the choice of electrical transmission is much wider. Since either ac or dc current may be selected. In addition, a choice may be made between a voltage or a current level, the range of signal levels has also to be decided.

A dc current transmission has the advantage over ac transmission in that only the resistance of the transmission lines need to be considered. The effects of line inductance and capacitance can be ignored. Certain types of measuring device operate on ac current, but the output from these can be converted into dc by relatively simple means.

A current rather than a voltage is normally used as the method of signal transmission, since the effects of line resistance have very little effect on the accuracy of the information transmitted. An output current is obtained from transducers and controllers by providing current feedback in the amplifier used. Wherever necessary, the current signal can be converted into a voltage signal by connecting a resistor in the line.

The range of current values used varies widely, typically ranges being 0–10 mA, 0–20 mA, and 4–20 mA. The live zero of 4 mA permits transducers to use the line characteristic as a source of power for its amplifier as shown in Fig. 21.17

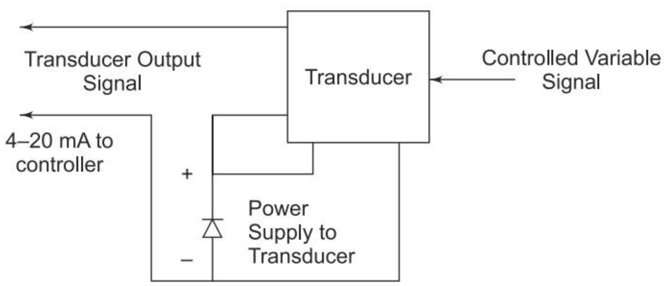


Fig. 21.17 Two-wire transmission system with a live zero

By this means, a two-wire transmission system is obtained. A disadvantage of the line zero is that the controlled variable zero does not correspond to a true zero of the transmitted current. This presents problems where mathematical operations, e.g. a square root function, have to be carried out on the current.

DIGITAL CONTROLLERS 21.12

Developments in digital electronics have led to many industrial processes being computer controlled. The first system used was Direct Digital Control (DDC), in which a computer measured each variable in the process, these signals being used to maintain the required set points in the process.

As digital electronics developed, minicomputers become more competitive, and with the advent of electronic Multiplexer (Mux) and Demultiplexer (Demux), it is possible to sample the state of many analog and digital transducers at high speed and also to control many values or devices in the plant.

A basic DDC system is shown in Fig. 21.18. In this system a large number of transducers are sited around the plant, each transducer being connected to one input of a Mux. (A Mux can be considered as the electronic equivalent of a switch with a contact or blade which rotates very rapidly so that it moves from one transducer to another, the blade remaining in contact with the transducer long enough for an ADC to sample and digitise or to quantise the analog signal. The quantised data are then transmitted along the data bus of the system to the CPU).

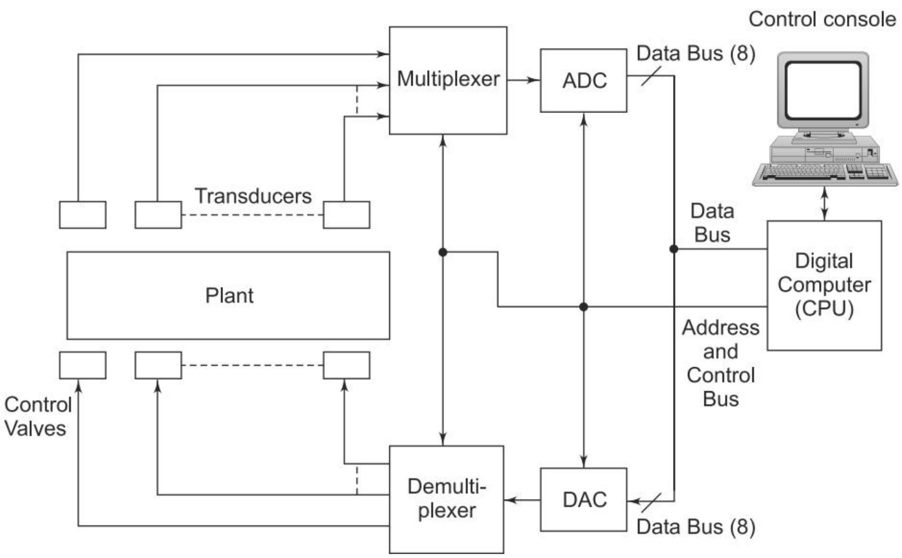


Fig. 21.18 Direct digital control of an industrial process

When the CPU has analysed the data from one or more transducers and has compared them with the appropriate set points in the computer program, it sends signals along the data bus to the values controlling the system as follows. The digital signal produced by the CPU is converted into an analog signal by a DAC and the analog signal is transmitted to the appropriate control value through a demux. As this is being performed, data are displayed on the operator's visual display unit and if necessary, he can remotely change the set points associated with various sections of the process.

An alternative control system known as Computer Supervisory Control (CSC) is used in microprocessor based systems. In this case, the process is controlled by a number of local feedback loops using individual process controllers, the main computer merely acting in a supervisory role in which it monitors the measured variable (although it has the capability of remotely controlling set points of the controllers). This method provides the local operator with his own controller.

A basic CSC system is shown in Fig. 21.19. Each digital controller controls one section of the plant by means of a closed loop involving its own sensors and control values. Each controller can either be operated manually to give the local operator complete control of his own section of the plant or it can be remotely controlled by the computer operator.

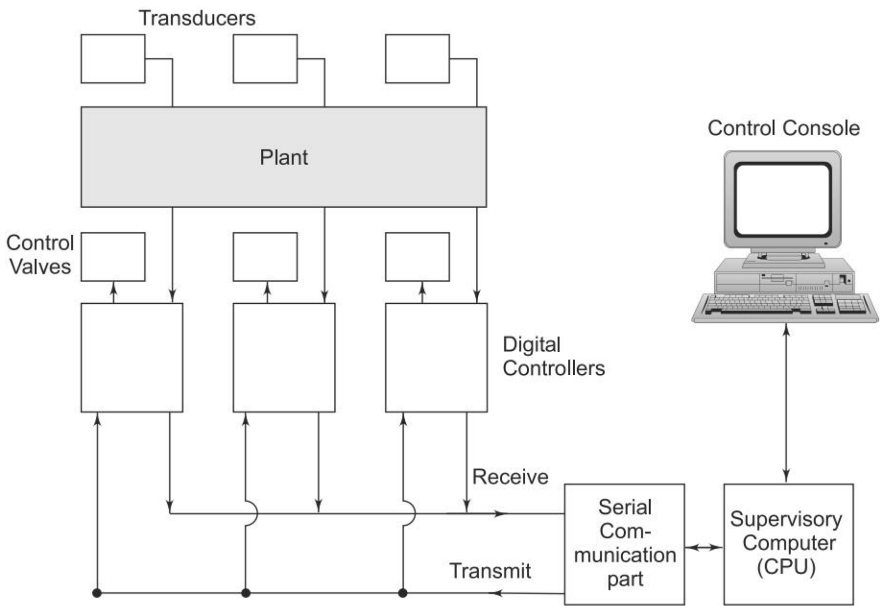


Fig. 21.19 Computer supervisory control (CSC) of an industrial process

The computer is connected to each controller by means of a communication port which allows the computer either to receive information from the process controllers or to transmit data to the controllers to change, say, a set point. To minimise the cost of cabling, the communication port operates in a serial mode, that is, the data are sent along the communicating link in the form of pulses (which requires only two wires—a "send" wire and a "receive" wire). This method is clearly more economical than using a 8-bit data bus throughout the plant, but it is slower in operation, since it takes more time to send eight individual bits along one wire then it takes to send 8-bits simultaneously along eight wires. The digital process controller (also called the central computer) must contain a special interface which converts the pulses it receives in a serial mode along the supervisory serial data bus into the parallel mode used in microprocessor chip itself. Such an interface is known as a UART (Universal Asynchronous Receiver/Transmitter).

DIGITAL PROCESS CONTROLLER 21.13

A digital process controller functionally needs to operate in much the same way as a conventional analog controller. That is, it must provide facilities to raise or lower the set point and to provide proportional, 2-term or 3-term control. It must be capable of operating in one of many modes including, manual, automatic, remote or ratio (in which the controller, or controllers, controls the ratio of two quantities such as air and gas in an air–gas mixture burner). It should accept analog signals from plant mounted transducers and signal conditioners and transmit analog signals to control valves, etc. The processor software must also provide for square root extraction from appropriate plant signals. A block diagram of a typical process controller is shown in Fig. 21.20.

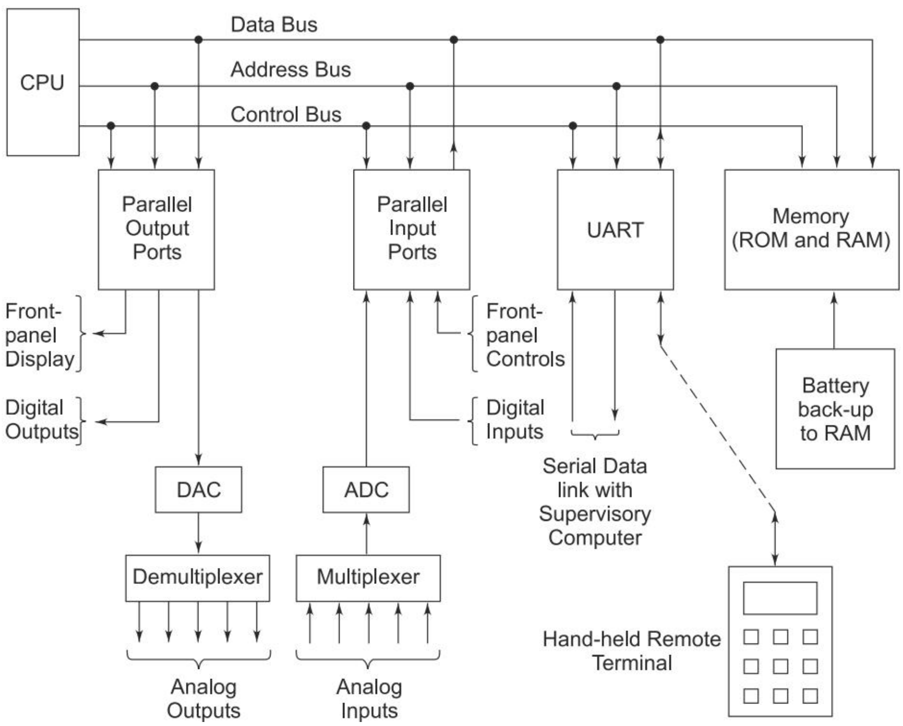


Fig. 21.20 Block diagram of a digital process controller (Based on the TCS 6350 process controller)

The process controller contains output ports which controls a number of functions as follows. The controller has a number of digital displays on its front panel including 7-segment displays, which can be used to display the measured value and/or deviation (although these could be displayed in the form of a bar using light emitting diodes), the front panel display could also indicate the operating mode of the controller. Other digital outputs can be used to provide remote indication of the status of the controller. Yet another output port is associated with a DAC which converts digital data from the microprocessor into analog data used to control the plant, a demux is contained in the controller to permit a number of analog devices to be controlled in this way.

The controller also contains a number of input ports to allow data to be fed to the controller. These data may be obtained from remote switches on the plant or from control buttons on the front panel of the controller. Another input port is associated with an ADC whose inputs are multiplexed so that a number of analog values of process variables can be measured.

The controller contains a UART which allows serial communication between the controller and the central computer. In addition, the UART enables the local operator to use a hand held computer terminal which allows him to communicate with the process controller.

The process controller also has its own memory, i.e. it is an intelligent controller, the main operating program being held in ROM while temporary data are stored in RAM. Since the data stored in RAM (e.g. a new value of set point), should not be lost in the event of a power supply failure which can be even for a short period. The RAM has a battery back up so that temporary data will be retained in the event of a power supply failure.

CASCADE PROCESS CONTROLLER WITH DIGITAL CONTROLLERS 21.14

A simplified system to control the temperature of a process employing a gas burner using two cascaded digital controllers is shown in Fig. 21.21.

Each controller has a number of digital inputs and outputs, and also a number of analog inputs and outputs, the actual inputs and outputs are limited to those used to control the system, although the inputs and outputs are also available to the user in a practical system.

In the case of the master controller, the remote set points enable input which is a digital input. If the controller has a logic 0 applied to it, this disables the analog 'remote set point' input of the master controller. The temperature of the process being controlled is sensed by a Thermocouple (TC). This signal, after amplification, is applied to the process variable analog input of the controller.

This signal is compared with the set point of the master controller, and the resulting deviation produces an output at the analog 3-term output.

This signal becomes the analog remote set point of the slave controller (The remote set point of the slave controller is enabled by the digital 'remote / ratio' status terminal of the master controller), hence the master controller controls the set point of the slave. The analog 3-term control output signal of the slave controller changes the relative rates of the flow of air and gas via a ratio controller which could also be a digital controller. A signal proportional to the flow rate

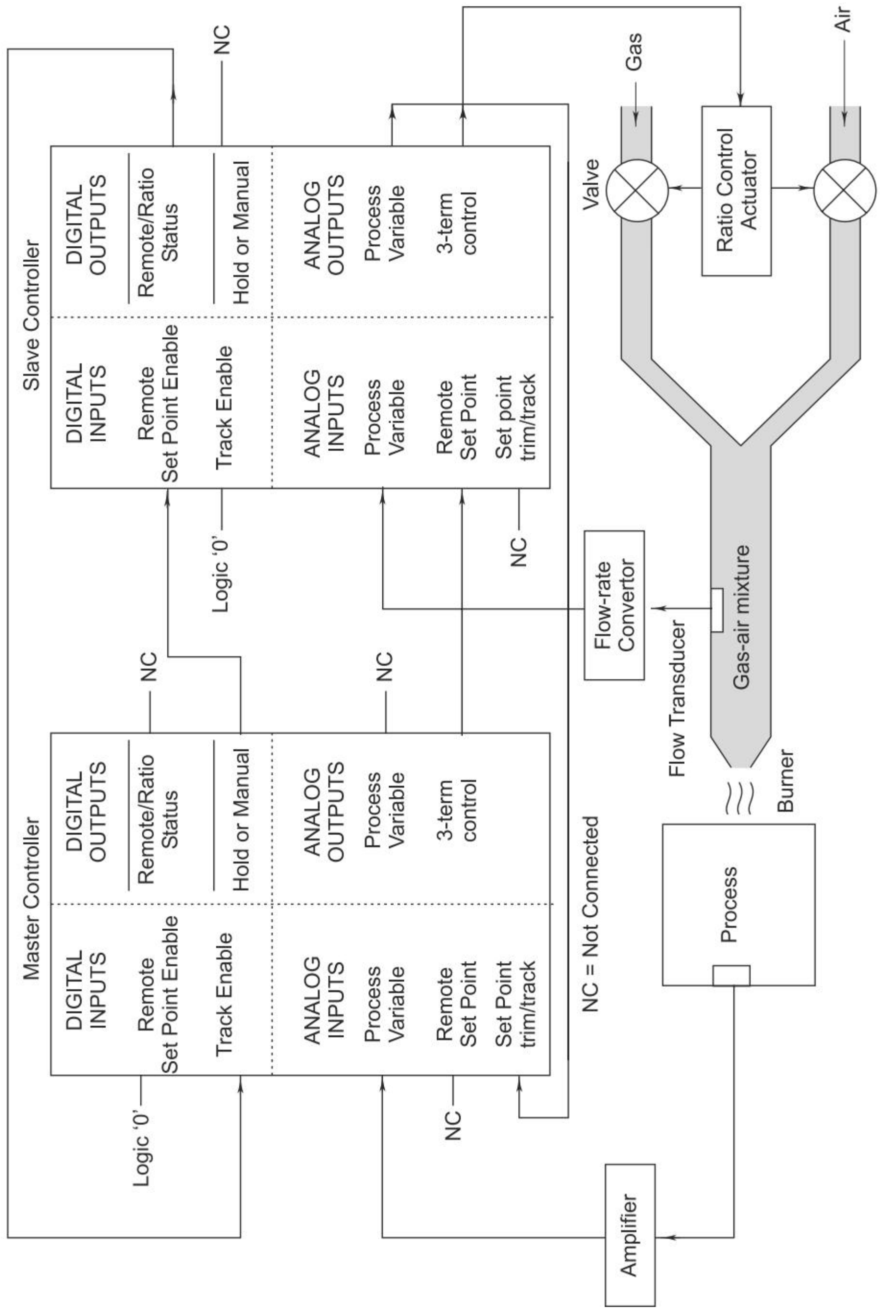


Fig. 21.21 A Cascade process using digital controllers (Based on the TCS 6350 process controller)

of the air-gas mixture is used as the analog process variable input of the slave controller. A copy of this signal appears at the process variable output terminal and is used as a track signal for the master controller. This ensures that, during start up, the 3-term output of the master is forced to follow the process variable signal of the slave, so that the remote set point of the slave is equal to its own process variable signal. The object of this arrangement is to ensure a 'bumpless' transfer from any operating mode into cascade. The track operation of the master controller is enabled by a logic signal applied to the 'track' enable digital input of the master controller.

As the temperature of the process varies, the master controller causes the slave controller to alter the flow rate of the air-gas mixture to result in a stable operating temperature.

PROGRAMMABLE LOGIC CONTROLLER 21.15

Programmable Logic Controller (PLC) or commonly simply called a Programmable Controller, is a solid state, digital, industrial computer.

It is a device that was invented to replace the necessary sequential relay circuits for machine control. The PLC basically operates by looking at its inputs and depending upon their state, turning on/off its output. The user enters a program, normally through software, that gives the desired results.

PLCs are used in many real world applications such as machining, packaging, material handling, automated assembly, etc. Almost any applications that need some type of electrical control has a need for a PLC.

Let's assume that when a switch turns ON, we want to turn a solenoid ON for 10 seconds and then turn it OFF regardless of how long the switch is ON for. This can be done by a simple external timer. But if the process included 10 or more switches and solenoids, then we would need 10 or more external timers. But if the processes also needed to count how many times the switches individually turned ON, then a lot of external counters would be needed. As can be seen, the bigger the process, the more important is the need of a PLC. The PLC can be simply programmed to count its inputs and turn the solenoids ON for the specified time.

Since the PLC is a computer it should be told what to do. The PLC knows what to do through a program that is developed and then entered into its memory. The PLC however, without a set of instructions, is just a black box consisting of electronic components only.

A PLC can control devices such as limit switches, push button, proximity or photo-electric sensors, float switches or pressure switches, etc. to provide the incoming control signals into the unit. The incoming control signal is called the INPUT. These control signals or inputs, interact with the instructions specified in the user program, telling the PLC how to react to the incoming signals. The user program also directs the PLC on how to control field devices such as motor starters, pilot lamps and solenoids. A signal going out of the PLC control to a field device is called an OUTPUT.

PLC can also be defined as per National Electrical Manufacturer Association (NEMA) as a digitally operated electronic system, designed for use in industrial environment, which uses a programmable memory for internal storage of user-oriented instructions for implementing specific functions such as logic, sequencing, timing, counting and arithmetic to control, through digital or analog inputs and outputs, various types of machines or processes. Both the PLC and its associated peripherals are designed so that they can be easily integrated into an industrial control system and easily used in all intended functions.

21.15.1 PLC History

PLC's were first introduced in the late 1960s. The primary reason for designing such a device was to eliminate the large cost involved in replacing the complicated relay based machine control systems.

As and when the production requirements changed, so did the control system. This became very expensive specially when the changes were so frequent. Since relay are mechanical devices, they also have a limited lifetime which required strict rules of maintenance schedules. With so many relays involved, troubleshooting was also quite tedious. A machine control panel include possibly hundreds or thousands of individual relays. The wiring of individual device was also complicated as these relays were individually wired to yield the desired outcome. These new controllers also had to be programmed by maintenance and plant engineers. The PLC was not only a long lifetime, the programming were to be easily performed. They also had to survive the harsh industrial environment. A programming technique was required that most people were already familiar with and that replaced the mechanical parts with solid state ones.

The Modular Digital Controller (Modicon) brought the world's first PLC into commercial productions.

In mid 1970's, the most dominating PLC technologies were sequencer state machines and bit-based PLC. Conventional microprocessor's lacked power to quickly solve PLC logic in all but the smallest PLCs. As conventional microprocessors evolved, larger and larger PLCs were based on them.

In approximately 1973, communication abilities began to appear. The first such system was Modicon's Modbus. The PLC could talk to other PLCs and could be far away from the actual machine they were controlling. They could also now be used to send and receive varying voltage, to allow them to enter the analog world. The lack of standardization coupled with continually changing technology has made PLC communication difficult because of incompatible protocols and physical networks.

In the 1980's, an attempt was made to standarise communications with General Motor's Manufacturing Automation Protocol (MAP). It was also the time when attempts were made for reducing the size of the PLC and making them software programmable through symbolic programming on PC instead of dedicated programming terminals or handheld programmers.

Today the world's smallest PLC is about the size of a single control relay.

The 1990's have seen a gradual reduction in the introduction of new protocols. The latest standard (IEC – 1131-3) has tried to merge PLC programming languages under one international standard. We have PLCs that are programmable in function block diagrams, instruction lists, C and structured text all at the same time.

21.15.2 PLC Structure

The PLC mainly consists of a CPU, memory areas, and appropriate circuits to receive input/output data as shown in Fig. 21.22. A PLC can be considered as a box full of hundreds of thousands of separate relays, counters, timers and data storage locations. (These counters, timers, etc. really do not exists physically but rather they are simulated and can be considered software counters, timers, etc). These internal relays are simulated through bit locations in registers.

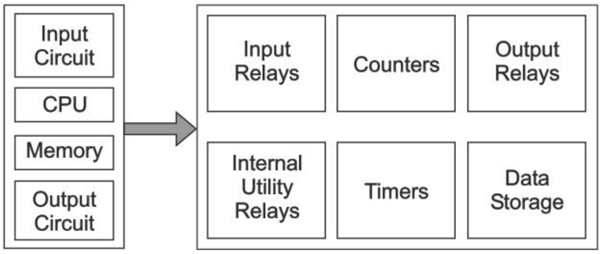


Fig. 21.22 PLC structure

The PLC structure consists of the following

1 *Input Relays:* (Contacts) These are connected to the outside world. They physically exist and receive signals from switches, sensors, etc. Typically they are not relays but are transistors.

2. *Internal Utility Relay:* (Contacts) These do not receive signals from the outside world nor do they physically exist. They are simulated relays and are what enables a PLC to eliminate external relays. There are also some special relays that are dedicated to performing only one task. Some are always ON while some are always OFF. Some are ON only once during Power-on and are typically used for initializing data that was stored.

3. *Counters:* These again do not physically exist. They are simulated counters and they can be programmed to count pulses. Typically these counters can be up-count, down count or both. Since they are simulated they are limited in their counting speed. Some manufacturers also include high speed counters that are hardware based.

4. *Timers:* These also do not physically exist. They come in many varieties and increments. The most common type is an ON-delay type. Others include OFF-delay and both retentive and non-retentive types. Increments vary from 1ms – 1s.

5. *Output Relays:* (Coils) These are connected to the outside world. They exist physically and send ON/OFF signals to solenoid, lamps, etc. They can be transistors, relays or triacs depending upon the type selected.
6. *Data Storage:* Typically there are registers assigned simply to store data. They are usually used as temporary storage for math or data manipulation. They can also be used to store data in case of a power failure. These registers ensure that there is no loss of contents owing to disconnection of power.

21.15.3 PLC Operation

A PLC works by continually scanning a program. This scan cycle can be considered as made up of three important states as shown in Fig. 21.23. In addition there are also more than three states and these are used for checking the system and updating the internal counter and timer values.

The three important states are:

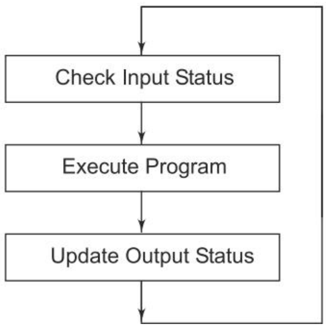


Fig. 21.23 PLC operation diagram

Step 1: Check Input Status First the PLC takes a look at each input to determine if it is ON or OFF. In other words, it checks and senses whether the sensor connected to the first input is ON, to the second input is ON, to the third input is ON… It records this data into its memory to be used during the next step.

Step 2: Execute Program The PLC next executes the program, one instruction at a time. For example, if the program says that if the first input was ON then it should turn ON the first output. Since it already knows which inputs are ON/OFF from the previous step, it will be able to decide whether the first output should be turned ON based on the state of the first input. It will store the execution results for use later during the third step.

Step 3: Update Output Status Finally the PLC updates the status of the outputs. It updates the outputs based on which inputs were ON during the first step and the results of executing the program during the second step. Based on example in step 2, it would now turn ON the first output because the first input was ON and the program said to turn ON the first output when this condition is true.

After the third step, the PLC goes back to step one and repeats the steps continuously.

The time taken to execute the above three steps or one instruction cycle is defined as the *scan time*.

21.15.4 Response Time of PLC

The PLC takes a certain amount of time to react to changes. The total response time of the PLC is a fact that has to be considered while selecting a PLC for some application where speed is a concern.

For example, if you take a moment off to look away from this text and see a picture on the wall. Your eyes actually see the picture before the brain responds to say "Oh there's a picture on the wall". In this example your eyes can be considered as the sensor. The eyes are connected to the input circuit of your brain. The input circuit of the brain takes a certain amount of time to realize that the eyes saw something, eventually when the brain realizes that the eyes have seen something and it processes the data. It then sends an output signal to your mouth which in turn receives this data and begins to response to it. Eventually your mouth comments on the picture.

In this example, we had to respond to three things:

Input It took a certain amount of time for the brain to notice the input signal from the eyes.

Execution It took a certain amount of time to process the information received from the eyes. Consider the program to be 'If the eyes see a good picture then output appropriate worth to the mouth'.

Output The mouth receives a signal from the brain and eventually gives the comment.

Hence, Input Response Time + Program Execution Time + Output Response Time = Total Response Time

Having understood the concept behind the response time, let us see what happens in a typical PLC application. The PLC can only see an input turn ON/OFF when its looking. In other words, it only looks at its inputs during the check input status part of the scan.

In the diagram shown in Fig. 21.24, input 1 is not seen until scan 2. This is because when input 1 is turned ON, scan 1 had already finished looking at the inputs. Similarly input 2 is not seen until scan 3. This is also because when the input 2 is turned ON, scan 2 had already finished looking at the inputs.

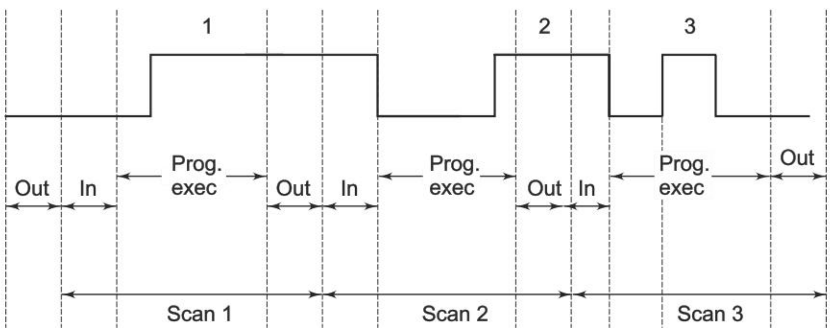


Fig. 21.24 Scan diagram

Input 3 is never seen, this is because when scan 3 was looking at the inputs, signal 3 was not ON yet. It turns OFF before scan 4 looks at the inputs. Therefore, signal 3 is never seen by the PLC.

To avoid this, the input should be ON for at least 1 input delay + one scan time as shown in Fig. 21.25.

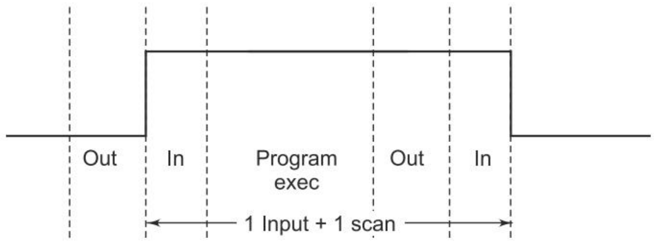


Fig. 21.25 Scan + delay diagram

If the input was not long enough, then the PLC does not see the input turn ON, which is not desirable, hence, there are two ways to overcome this disadvantage. One is using Pulse Stretch function shown in Fig. 21.26. This function extends the length of the input signal until the PLC looks at the input during the next scan, i.e. it stretches the duration of the pulse. The other method is the Interrupt function as shown in Fig. 21.27. This function interrupts the scan to process a special routine that has been written, that is, as soon as the input turns ON, regardless of where the scan currently is; the PLC immediately stops what it is doing and executes an interrupt routine. After it has done executing the interrupt routine, it goes back to the point it stopped or was interrupted or left OFF at and continues on with the normal scan process.

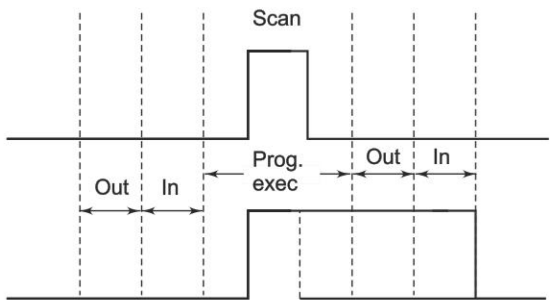


Fig. 21.26 Pulse stretch function

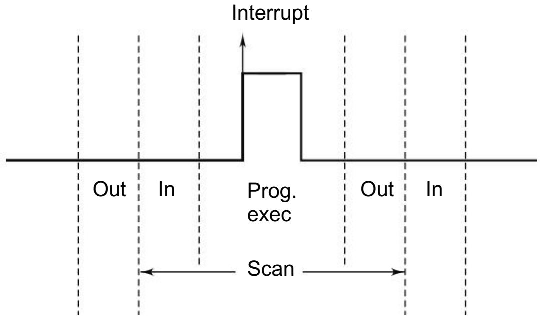


Fig. 21.27 Interrupt function

21.15.5 Relays

The main purpose of a PLC is to replace real world relays.

A relay is basically an electro-magnetic switch. When a voltage is applied to the coil, a magnetic field is generated. This magnetic field attracts the contact of the relay in, causing them to make a connection. These contacts act like a switch and allow current to flow between 2 points thereby closing the circuit.

Let us consider the following example in which we will simply turn ON a bell, whenever the switch is closed. A switch, relay and a bell is connected as shown in Fig. 21.28.

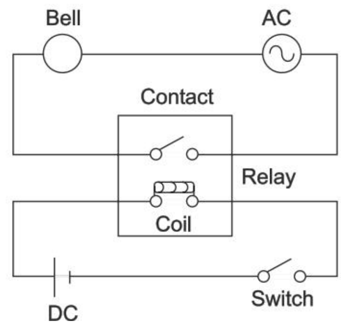


Fig. 21.28 Bell circuit diagram

When the switch closes, current is applied to a bell causing it to sound.

In Fig. 21.28, it is seen that it consists of two separate circuits. One circuit is the dc part and the other circuit is the ac part.

In this case we are using a dc relay to control an ac circuit. When the switch is open, no current flows through the coil of the relay. As soon as the switch is closed, current starts to flow through the coil causing a magnetic field to build up. This magnetic field causes the contacts of the relay to close. Hence, ac current flows through the bell and the sound of the bell can be heard.

Let us now replace the relay by a PLC. The first process is necessary to create what is called a *ladder diagram*. (A ladder diagram consists of vertical lines called the bus bars and within these vertical lines are placed various horizontal lines consisting of input contacts and output. These horizontal lines are called as rungs.) We have to create a ladder diagram, but a PLC does not understand a schematic diagram. It only recognizes code. Fortunately most PLCs have software which convert ladder diagrams into code.

First step We have to translate all of the items we are using into symbols the PLC understands. The PLC does not understand terms like switch, relay, bell etc. It prefers input, output, coil, contact, etc. It does not care what actual input or output device actually is. It only cares that it is an input or an output.

The batteries or power supply is replaced by a symbol. This symbol is common to all ladder diagrams. These are called the bus bars. These look like two vertical lines on either side and the input and output are placed within these bars. The left side can be considered as the voltage and the right side as the ground and the current flow from left to right.

The inputs and outputs each are also given a symbol. The input, that is, the switch will be connected by a symbol, shown in Fig. 21.29. This symbol can also be used as the contact of the relay.

Only one output is normally used, e.g. a bell. The output that the bell will be physically connected in the circuit by the symbol is shown in Fig. 21.30. This symbol is used as the coil of a relay.

Fig. 21.29

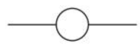


Fig. 21.30

The ac supply is an external supply hence it is not put in the ladder diagram. The PLC only knows and cares about which output it has to turn on.

The PLC must know where everything is located. In other words we have to give all the devices an address. The location where the switch is going to be physically connected to the PLC. Each inputs and outputs used have an address. The PLC has a lot of inputs and outputs but the PLC has to figure out which device is connected where.

The final step is to convert the schematic into a logical sequence of events. The program written tells the PLC what to do when certain events take place. The PLC should be told what to do when the operator turns ON the switch. The diagram shown in Fig. 21.31 is the final converted diagram. In Fig. 21.31, the input is called as '0000' and output is called as '0500'.

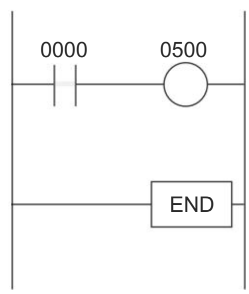


Fig. 21.31 Basic ladder diagram

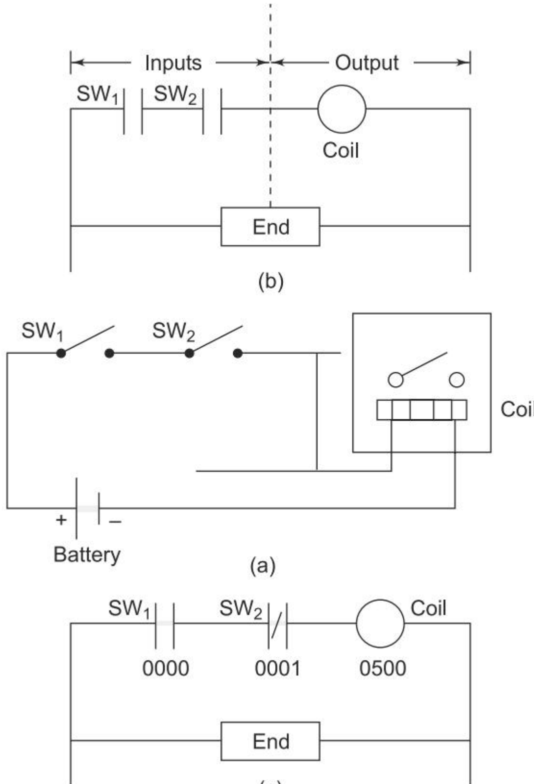
21.15.6 PLC Registers

Let us consider a simple example and compare the ladder diagram with its real world external physically connected relay circuit.

In Fig. 21.32 (a), the coil circuit will be energized when there is a closed loop between the '+' and '–' terminals of the battery. The same circuit can be drawn using ladder diagram. A ladder diagram consists of individual rungs. Each rung must contain one or more inputs and one or more outputs. The first instruction on a rung must always be an input instruction and last instruction on a rung should always be an output coil. The ladder diagram of Fig. 21.32(a) is shown in Fig. 21.32(b).

The register in use can be explained by using Fig. 21.32(b) and changing SW2 from normally open to normally closed as shown in Fig. 21.32(c).

Hence, in Fig. 21.32(c), SW1 will be physically OFF and SW2 will be physically ON initially. Each symbol or instruction has been given an address. This address sets aside a certain storage area in the PLC data files so that the status of the instruction (i.e. true/false) can be stored. Most PLCs uses 16 slots or bit storage locations. In the example given above, two different storage locations or registers are used.



(c)

15	14	13	12	11	10	Register 00									
15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00
														1	0
Register 05															
15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00
															0

(d)

Fig. 21.32

In the tables of two registers 00 and 05 shown in Fig. 21.32(d), we can see that in register 00, bit 00 corresponding to input 0000 was a logic 0 and bit 01 corresponding to input 0001 was a logic 1. Register 05 shows that bit 00 corresponding to output 0500 was a logic 0. The logic 0 or 1 indicate whether an instruction is False or True.

The PLC will only energise an output when all conditions on the rung are TRUE. Hence, in the above example, SW1 must be logic 1 and SW2 must be logic 0, then and only then the output (coil) will be True, that is energized.

If any instruction on the rung before the output (coil) is false, then the output (coil) will be false (not energized).

21.15.7 Timers

Timer is an instruction that waits a set amount of time before doing something. The different kinds of timers are available with different manufacturers. Most of the timers are:

1. *ON-delay Timer:* This type of timer simply "delays turning on". In other words after our sensor (input) turns on, the timer waits for some seconds say x secs before activating a solenoid valve (output). This is the most common timers. It is often called TON (timer on-delay), TIM (Timer) or TMR (Timer).
2. *OFF-delay Timer:* This type of timer is the opposite of the on-delay timer. This timer simply "delays turning OFF". After our sensor (input) sees a target, the solenoid (output) is turned on. When the sensor no longer sees the target, the solenoid is held for some seconds, say x secs before turning it OFF. It is called a TOF (timer OFF-delay) and is less common than the on-delay type listed above.
3. *Retentive or Accumulating Timer:* This type of timer needs 2 inputs. One input starts the timing event (i.e. the clock starts ticking) and the other resets it. The ON/OFF delay timers above would be reset if the input sensor was not ON/OFF for the complete timer duration. This timer however holds or retains the current elapsed time when the sensor turns off in mid-stream. For example, if we wish to know how long a sensor is on during a 1 hr period. If we use one of the above timers they will keep resetting when the sensor turns OFF/ON. This timer however, will give us a total or accumulated time. It is often called an RTO (Retentive timer) or TMRA (accumulating timer).

To use a timer, we typically need to know two things.

1. What will enable the timer? There is typically one of the input (sensor).
2. How long we want to delay before we react. For example, let us say, 5 sec, before a solenoid is turned on.

When instructions before the timer symbol are true, the timer starts "ticking" when the time elapses the timer will automatically close its contacts. When the program is running on the PLC, the program typically displays the elapsed or 'accumulated' time, so that the current value can be seen. Typically timers can tick from 0 to 9999 or 0-65535 times. Most of the PLC used have 16 bit timers. Hence, a 0-9999 is for a 16 bit BCD and 0-65535 for a 16 bit binary. Each tick of the clock is equal to x seconds.

Typically each manufacturer offers several different ticks. Most manufacturers offer 10 ms and 100 ms increment or tick of the clock. Several manufacturers also offer 1 ms as well as 1 second increments. These different increment timers work the same as above but sometimes have different names to show their timebase. Some of them are TMH (high speed timer), TMS (super high speed timer) and TMRAF (accumulating fast timer).

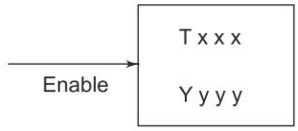


Fig. 21.33 Timer symbol

A typical timer instruction symbol is shown in Fig. 21.33. This timer is the on-delay type and is named Txxx. When the enable input is on the timer starts. When it ticks up-to yyyy (the preset value) times, it will turn on its contacts. The duration of a tick (increment) varies with vendor or user and the time base used (for a tick may be of 1ms or 1 sec or …).

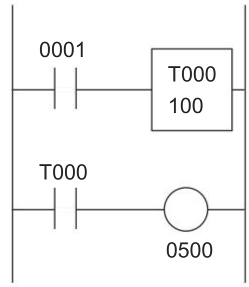


Fig. 21.34 Ladder diagram using timer

In the ladder diagram shown in Fig. 21.34, when the input 0001 is turned on. The timer T000 (a 100 ms increment timer) starts ticking. It will tick 100 times. Each increment is of 100ms, hence the timer will be a 100 tick × 100 ms = 10000 ms (i.e. 10 second) timer. When 10 secs have elapsed the T000 contacts closes and the output 0500 is turned on. When the input 0001 turns off, the timer T000 will reset back to 0 causing its contacts to turn off thereby output 0500 is turned back off.

An accumulating timer, is as shown in Fig. 21.35. Let this timer be named Txxx. When the enable input is ON the timer starts to tick. When it ticks yyyy (the preset value) times, it will turn on its contacts that is used later in the program. The duration of a tick (increment) varies with the user and the timebase used. (A tick can be a 1 ms or 1 sec or ...) If however the enable input turns off before the timer has completed, the current value will be retained. When the input turns back on, the timer will continue from where it left off. The only way to force the timer back to its preset value to start again is to turn on the reset input.

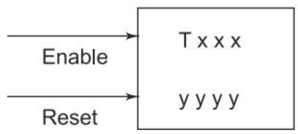


Fig. 21.35 Accumulating timer

A ladder diagram is shown in Fig. 21.36 consisting of the symbol of the timer. When the input 0002 turns on, the Timer T000 (a 10 ms increment timer in this case) starts ticking. It will tick 100 times. Each tick or increment is 10 ms so the time will be a 100 × 10 ms = 1000 ms = 1 second timer. When 1 second is elapsed, the T000 contacts closes, output 0500 turns on. If input 0002 turns back OFF, the current elapsed time will be retained. When the 0002 turns back on, the timer will continue when it left off. When input 0001 turns on (true) the timer T000 will reset back to 0 causing its contacts to turn off (false) thereby making output 0500 turn back off.

When creating a timer that lasts a few seconds, or more, the precision or accuracy is of not much significance. But in the timers that have duration in ms range, precision and accuracy is very significant.

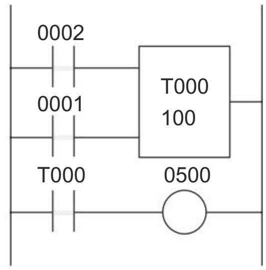


Fig. 21.36 Ladder diagram using accumulating timer

Basically there are two types of errors in general when using a timer. The first is called an *input error* and the other is called the *output error*. The total error is the sum of both the input error and output errors.

Input Errors An input error occurs depending upon when the timer input turns on during the scan cycle. When the input turns on immediately after the PLC looks at the status of the inputs during the scan cycles, the input error will be at its largest (i.e. More than 1 full scan time). This is because as seen in the scan time, the inputs are looked at once during a scan. If it was not on when the PLC looked or scanned and turns on later in the scan, obviously an error would occur. Further, we have to wait until the timer instruction is executed during the program part of the scan. If the timer instructions is the last on the rung then the error could be quite large.

Output Error An output error occurs depending upon, when in the ladder the timer actually 'times out' (ceases) and when the PLC finishes executing the program to get to the part of the scan when it updates the output. This is because the timer finishes during the program execution but the PLC must first finish executing the remainder of the program before it can turn on the appropriate output.

Figure 21.37 illustrates the worst possible input error.

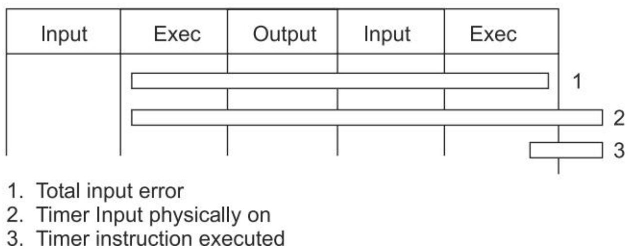


Fig. 21.37 Worst possible input error

1. Total input error
2. Timer input physically on
3. Timer instruction executed

As seen from the Fig. 21.37, the worst possible input error would be 1 complete scan time + 1 program execution time. (A program execution time varies from program to program.)

Figure 21.38 (a) illustrates the worst possible output error. Again the worst possible output error can be 1 complete scan time.

Based on the figure, we can see that the total worst possible timer error would be equal to : 1 scan time + 1 program execution time + 1 scan time = 2 scan times + 1 program execution time.

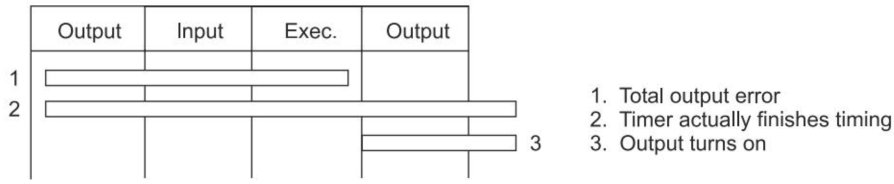


Fig. 21.38 (a) Worst possible output error

Even though most manufacturers currently have timers with 1 ms increments, they really should not be used for duration less than a few ms. This assumes that your scan time is 1 ms. For example if your scan time is 5 ms, then a timer with a duration less than about 15 ms should not be used.

In most applications these errors are insignificant, but in some high speed or precise applications, these errors can be very significant. These errors are software errors. There is also a hardware input error as well as hardware output error.

The hardware input error is caused by the time it takes for the PLC to actually realize that the input is ON when it scan its inputs. Typically this duration is about 10 ms. This is because many PLCs require that an input should be physically on for a few scans before it determines its physically on (to eliminate noise or bouncing inputs).

The hardware output error is caused by the time it takes when the PLC tells its output to physically turn on until the moment it actually does. Typically a transistor takes about 0.5 ms whereas a mechanical relay takes about 10 ms.

21.15.8 Level Application

We have already discussed the working of registers. In order to enhance the understanding of how the program gets scanned, let us process a program like the PLC does with an application.

Let us consider the following application to control lubricating oil being dispensed from a tank. This is possible by two sensors. One near the bottom and one near the top as shown in Fig. 21.38(b).

In this application, we want the fill motor to pump lubricating oil into the tank until high level sensor turns on. At this point, the motor is to be turned OFF. As the level falls below the low level sensor. The fill motor is again turned ON the process is repeated.

Here a three I/O devices are needed—two inputs (sensors) and one output (the fill motor). Both of the inputs will be NC (normally closed) fiber optic level sensor. When they are NOT immersed in liquid they will be ON but when immersed in liquid they will be OFF.

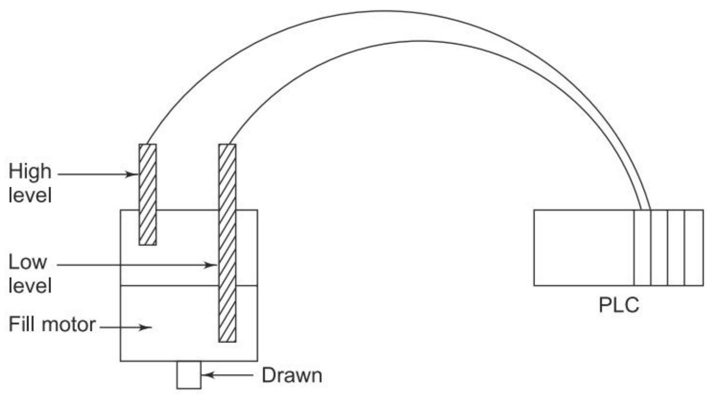


Fig. 21.38 (b) Dispensing oil from a tank

Each input and output device is given an address. This will let the PLC know where they are physically connected. The addresses are shown in Fig. 21.39.

Inputs	Address
Low	0000
High	0001

Outputs	Address
Motor	0500

Internal Utility Relay
1000

Fig. 21.39

The ladder diagram for this application is as shown in Fig. 21.40. Note that in this application an internal utility relay is used. The contacts of these relays can be used as many times as required. In this case, they are used twice to simulate a relay with two sets of contacts. These relays do not physically exist in the PLC but rather they are bits in a register that you can use to simulate a relay.

The most common reason for using a PLC in applications is for replacing real world relays. The internal utility relay make this action possible.

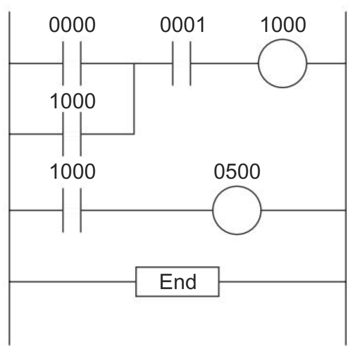


Fig. 21.40 Ladder diagram

The Program Scan Let us analyse in this program scan by scan. Initially the tank is empty. Therefore, input 0000 and 0001 is true shown in Fig. 21.40.

Figure 21.41(a) indicates the condition of the inputs at Scan 1 where the internal utility relay is False. The utility relay becomes active, i.e. True from Scan 2 to 100 as shown in Fig. 21.41 (b). The fill motor is activated and starts filling the tank. Gradually the tank fills because 500 (fill motor) is on.

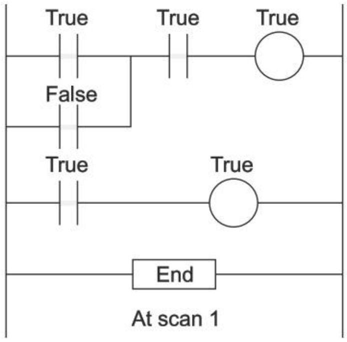


Fig. 21.41 (a)

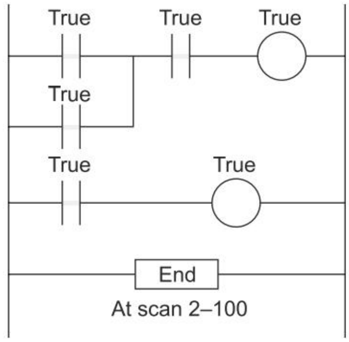


Fig. 21.41 (b)

After 100 scan the oil level rises above the low level sensor and it becomes open (False) shown in Fig. 21.41(c)

It can be seen that even when the low level sensor is false, there is still a path of true logic from left to right shown in Fig 21.41 (c). This is why an internal utility relay is used. Relay 1000 is latching the output (0500) on. It will remain in this condition until there is no true logic path from left to right, i.e. when 0001 becomes false shown in Fig. 21.42 (a).

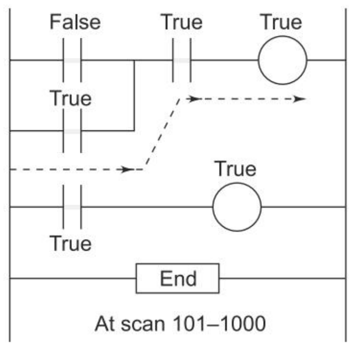


Fig. 21.41 (c)

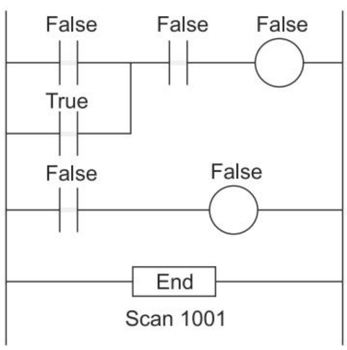


Fig. 21.42 (a)

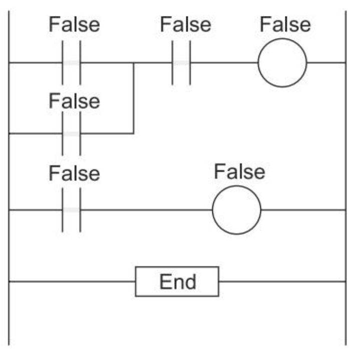


Fig. 21.42 (b)

After 1000 scans, the oil level rises above the high level sensor as it also becomes open i.e. false.

Since there is no more true logic path output 0500 is no longer energized (true) and therefore motor turns off as shown in Fig. 21.42(b)

After 1050 scans the oil level below the high level sensor falls and it will become true again. Even though the high level sensor became True, there still is NO continuous true logic path and therefore coil 1000 remains false.

After 2000 scans the oil level falls below the low level sensor and it will also become true again. At this point the logic will appear the same as scan 1 and the logic will repeat as described earlier.

21.15.9 Counters

A counter is a simple device used for counting. There are different types of counters.

There are up-counters (that only count up 1, 2, 3 …). These are called CTU (Count up), CNT, C or CTR.

There are down counters (that only count downwards, that is, reversed 9, 8, 7). These are typically called CTD (Count down).

There are also up-down counters (they count up and/or down). These are typically called UDC (up-down counter).

Most manufacturers also include a limited number of high speed counters. These are commonly called HSC (High Speed Counter), CTH (Counter High Speed). Typically a high speed counter is a hardware device. The normal counters listed above are software counters. In other words, they do not physically exist in the PLC but rather they are simulated in software. Hardware counters do exist in the PLC and are not dependant on scan time.

As a rule, it is important to use the normal (software) counters unless the pulses you are counting arrive faster than twice the scan time, e.g. If the scan time is 2 ms, and pulses arrive for counting every 4 ms or longer, then use a software counter. If they arrive faster than every 4 ms (3 ms, for example) then use the hardware (high speed) counters. To use the counter the following should be known.

1. The source of the pulse to be counted typically, this is from one of the inputs, i.e. a sensor connected to input 0000.
2. How many pulse to be counted before activation. For example, let us say 5 objects are to be counted.
3. When/how the counter will be reset so that it can count again. For example, let us say after 5 objects let us reset the counter.

When the program is running on the PLC, the program typically displays the current or accumulated value so that we can see the current count value.

Typically counters can count from 0 – 9999, – 32,768 to +32,767 or 0 – 65535, since most PLCs have a 16 bit counters.

0 – 9999 is a 16 bitBCD and –32768 – + 32767 and 0 – 65535 are 16 bit Binary.

Figure. 21.43 shows a counter with 2 inputs. One goes to the reset line and when this input in turned on, the current (accumulated) count value will return to zero. The second input in the address from where the pulses arrive that are used for counting. Let us consider an example. If we are counting the number of objects that pass in front of the sensor that is physically connected to input. 0001, then normally open contacts with the address 0001 would be connected in front of the pulse line.

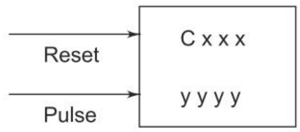


Fig. 21.43 Symbol of counter

In Fig. 21.43 Cxxx is the name of the counter. If the counter is called 000 then C000 would be written.

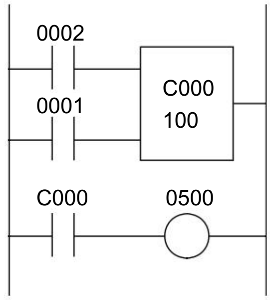


Fig. 21.44 Ladder diagram using counter

YYYY is the number of pulses that is used for counting. If 10 objects are to be counted before turning on a physical output to box these, then 10 would be placed. If 100 is to be counted then 100 would be placed here. When the counter has finished counting YYYY objects, it will turn on a separate set of contacts that are also labelled Cxxx. It is important to note that the counter accumulated value only changes at the off to on transition of the input pulse. Figure. 21.44 shows, the symbol on a ladder and how the counter can be set up. Let the counter be called C000 to count 100 objects from 0001 input before turning on output 0500. Sensor 0002 resets the counter.

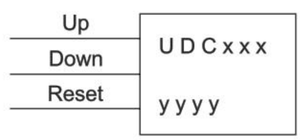


Fig. 21.45 Symbol of up-down counter

Different types of counters are used for up counting, down counting or up-down counting.

The symbol for an up-down counter (UDC) is shown Fig. 21.45. Let us name it UDCxxx and yyyy as before.

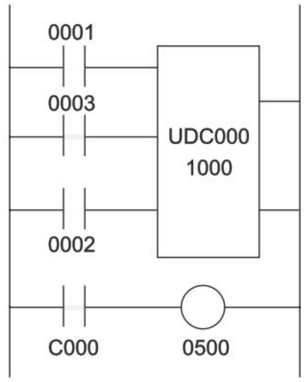


Fig. 21.46 Ladder diagram using up-down counter

A UDC needs three inputs. The reset input has the same function as described earlier. However, now instead of having only one input for pulse counting, two inputs are used for counting. One pulse input is for counting up and the other is for counting down. Let us consider an example in which the counter is named UDC000 and give a preset value of 1000, i.e. it will count 1000 total pulses. For inputs, sensor are used, one of which turns, input 0001 on when it sees a target and another sensor at 0003 also turns on when it sees the target. When input 0001 turns on, the counter counts up and when input 0003 turns on, the counter counts down. When 1000 pulses is reached the output 0500 will be turned ON. The ladder is as shown in Fig. 21.46.

In most PLCs, it is important to note that counters and timers cannot have the same name.

21.15.10 DC Inputs

The, dc input modules commonly available will work with 5, 12, 24 and 48 volts.

The dc input modules allow us to connect either PNP (sourcing) or NPN (sinking) transistor type devices to them. If using a regular switch (i.e. a toggle or Push Button, etc.) then it is not important whether it is wired as NPN or as PNP. It is to be noted that most PLCs do not mix NPN and PNP devices on the same module. But when using a sensor e.g. (photo-sensor, proximity, etc.) then it is important to know its output configuration. Hence it always advisable to verify before connection whether its PNP or NPN. The difference between the two types is whether the load (in our case PLC is the load) is switched to ground or positive voltage. An NPN has the load switched to ground whereas a PNP device has the load switched to positive voltage. Figure 21.47 shows an NPN sensor (Sinking mode). In this NPN sensor one output is connected to the PLCs input and the other to the power supply ground.

If the sensor is not powered from the same power supply as the PLC, then both the grounds should be connected.

For a PNP sensor, the load is switched to ground. On the PNP sensor, one output is connected to positive voltage and the other to the PLC's input as shown in Fig. 21.48.

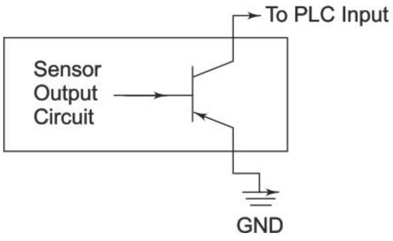


Fig. 21.47 NPN sensor

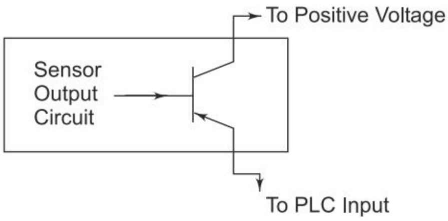


Fig. 21.48 PNP sensor

If the sensor is not powered from the same supply as the PLC, then both positive voltages (V+) should be connected together. Inside the sensor the transistor acts as a switch. The sensors internal circuit tells the output transistor to turn on when a target is present. The transistor then closes the circuit between the two connections shown above, i.e. The V + and the PLC input.

The only thing that is accessible to the user are the terminals labelled common, input 0000, input 0001, input xxxx as shown in Fig. 21.49.

The common terminal either gets connected to V + or ground. The type of sensor used decides where the common terminal is to be connected. When using a NPN sensor this terminal is connected to V +. When using a PNP sensor this terminal is connected to zero voltage (ground).

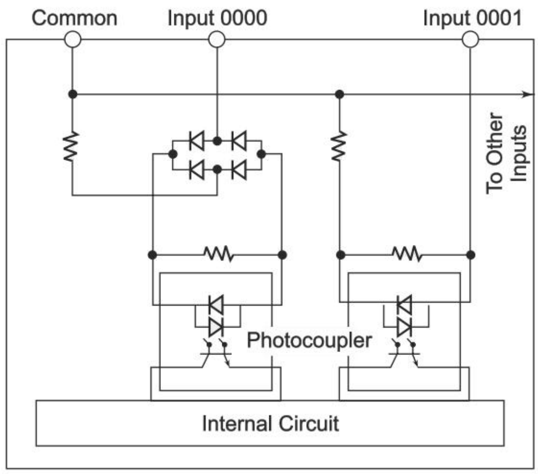


Fig. 21.49 Internal circuit using photocouplers

A common switch (i.e. limit switch, push button, toggle, etc.) would be connected to inputs in a similar fashion. One side of the switch would be connected directly to V + and the other end goes to the PLC input terminal. This assumes that the common terminal is connected to 0V (ground). If the common terminal is connected to V+ then simply connect one end of the switch to 0V (ground) and the other end to the PLC input terminal.

The photocouplers shown in Fig. 21.49 are used to isolate the PLC's internal circuit from the inputs. This eliminates the chance of any electrical noise entering the internal circuitry. They operate by connecting the electrical input signal to light and then by converting the light back to an electrical signal to be processed by the internal circuit.

21.15.11 AC Inputs

The, ac input modules that are commonly available work with 24, 48, 110 and 220 V. It is important to use the one that fits the needs based upon the inputs devices (voltage) used.

These day, ac input modules are less common than dc input modules. This is because most of the sensors used today have transistor outputs. A transistor does not operate with ac voltages. Most commonly, the ac voltage is being switched through a limit switch or other switch type. If a sensor is used it probably is operating on a dc voltage. An ac device is connected to the input module as shown in Fig. 21.50, while the dc hot wire is connected to the switch while the neutral goes to the PLC. The ac ground (third wire wherever applicable) should be connected to the frame ground terminal of the PLC. The ac connections are color coded so that the individual wiring the device knows which wire is which.

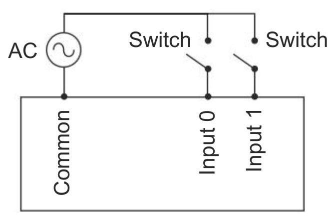


Fig. 21.50 ac input module

The PLC's ac input module is as shown in Fig. 21.51. The only thing accessible to the user are the terminals labeled common, input 0000, input xxxx. The common terminal gets connected to the neutral wire.

A common switch such as limit switch, push button, toggle, etc. would be connected to the inputs directly. One side of the switch would be connected directly to input xxxx. The other end goes to the ac hot wire, assuming that the common terminal is connected to neutral.

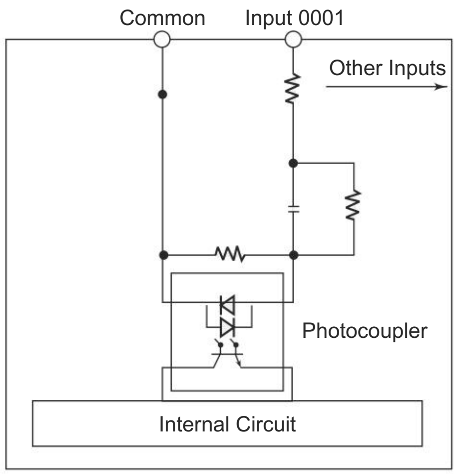


Fig. 21.51 PLC's ac input module (Internal diagram)

Typically ac input takes longer than a dc input for a PLC to scan. It does not matter from the programmers point of view because an ac input device is typically a mechanical switch and mechanical devices are very slow. It is quite common for a PLC to require that the input be on for 25 or more milliseconds before it is seen. This delay is required because of the filtering which is needed by the PLC internal circuit.

21.15.12 Master Controls

Master controls can be considered the emergency stop switches. An emergency stop switch typically is a big red button on a machine that will shut it off in cases of emergency. It is commonly abbreviated as MC/MCR (Master control / Master Control Reset), MCS/MCR (Master Control Set/ Master Control reset) or just simply MCR (Master control reset)

The master control instruction is used in pairs with a master control reset. Some use MCR in pairs instead of putting together with another symbol. Many manufacturers use them differently. Consider the ladder diagram using MCR shown in Fig. 21.52. Let us see how this program will run using differentPLCs.

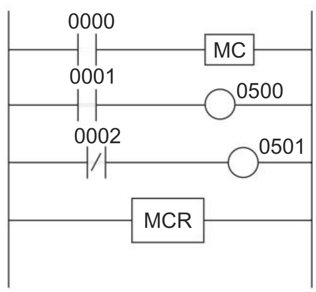


Fig. 21.52 Ladder diagram using master control (MC/MCR)

Manufacturer A In this example, rungs 2 and 3 are only executed when input 0000 is on (true). If input 0000 is not true the PLC pretends that the logic between the MC and MCR instructions does not exist, it would therefore by pass this block of instructions and immediately go to the next rung after the MCR instructions. Conversely, if the input 0000 is true, the PLC would execute rungs 2 and 3 and updates the status of outputs 0500 and 0501 accordingly, so if input 0000 is true, program execution goes to rung 2. If input 0001 is true, 0500 will be true and hence it is on, when the PLC updates the outputs. If input 0002 is true (i.e. physically off) 0501 will be true and therefore it will turn on when the PLC updates the outputs. MCR just tells the PLC "that's the end of MC/MCR block".

In this PLC, scan time is not extended when the MC/MCR block is not executed because the PLC pretends the logic in the block does not exist. In other words, the instructions inside the block is not seen by the PLC and therefore it doesn't execute them.

Manufacture B In this example, rungs 2 and 3 are always executed immaterial of the status of the input 0000. If the input 0000 is not true, the PLC executes the MC instruction (i.e. MC becomes true) it then forces all the input instruction inside the block to be OFF. If input 0000 is true the MC instruction is made to be false.

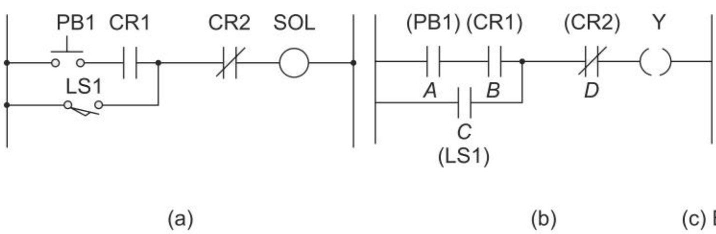
Then, if input 0000 is true, program execution goes to rung 2, if input 0001 is true, 0500 will be true and hence it will turn on when the PLC updates the output. If input 0002 is true (i.e. physically off), 0501 will be true and therefore it will turn on when the PLC updates the outputs. MCR just tells the PLC "that's the end of the MC/MCR block''. When input 0000 is false, inputs 0001 and 0002 are forced off regardless whether they are physically ON or OFF. Therefore outputs 0500 and 0501 will be false.

The difference between manufactuers A and B is that in B scheme the scan time will nearly be the same, whether the block is ON or OFF. This is because the PLC sees each instruction, whether the block is ON or OFF.

21.15.13 PLC Programming Languages

The term PLC programming language refers to the method by which the user communicates information to the PLC. The two most common language structures are ladder diagram language and Boolean language.

Ladder diagram language is by far the most commonly used PLC language. Figure. 21,53, shows a comparsion of the ladder diagram language and the Boolean language programming.



Start	PB1
And	CR1
Or	LS1
And	
Not	CR2
Out	SOL

(c) Boolean equation: $Y = [(AB) + C]\,\overline{D}$

Fig 21.53 PLC ladder and boolean language

Figure 21.53(a) shows the original relay ladder diagram drawn as it were to be hard-wired. Figure 21.53 (b) shows the equivalent logic ladder diagram that is programmed into the controller. The addressing format shown for input and output devices is generic in nature and varies for different PLC models. Figure 21.53(c) shows a typical set of generic Boolean statements that could also be used to program the original circuit. This statement refers to the basic AND, OR and NOT logic gate functions.

Relay Type Instructions

The ladder diagram language is basically a symbolic set of instructions used to create the controller program. These ladder instruction symbols are arranged to obtain the desired control logic that is to be entered into the memory of the PLC. Because the instruction set is composed of contact symbols ladder diagram is also referred to as *contact symbology*.

The main function of the logic ladder diagram program is to control outputs based on input conditions. This control is obtained through the use of what is referred as a ladder rung. In general, rung consists of a set of input conditions, represented by contact, instructions and an output instruction at the end of the rung represents by the coil symbol as shown in Fig. 21.54.

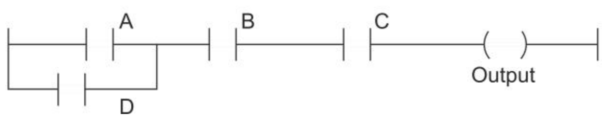


Fig 21.54 Ladder rung

Each contact or coil symbol is referenced with an address number that identifies what is being evaluated and what is being controlled. The same contact instruction can be used throughout the program whenever the conditions need to be evaluated. For an output to be activated or energized at least one left to right path of contacts most be closed. A complete closed path is referred to as having logic continuity. When logic continuity exists in at least one path, the rung condition is said to be true. The rung condition is false, if no path has continuity.

Instruction Addressing

To complete the entry of relay type instruction, an address number must be assigned to it. The number will indicate which PLC input is connected to which input device and which PLC output will drive which output device.

The addressing of real inputs and outputs, as well as internals, depends upon the PLC model used. These addresses can be represented in decimal, octal or hexadecimal number depending upon the number system used by the PLC.

Figure 21.55 shows a typical addressing format. Again, the programming manual of the PLC being used should be consulted to determine the specific format used, as this can vary from model to model, as well as from manufacturer to manufacturer.

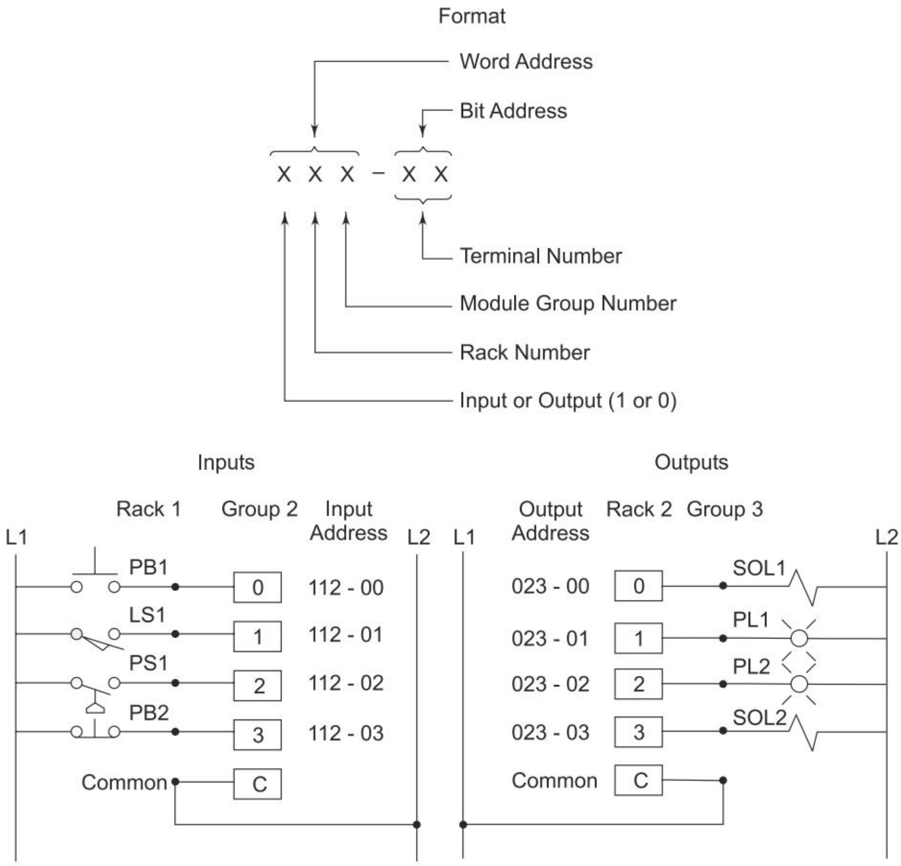


Fig. 21.55 Addressing formats

The address identifies the function of an instruction and links it to a particular status bit in the data table portion of the memory. Figure 21.56 shows the structure of a 16 bit word and its assigned bit values. The breakdown of the word and its addressing are in the decimal numbering system.

Bit No.	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
On/Off Status of the Bits	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1

Fig. 21.56

21.15.14 Basic Process PLC

A PLC is basically made up of the following sections, each of which has a unique function to perform in its operation. The sections are:

1. Sensing inputs or controlling hardware
2. PLC input hardware

3. Controller or CPU
4. Hand held programming device or personal computer
5. Output PLC hardware
6. Hardware output devices

1. Sensing Input The sensing section consists of limit switches, pressure switches, photoelectric sensors, push buttons, etc. These incoming hardware devices provide input signals. Devices such as the push button, limit switch or photoelectric sensors are field input devices. The term *field input* refers to hardware items which provide the incoming signals that you physically connect to the PLC.

2. Input Section The input section of the PLC contains two major areas. First, the physical contact (screw) terminals where the incoming signals (input) from field input devices such as a pressure switches are attached to the PLC.

Figure 21.57 illustrates a product sitting on a conveyor. When the conveyor moves the product into position the sensor will send an input signal into input screw terminal number 0 on the PLC input section.

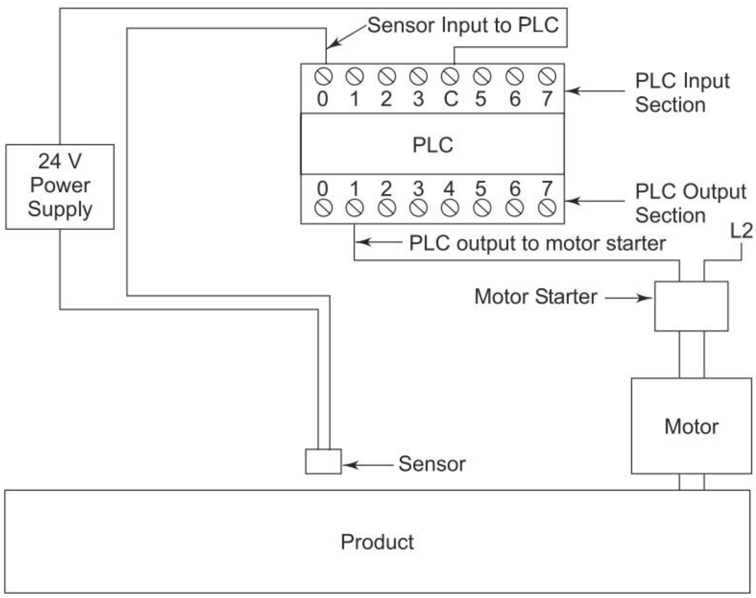


Fig. 21.57 Product on conveyor sensed by 24 v input sensor. PLC will stop the conveyor motor so that the product can be assembled.

The second part of the input section is the PLC's internal conversion electronics. The function of the input section electronic component is to convert and isolate the high voltage input level from field devices.

High voltage signals from the field devices are converted to +5 V direct current

V DC for a valid ON input signal and OV DC for a valid OFF input signal.

Incoming signal conversion and isolation are necessary as the solid state microprocessor components operate on +5V dc, where as input signal may be 24 V dc, 110V ac or 220V dc.

3. Controller The controller, commonly called CPU (Central Processing Unit) or simply the processor. The processor controls or supervises the entire process. The CPU solves the user program and updates the status of the outputs.

4. Programmer The programmer is the device whereby the programmer or operator can enter or edit program instruction or data. The programmer can be a handheld unit, a PC or an industrial computer programming terminal.

5. Output Section The result of the process whereby the CPU will look at, or read the status of inputs and then use this information to solve user program instructions is a series of outgoing signals from the CPU to an outside device such as a motor starter. These outgoing ON or OFF signals are known as outputs. Output signals are electronically transferred from the CPU to the output section electronics controlling the physical screw terminals on the output section of the PLC. (The PLC program for the application of Fig. 21.57 will be written to instruct the PLC to turn OFF the conveyor motor when a part is in position. The PLC will turn OFF the motor by turning OFF the output signal for the output Number1 in the PLC output section.

6. Field Hardware Devices The PLC will stop the conveyor motor when the product is sensed. The conveyor motor starter's coil, which is controlled by the PLC, will be connected to the proper screw terminals on the output modules as shown in Fig. 21.58.

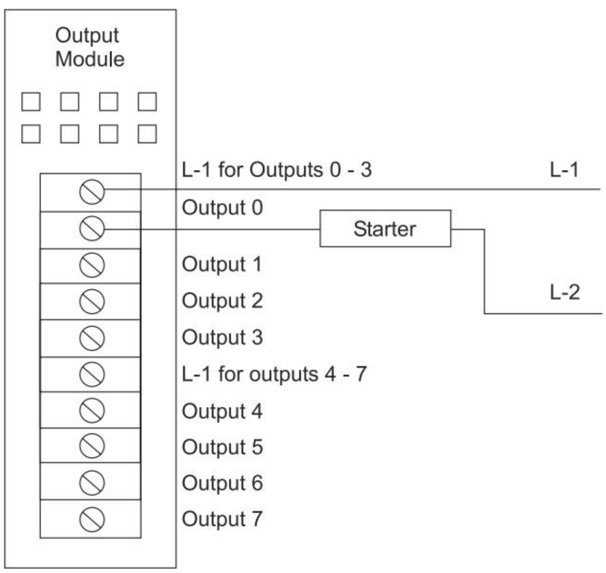


Fig. 21.58 Output module wiring to motor starter coil

21.15.15 PLC Hardware

PLC hardware falls into the following physical configuration:1. Fixed Input/ Output (I/O) and 2. Modular I/O.

1. Fixed I/O PLCs A fixed PLC consists of a fixed, or built-in, input/output section. There is one fixed or built-in, non removable screw terminal strip containing all input signal screw terminal connections and another terminal strip containing all output control signal screw terminals.

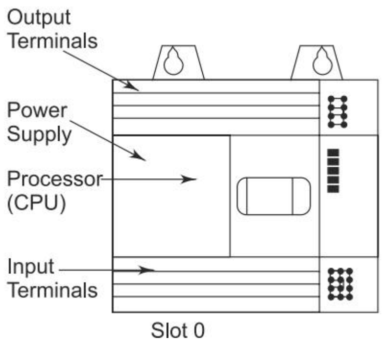


Fig. 21.59 Allen-Bradley SLC 500 modular PLC with the major parts labeled. (Courtesy Rockwell Automation/Allen-Bradley)

Figure 21.59 shows an Allen Bradley fixed SLC 500 PLC. The plastic door labeled output terminals is hinged at the bottom. On opening the door, all of built-in screw terminals can be seen. The area labeled Input terminals is also a plastic door hinged at the top. On opening the door all the built-in input screw terminals can be seen.

2. Modular I/O PLC A modular PLC does not have a terminal strip built into the processor unit. Modular PLC have their I/O points on plug-in type, removable units called I/O Modules. PLC with modular inputs and outputs consist of a chasis, rack or base plate where the power supply, CPU and all input and output modules are present as separate hardware item. The diagram shown in Fig. 21.60 is assembled to make a working PLC.

When I/O is modular, the user can mix input and output types in the rack or base plate to meet the specific needs. There are usually few limitations on the mix or positioning of I/O modules.

Common racks, chasis or base plates hold 4, 5, 7, 8, 10, 13 or 16 I/O modules. Typical modules will contain 4, 8, 12, 16 or 32 I/O points. Electrical connections between each modules and CPU are made by two mating plugs.

One plug is located on a printed circuit board at the back of each module. A PCB that the modules plug into runs along the back of the rack or is built into a base plate, called the *backplane*. The backplane conveys signals from the CPU to each module and from module to the CPU.

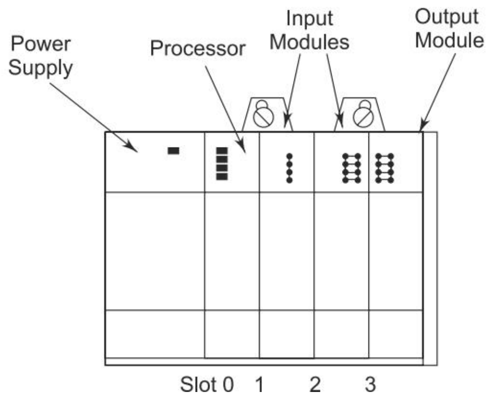


Fig. 21.60 Allen-Bradley SLC 500 modular PLC. (Courtesy Rockwell Automation/Allen-Bradley)

A *DIN rail* is a metal track or rail attached to the back of an electrical panel, where devices can be easily clipped to or removed from the rail. The advantage of the DIN rail is that it permits easy snap-on and removal of hardware devices either at installation or for maintenance or replacement.

The power supply, which is also modular, is hooked up to line voltage. The function of the PLC power supply is to convert line voltage to low voltage dc and then isolate it for use to operate the CPU and any associated I/O module electronics in the rack or chasis. In most cases power must be supplied to all field input and output devices from a source other than the PLC's power supply.

The CPU works like a human brain. It is a solid state microprocessor integrated chip, which is placed on PCB with other supporting and interface chips to build the PLC's CPU or processor module. The CPU basically consists of two components. The controller and memory system.

The controller is the microprocessor or The brain that supervises all operations in the systems. The CPU reads or gathers information from external sources such as the input devices and stores this information in memory for later use by the CPU. When solving the user program is done, the CPU will write or send, data out to the external devices such as output modules and field hardware devices.

The memory system used has to provide the following:

1. **Storage for the user program.**
2. **Storage of input status file data.** The input status file consists of memory locations, that stores the ON-OFF status of each field input device.
3. **Storage of output status file data.** The output status file consists of memory locations that store ON or OFF status of field hardware devices as the result of solving the user program. Data in the output status file is waiting to be transferred to the output module switching device. This device for each output point will turn power ON or OFF to each field output device.

4. **Data Storage:** The data storage area of the memory is used to store numerical data that may be used in mathematical calculations, bar code data being input to the PLC's similar function.

21.15.16 PLC Configuration

A PLC is a solid state device designed to perform logic functions previously accomplished by electro-magnetic relays.

Basically, the PLC is an assembly of solid state digital logic decisions which provides outputs. PLCs are used for the control and operation of manufacturing, process equipment and machinery.

A typical PLC can be divided into three sections as illustrated in Fig. 21.61.

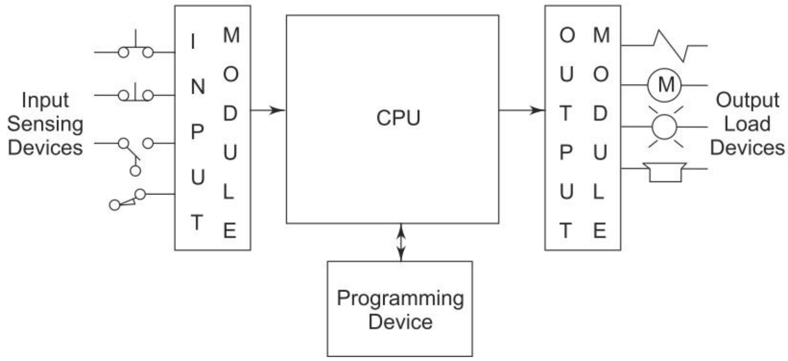


Fig. 21.61 Basic PLC components

These sections are the input/output (I/O) section, Central Processing Unit (CPU) and the programming device.

The CPU is the important section of a PLC and is the brain of the system. It contains internally a number of logic gates. The CPU shown in Fig. 21.62 is a microprocessor based system that replaces control relays, counters, timers and sequencers. It is designed so that the user can enter the desired circuits in relay ladder logic. The CPU accepts (reads) input data from various sensing devices, executes the stored user program from memory and sends the appropriate output commands to control devices.

A dc Power source is required to produce dc low level voltage used by the processor and the I/O modules. This power supply can be housed in the CPU unit or may be a separately mounted unit depending on the PLC manufacturer.

The I/O section consists of input modules and output modular. The I/O forms the interface by which field devices are connected to the controller. The purpose of this interface is to condition the various signals received from or sent to external field devices. Input devices such as Push Button, limit switches, sensors, selector switches and thumb wheel switches are hard wired to the terminals on the input modules.

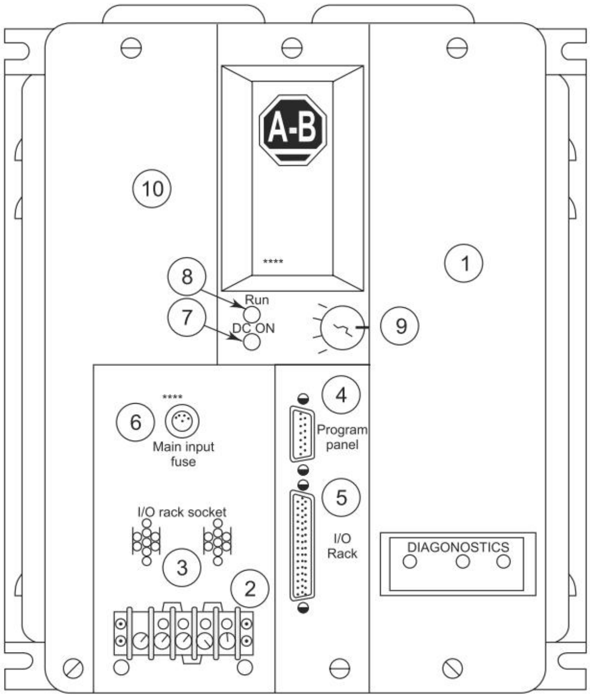


Fig. 21.62 Typical process unit

Output devices such as small motors, motor starters, solenoid valves and indicator lamps are hard wired to the terminals on the output modules. These devices are referred to as field or real world inputs and outputs.

The terms field or real world are used to distinguish actual external devices that exists and must be physically wired from the internal user program that duplicates the function of relays, timers and counters.

The programming device, or the terminals are used to enter the desired program into the memory of the processor. This program is entered using relay ladder logic. The program determines the sequence of operation and ultimate control of the equipment or machinery. The programming device must be connected to the controller only when entering or monitoring the program. By designing the controller to be "electrician friendly''. The PLC can be programmed by people without extensive computer programming experience. The programming unit may be a hand held unit, a LED display or a desktop unit with a CRT display.

21.15.17 Principles of Operation

To understand the operation of a PLC, consider the simple process shown in Fig. 21.63. In this process a mixer motor is to be used to automatically stir the liquid in a vat when the temperature and pressure reach a preset values. In addition, direct manual operation is also provided by means of a separate push button. The process is monitored with temperature and pressure sensor switches that close their respective contacts when conditions reach their preset values.

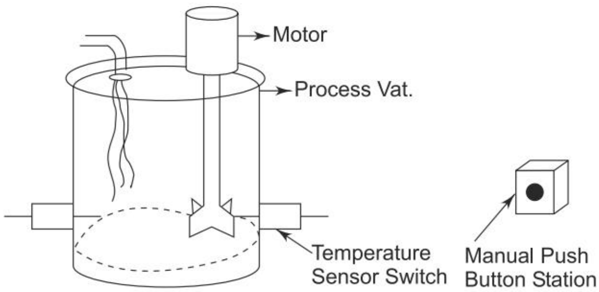


Fig. 21.63 Mixer process control problem

This control problem can be solved using the relay method for motor control shown in the relay diagram Fig. 21.64 (a).

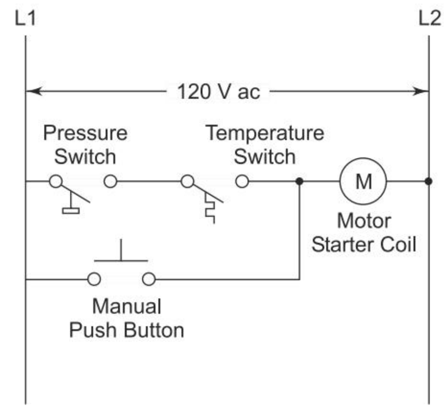


Fig. 21.64 (a) Process control relay ladder diagram

The motor starter coil (M) is energized when both the pressure and temperature switches are closed or when the manual push button is pressed.

To use a PLC for this process the same input field devices such as pressure switch, temperature switch and push button are used. These devices are to be hard wired to an appropriate input module according to the manufacturers labeling scheme. Typical wiring conn-ections for a 120V AC input module are shown in Fig. 21.64(b).

The same output field device (motor starter coil) would also be used. This device would be hard wired to an appropriate output module according to the manufacturers labeling scheme. Typical wiring connections for a 120V ac output module are shown in Fig. 21.64(c).

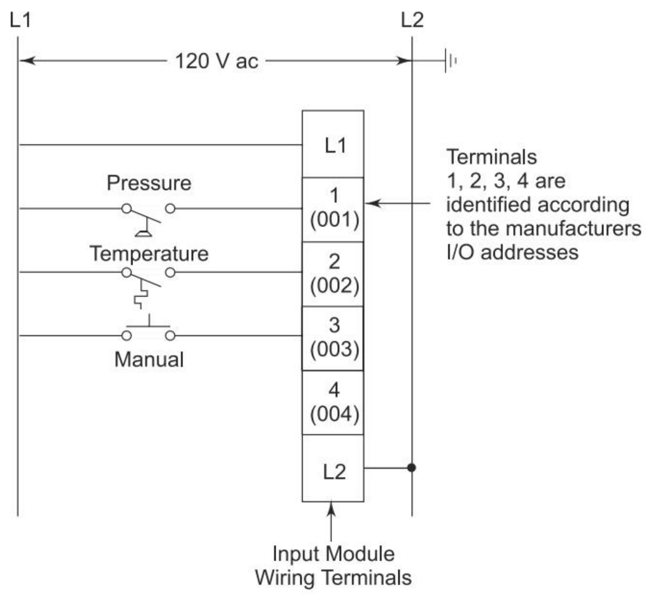


Fig. 21.64 (b) Typical input module wiring connection

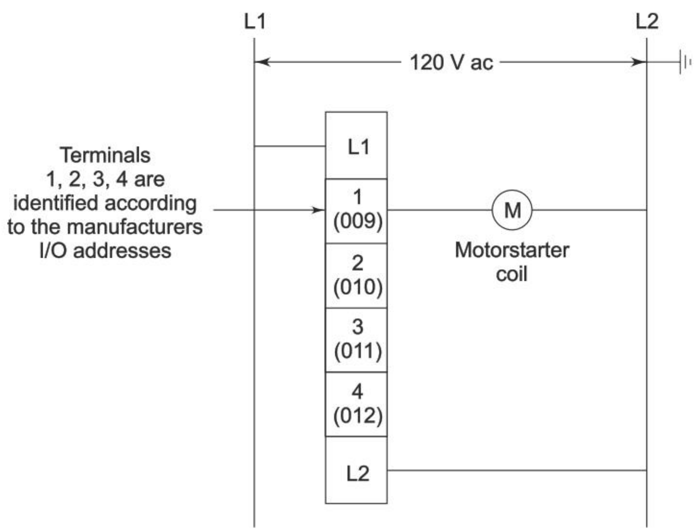


Fig. 21.64 (c) Typical output module wiring connection

Next, the PLC ladder logic diagram would be constructed and programmed into the memory of the CPU. A typical ladder logic diagram for this process is as shown in Fig. 21.65.

The format used is similar to the layout of the hardwired relay ladder circuit. The individual symbols represent instructions while the number represent the instruction addresses. When programming the controller these instructions are entered one by one into the processor memory from the operator terminal keyboard. Instructions are stored in the user program portion of the processor memory.

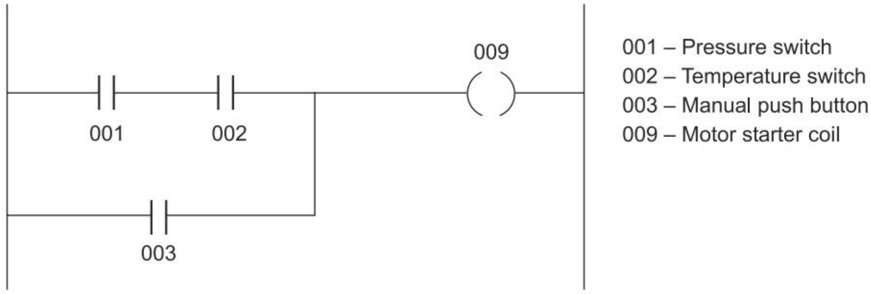


Fig. 21.65 Process control PLC ladder logic diagram

To operate the program, the PLC is placed in the RUN mode, or operating cycle. During each operating cycle, the controller examines the status of input devices, executes the user program, and changes output accordingly.

Each –| |– can be thought of as a set of normally open (NO) contacts. The –( )– can be considered to represent a coil that, when energized will close a set of contacts.

In the ladder diagram shown in Fig 21.65, the coil 009 is energized when contacts 001 and 002 are closed or when 003 is closed. Either of these conditions provides a continuous path from left to right across the rung that include the coil.

The RUN operation can be described by the following sequences of events. First the inputs are examined and their status is recorded in the controllers memory (a closed contact is recorded as a signal of logic 1 and an open contact by a signal of logic 0). The ladder diagram is then executed, with each internal contact given OPEN or CLOSED status according to the record of these contacts, provide a current path from left to right in the diagram, the output coil memory location is given a logic 1 value and the output module interface contacts will close. If there is no conducting path in the program rung, the output coil memory location is set to logic '0' and the output module interface contact will be open.

The completion of one cycle of this sequence by the controller is a scan. The time required for one full scan provides a measure of the speed of response of the PLC and is called the scan time.

21.15.18 Basis of PLC Programming

1. Processor Memory Organisation

The term processor memory organization refers to how certain areas of memory in a given PLC are used. Different PLC manufacturers organize their memories in different ways. Even though they do not use the same memory make up and terminology, the principles involved are the same.

Figure 21.66 shows an illustration of memory organization known as a *memory map.* Every PLC has a memory map. It need not be the same as the one illustrated. The memory space can be divided into two broad categories the user program and the data table.

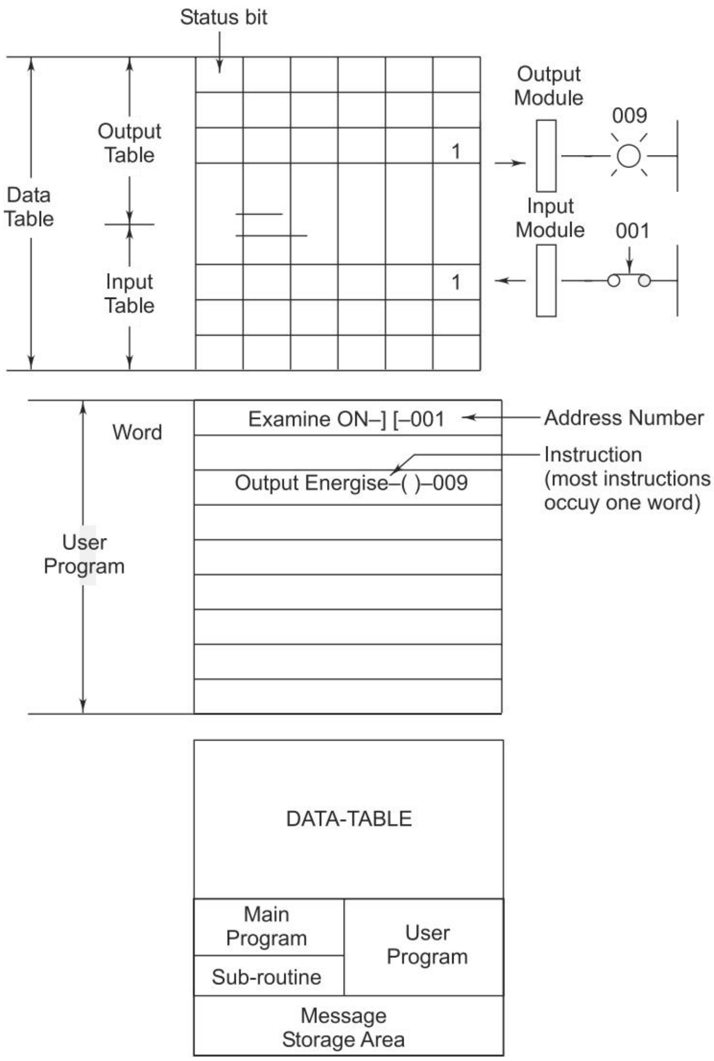


Fig. 21.66 Memory map of PLC

The user program is where the programmed logic ladder diagram is entered and stored. The user program will account for most of the total memory of a given PLC system. It contains the logic that controls the machine operation. The logic consists of instructions that are programmed in a ladder format. Most instructions require one word of memory.

The data table stores the information needed to carry out the user program. This include, such information as the states of input and output devices, timer and counter values, data storage and so on.

Contents of the data table can be divided into two categories—status data and numbers or codes. Status is ON/OFF type of information represented by 0's and 1's stored in unique bit locations. Number or code information is represented by group of bits which are stored in unique bytes or word locations.

The data table can be divided into the following three sections according to the type of information to be remembered—The input image table, the output image table, and timer and counter storage.

The input image table stores the status of the digital inputs which are connected to input interface circuits.

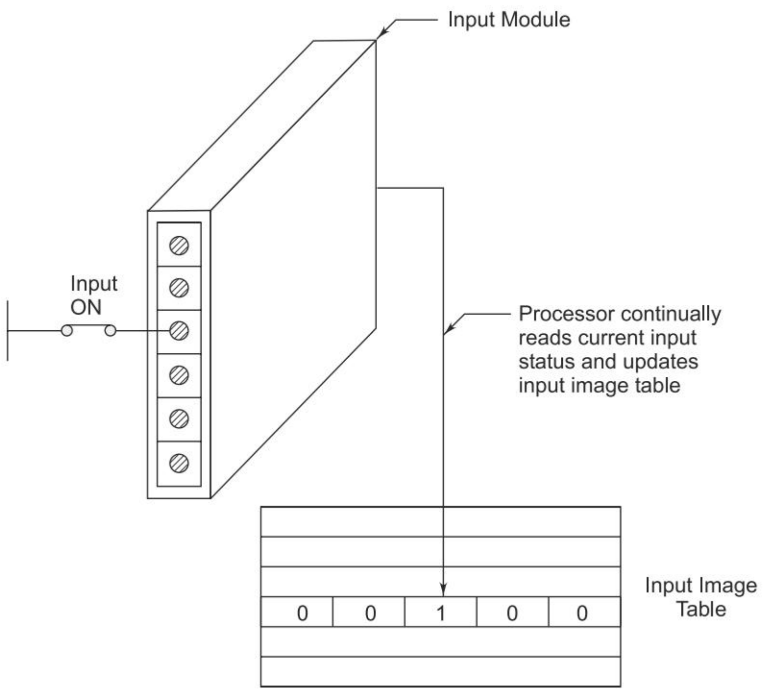


Fig. 21.67 (a) Typical input image table connection

Figure 21.67(a) shows a typical connection of a switch to the input image table through the input module. When the switch is closed the processordetects a voltage at the input terminal and records that information by storing abinary 1 in the proper bit location. Each connected input has a bit in the input image table that corresponds exactly to the terminal to which the input is connected. The input image table is constantly being changed to reflect the current status of the switch. If the input is ON (switch closed), its corresponding bit in the table is set to 1. If the input is OFF (switch open) the corresponding bit is cleared or reset to zero.

The output image table is an array of bits that controls the status of digital output devices which are connected to output interface circuits. Figure 21.67(b) shows a typical connection of a light to the output image table through the output module.

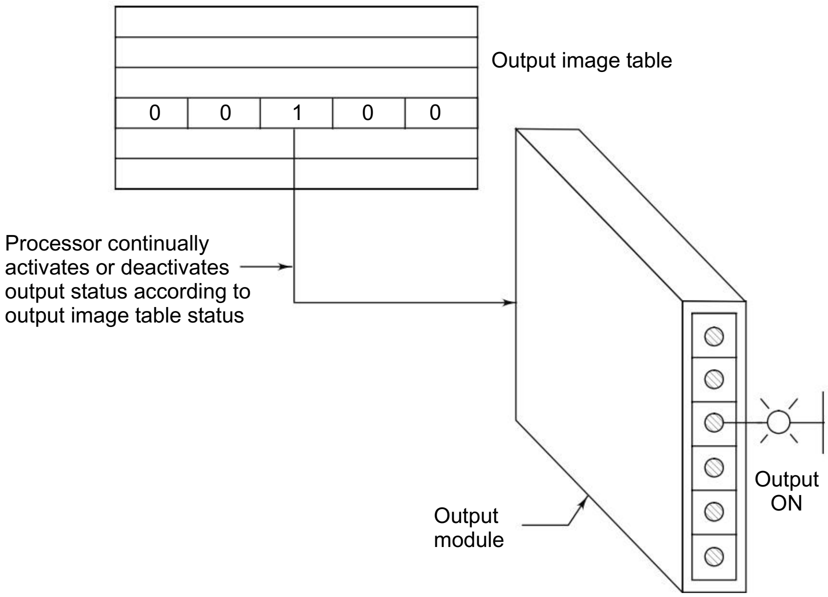


Fig. 21.67 (b) Typical output image table connection

The status of this light (ON-OFF) is controlled by the user program and indicated by the presence of 1's (ON) and 0's (OFF). Each connected output has a bit in the output image table that corresponds exactly to the terminal to which output is connected. If the program calls for a specific output to be ON, it's corresponding bit in the table is set to 1, if the program calls for the output to be OFF, its corresponding bit in the table is set to 0.

2. Program Scan

During each operating cycle, the processor reads all the inputs takes their values and according to the user program energises or de-energises the outputs. This process is known as a scan. Figure. 21.68 illustrates a single PLC scan, which consists of the I/O scan and the program scan.

The PLC scan time specification indicates how fast the controller can react to changes in inputs. Scan time varies with program content and length. The time required to make a single scan can vary from 1 ms to 100 ms. If a controller has to react to an input signal that changes states twice during the scan time it is possible that the PLC will never be able to detect this change.

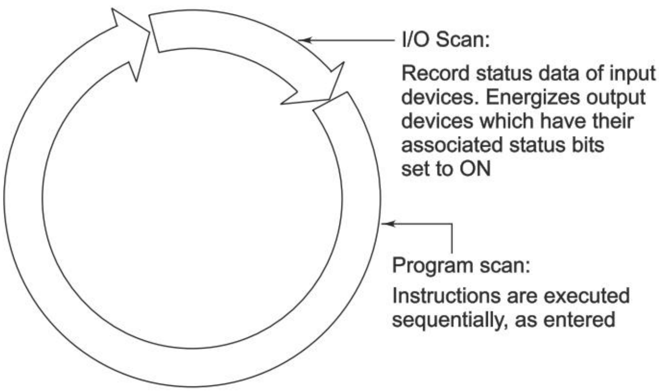


Fig. 21.68 Single PLC scan

The scan is normally a continuous and sequential process of reading the status of inputs, evaluating the control logic and updating the outputs. Figure. 21.69 illustrates this process. When the input device connected to address 101-14 is closed, the input module circuit senses a voltage and a 1 (ON) condition is entered into the input image table bit 101–14.

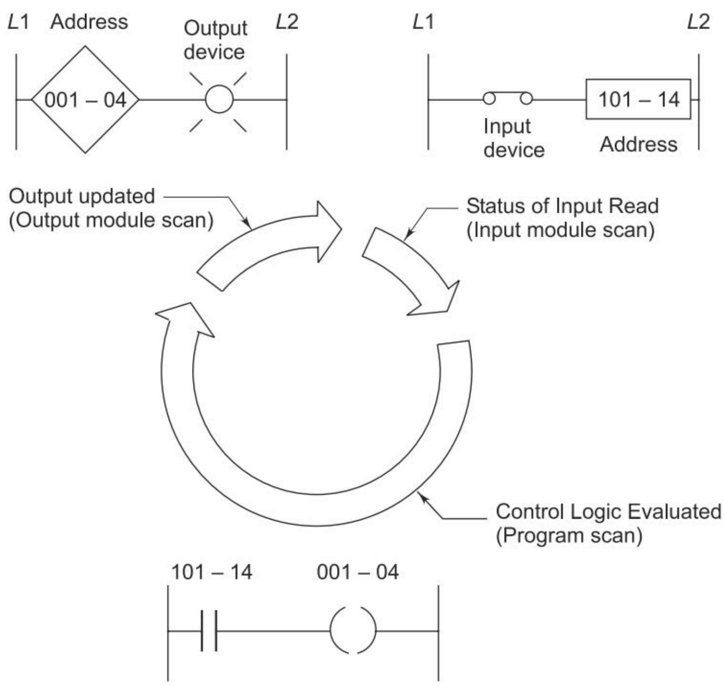


Fig. 21.69 Typical scan sequence

During the program scan, the processor examines bit 101–14 for a 1 (ON) condition. In this case, since input 101–14 is 1, the rung is said to be TRUE. The processor then sets the output image table 001–04 to 1. The processor turns on output 101–14 during the next I/O scan, and the output device (light) wired to this terminal becomes energized. This process is repeated as long as the processor is in the RUN mode. If the input device were to open, a 0 would be placed in the input image table. As a result, the rung would be said to be False. The processor would then set the output image table bit 001–04 to 0 causing the output device to turn off.

21.15.19 PLC Hardware Components

1. I/O Section

The input and output interface modules consists of an I/O rack and individual I/O modules. Input interface modules, accept signals from the machine or process devices (120V ac) and convert them into signals (5V dc) that can be used by the controllers. Output interface modules convert controller signals (5V dc) into external signals of 120V ac used to control the machine or process.

In large PLC systems, I/O sub system can be remotely located from the CPU. A remote subsystem is usually a rack type enclosure in which the I/O modules are installed. An interconnecting cable allows communication between the processor and the remote I/O rack.

The location of a module within a rack and the terminal number of a module to which an input or output device is connected will determine the device,s address. Each input and output device must have a specific address.

This address is used by the process to identify where the device is located in order to monitor or control it. In addition, there is some means of connecting field wiring on the I/O module housing. Connecting the field wiring to the I/O housing allows easier disconnection and reconnection of the wiring in order to change modules. Lights are also added to each module to indicate the ON or OFF status of each I/O circuit. Most output modules also have blown fuse indicators.

A standard I/O module consists of a printed circuit board (PCB) and terminal assembly. The PCB contains the electronic circuitry used to interface the circuit of the processor with that of the input or output device. It is designed to plug into a slot or connector in the I/O rack or directly into the processor. The terminal assembly, which is attached to the front edge of the PCB is used for making field wiring connections.

2. Discrete I/O Modules

The most common type of I/O interface module is the discrete type. This type of interface connects field input devices of the ON/OFF nature such asselector switches, push buttons and limit switches. Similarly output control is limited to devices such as lights, small motors, solenoids and motor starters that require simple ON/OFF switching.

Each discrete I/O module is powered by some field supplied voltage source. Since these voltages can be of different magnitude or type, I/O modules are available at various ac and dc voltage rating as listed in Table 21.1.

Table 21.1

Input Interfaces	*Output Interfaces*
24V ac /dc	12-48V ac
48V ac/dc	120V ac
120V ac/dc	230V ac
240V ac/dc	120V dc
5V dc (TTL level)	230V dc
	5V dc (TTL level)

Figure. 21.70 shows a block diagram for one input of a typical ac interface input module.

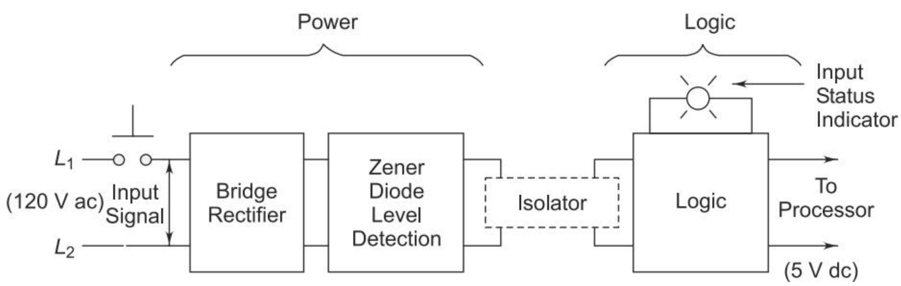


Fig. 21.70 Block diagram of ac interface input module

The input circuit is composed of two basic sections—the power section and the logic section. The power and logic sections are normally coupled together in a circuit, which electrically separates the two.

A simplified schematic and wiring diagram for one input of a typical ac input module is shown in Fig. 21.71(a) and (b). When the push button is closed, 120V ac is applied to the bridge rectifier through' R_1 and R_2. This produces a low level dc voltage which is applied across the LED of the optical isolator. The Zener diode (ZD) voltage rating sets the minimum level of voltage that can be detected . When the light from the LED strikes the photo transistor, it switches into conduction and the status of the push button is communicated in logic or low level dc voltage to the processor. The optical isolator not only separates the higher ac input voltage from the logic circuits, but also prevents damage to the processor due to line voltage transients.

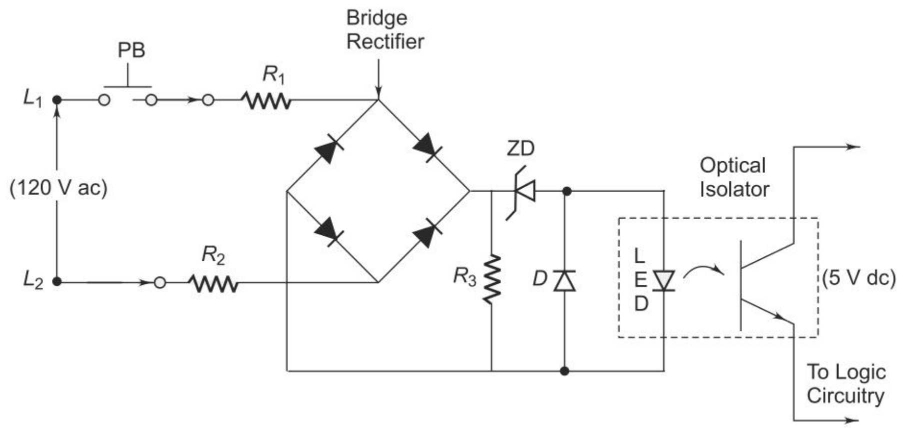


Fig 21.71 (a) Simplified schematic of an ac input module

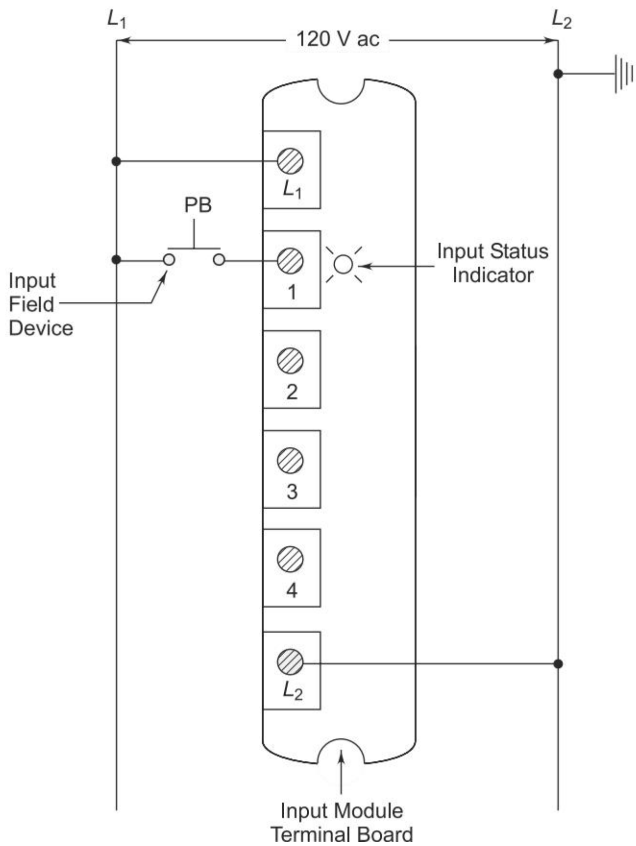


Fig. 21.71 (b) Typical input module wiring connection

Optical isolation also helps reduce the effects of electrical noise, common in industrial environment which can cause erratic operation of the processor. Coupling and isolation can be accomplished by use of a pulse transformer.

Figure. 21.72 shows a diagram for one output of a typical interface output module. Similar to the input module, it is composed of two basic sections, the power section and the logic section, coupled by an isolation circuit. The output interface can be considered of as a simple electronic switch to which power is applied to control output device.

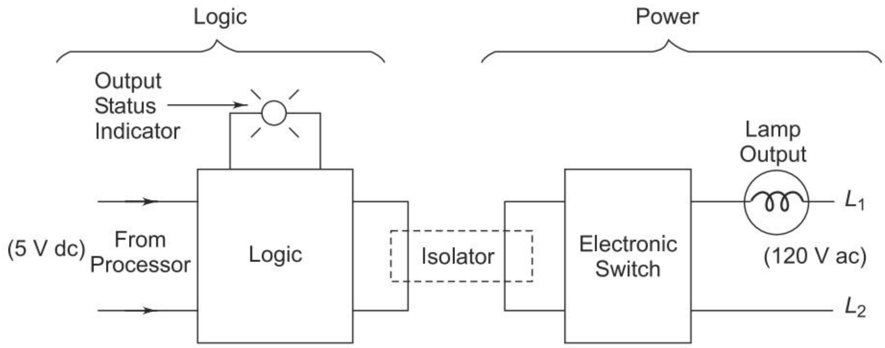


Fig. 21.72 Block diagram of ac interface output module

A simplified schematic and wiring diagram for one output of a typical ac output module is shown in Fig. 21.73. As part of its normal operation, the processor sets the output states according to the logic program. When the processor calls for an output, a voltage is applied across the LED of the isolator. The LED then emits light which switches on the phototransistor into conduction. This, in turn, switches a semiconductor switch such as Triac into conduction which turns on the lamp. Since the triac conducts in either direction, the output to lamp is ac. The triac rather than having ON and OFF status, actually has low and high resistance levels respectively. In its OFF state (high resistance) a small leakage current of a

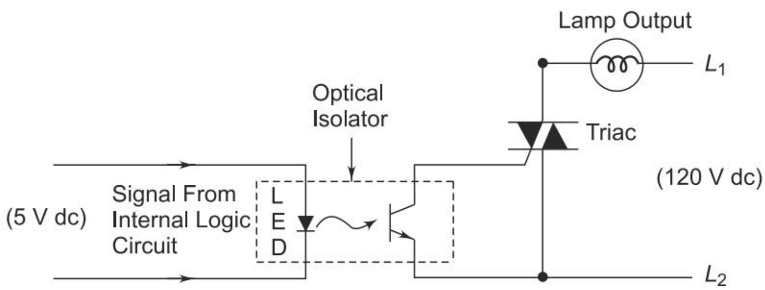


Fig. 21.73 Simplified schematic of an ac output module

few mA still flows through the triac. As with input circuits, the output interface is usually provided with LED's that indicate the status of each output. In addition, if the module contains a fuse, a fuse status indicator may also be used as shown in Fig. 21.74.

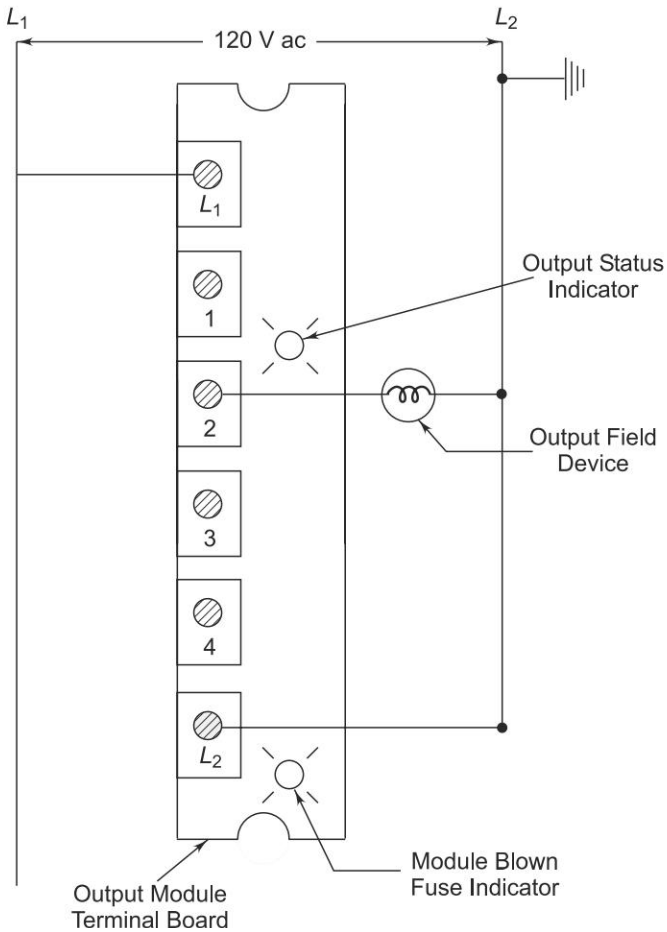


Fig. 21.74 Typical output module wiring connection

Individual ac outputs are usually limited by the size of the triac to 2 or 3A. The maximum current load for any one module is also specified. To protect the output module circuits, specified current ratings should not be exceeded.

For controlling large loads, such as large motors, a standard control relay is connected to the output module. The contacts of the relay can then be used to control a large load or motor starter as shown in Fig. 21.75. When a control relay is used in this manner it is called an *interposing relay*.

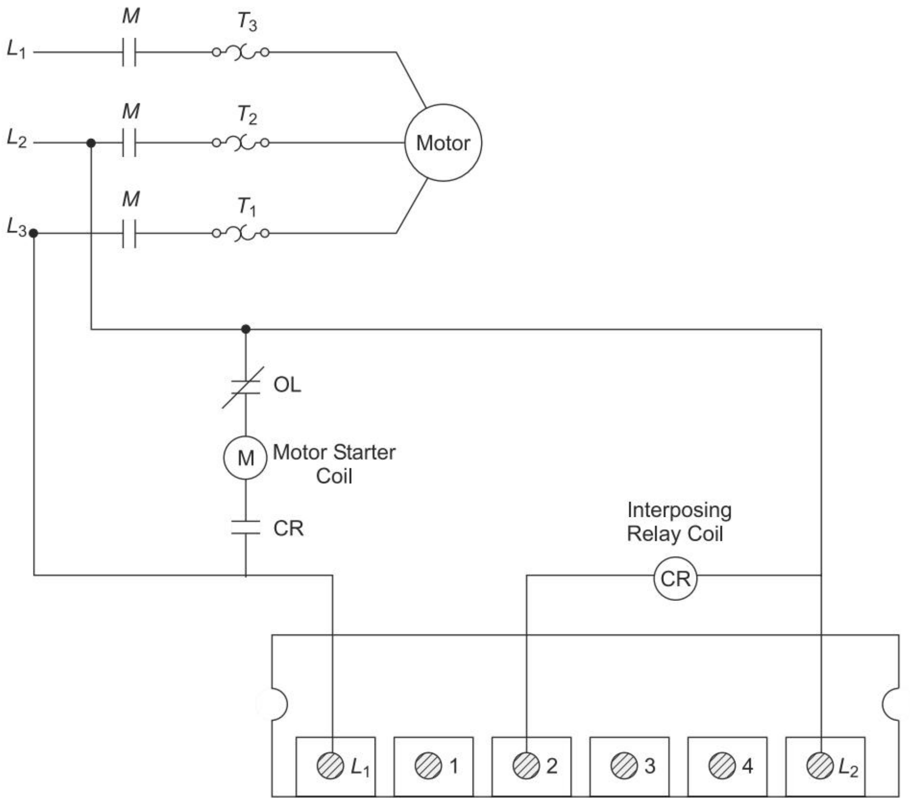


Fig. 21.75 Interposing relay connections

3. Analog I/O modules

Earlier PLCs were limited to discrete I/O interfaces, which allowed only ON/OFF type devices to be connected. This limitation meant that the PLC could have only partial control of many process applications. Today, however a complete range of both discrete and analog interfaces are available that will allow the controller to be applied to practically any type of the control process.

Analog input interface modules contain the circuitry necessary to accept analog voltage or current signals from analog field devices. These inputs are converted to digital by the use of an analog to digital converter (ADC). The conversion value, which is proportional to the analog signal, is expressed as a 12 bit binary or as a 3-digit BCD for the use by the processor. Analog input sensing devices include temerature, light, speed, pressure and position transducers. Figure 21.76 shows a typical analog input interface module connection to a thermocouple. A varying dc voltage in mV range proportional to the temperature being monitored, is produced by the thermocouple. This voltage is then amplified and digitized by the analog input module and then sent to the processor on command from a program instruction. Because of the low voltage level of the input signal, a shielded cable is used in wiring the circuit to reduce unwanted electrical noise signals that can be induced in the conductors from other wiring. This noise can cause temporary operating errors that can lead to hazardous or unexpected machine operation.

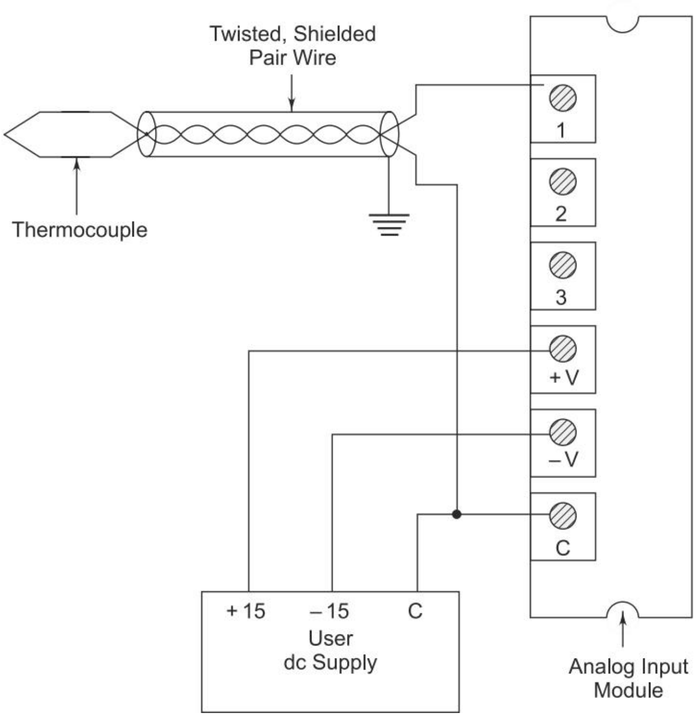


Fig. 21.76 Typical thermocouple connection to an analog input module

The analog output interface module receives from the processor digital data which is converted into a proportional voltage or current to control an analog field device. The digital data is passed through a DAC circuit to produce the necessary analog form. Analog output devices include small motors, valves, analog meters and seven segment displays.

4. T/O Specifications

Manufacturer's specifications provide much information about how an interface device is correctly and safely used. The specification places certain limitation, not only on the module, but also on the field equipment that it can operate. The following is a list of some typical manufacturer, I/O specifications.

1. *Nominal input voltage* This ac or dc value specifies the magnitude and type of voltage signal that will be accepted.
2. *On state input voltage range* This value specifices the voltage at which input signal is recognized as being absolutely on.

3. *Nominal current per input* This value specifies the minimum input current that the input devices must be capable of driving to operate the input circuit.
4. *Ambient temperature rating* This value specifies what the maximum temperature, of the air surrounding the I/O module, should be for best operating conditions.
5. *Input delay* This value specifies the time duration for which input signal must be ON before being recognized as a valid input. This delay is a result of filtering circuitry provided to protect against contact bounce and voltage transients. This input delay is typically in the 9 ms to 25 ms range.
6. *Nominal output voltage* This ac or dc value specifies the magnitude and type of voltage source that can be controlled by the output.
7. *Output voltage range* This value specifies the minimum and maximum output operating voltages. An output circuit rated at 120V ac, for example may have an absolute working range of 92 V ac (min) to 138 V ac max.
8. *Maximum output current rating per output and module* These values specify the maximum current that a single output and the module as a whole can safely carry under load (at rated voltage).
9. *Maximum surge current per output* This value specifies the maximum in rush current and duration (for example 20A for 0.1 second) for which an output circuit can exceed its maximum continuous current rating.
10. *Off-state leakage current per output* This value specifies the maximum value of leakage current that flows through the output in its OFF state.
11. *Electrical isolation* This maximum value (volts) defines the isolation between the I/O circuit and controller's logic circuitry. Although this isolation protects the logic side of the module from excessive input or output voltage or current, the power circuitry of the module may be damaged.

DISTRIBUTED CONTROL SYSTEMS 21.16

Since the beginning of process control in late 1970s, it has made great progress in every control activity. Similar to the other electronic equipment, its size is decreasing and its power is increasing. The conventional instrumentation is now outdated. The speedy development in the field of electronics has left the instruments working on mechanical lever, bellows and pointers far behind. The so-called miniature instruments of the size of 6”x3” are no longer miniature as compared to CRT console which is capable of handling hundreds of loops. The concept of control room has been totally changed.

Instrumentation is no longer limited to the measurement of temperature, pressure, flow, level, etc. parameters etc parameter, as an independent paramter using individual controller for each parameter. This process of using individual controller is known as parallel control concept. The instrument system has now been enhanced to complex high speed machinery data, integrating control of various plants data presentation for the management, etc. through the concept of distributed control in which the controllers are distributed around the plant graphically and also functionally.

Since instrumentation is a highly responsible task nowadays, the system should be quite sufficient, reliable and supported with effective subsystem for carrying the task.

With the introduction of microprocessor, the process control has become a window to the process enhancement and invaluable data generated as a byproduct has widened the field of process control to include process optimization and computer integrated manufacture.

The present trend in process control is to evolve a complete system with process control as basic function over which the interactive operator interface, process optimization, production planning and control can be built to suit the market needs. Distributed control systems open the gateway to better realization of process design with host computers and process models gainfully utilizing the data collected by the DCS. A distributed control system can briefly be defined as a control system with communication capabilities and it is a powerful and sophisticated tool in the hands of process control engineers.

The utilization of digital computer to start with and development of control strategy based on microprocessor embedded distributed control system for the plant makes it interesting development in process control. The latest trend is distributed control systems with intelligent control station physically distributed over the plant tropology.

21.16.1 Advantages (Added Facilities) in DCS Technique

Following are the advantages or the added facilities as a result of the DCS:

1. Data presentation is in a systematic format enabling easy comparison of various parameters and taking decision by a printer.
2. Logging of data is done by a printer thereby eliminating human error.
3. It is possible to control through dynamic graphic.
4. Operator's action can be logged, thereby eliminating confusion.
5. The alarm system can be regrouped.
6. Complex computations, analysis, etc. can be carried out easily.
7. Management information can be generated at regular intervals.
8. The super-imposed trends helps in the analysis of plant parameters and behaviour.
9. Hardcopy gives the actual dynamic printout at a particular instant.
10. The controlling software used is very simple and the application is readily understood. The software changed in one unit has no impact on other units and hence the system becomes very flexible. User's risk in software is minimum.

21.16.2 Disadvantage of DCS Technique

1. All information and data though presented in a systematic format is hidden behind the CRT. Hence, it requires a skilled operator.

2. In an emergency, decisions have to be taken single handedly, as few operators are there in the control room.
3. Failure of one controller effects more than one loop. Hence it calls for very high MTBF (Mean Time Between Failures) and high degree of redundancy.

21.16.3 System Configuration

The system configuration available in nearly all types of DCS present today in the market is composed of the following fundamental blocks.

1. Operator station
2. Field control station
3. Communication bus
4. Gateway unit
5. Programmable logic controller

Different brands specify their system differently having varied combination of the above blocks with addition/deletion of features.

21.16.4 DCS Process Control

A DCS is preferred over a central computer system for the following reasons.

Central Computer System (CCS) exposes the plant to the risk that the system might fail. This can be minimized by having one or more back up schemes.

(a) A back up computer
(b) Analog controller
(c) Manual loading stations
(d) Various combinations of the above.

These back up methods are expensive. The DCS also needs a backup but it distributes the risk of failure into smaller sub-sections. The failure of one node should not cause the failure of another portion. It decreases cost, kinds of redundancy and backup needed. In short, it has the capability to fail gracefully.

Another advantage is the substantial reduction of wiring costs. The data highway is generally a twisted shielded pair of wires or some type of coaxial cable which carries all the processor and computer system information required for efficient plant operation via a high speed digital data communication system. It replaces many miles of copper cable for sensor inputs and final control element outputs. Hence, a large saving in a plant with many inputs and outputs can be achieved. Routine process control is performed in the distributed micro-computers with their associated Input/Output (I/O) cards, multiplexer, ADC and DAC converters. The failure of computer at this node will leave, for example, the control valves at their last position due to hard circuitry on the output card. If backup microcomputer is installed, it will automatically take over control without a bump to the process. If no backup microcomputer is installed, the manual analog loading station may be switched in to maintain control with an operator in attendance.

21.16.5 Design features of TDC 3000 Architecture

Perhaps it would be better if we give an introudction as to what is TDC 3000 Architecture?

(a) The TDC3000 system It is designed to provide dependable service in an industrial environment. Failures are infrequent and contained. Failure effects are minimized by a fault tolerant design that also provides for optional redundancy of critical system modules. Isolation, part replacements and the return to operation of failed modules is speeded up by built- in hardware and software fault checks.

One of the major design goals of the TDC 3000 has been to ensure a high level of system availability. First, by increasing the mean time between failures and second, by lowering the mean time to repair.

The first objective has been met by the use of conservation design and use of quality components. The second objective is met by providing a variety of built in diagnostic tools designed to pinpoint faults to an Optimum Replacement Unit (ORU).

(b) Design Benefits The TDC 3000 design philosophy of full distribution of architecture and full integration of functions give important benefits to the user. The manufacturer of TDC 3000, Honeywell, has provided expanded capabilities much more that what is available with any other system.

(c) Incremental Expandability The functional specialization of the physical modules allows the purchaser to start with the minimal devices that match the initial requirements and the available funds. When it is required to expand, as a result of either an increase in process dimensions or the fundamental needs, it can be easily achieved by the simple addition of hardware modules, without any significant effect on the existing system, i.e. no reprogramming of the software is needed.

(d) Global Database No matter, whatever is the size of your TDC 3000 system, it functions with a single database. This database is accessible to all users by way of a single access mechanism. Since multiple access mechanis's are not required, expansion engineering costs as well as initial engineering costs are minimized.

(e) Network File Management System For large systems, with more than one process or application engineer having frequent need for simultaneous access to the system in order to build and edit files, the online network file management capability is invaluable.

For example, one engineer seated at a universal workstation in an office can design and implement a new control strategy in control language on an application module, while at the same time another engineer at a different universal workstation can update a graphic display for use by the operator. Also at the same time, a maintenance technician might be accessing data required to analyse a problem that has been logged by the system. While all this is taking place, the process is under control and the operators are performing their normal duties at the operator console.

(f) Distributed Risk In a control system with centralized architecture, failure of a component can bring down the entire system. With the distributed architecture of the TDC 3000 graceful degradation, a long time feature of Honeywell's Data Hiway-based systems will be in effect. A failure will, in the worst case, take the affected module with its assigned function, out of service depending on the importance of the failed module and whether or not it is redundant, the remainder of the system may continue to perform with minimum impact resulting from the failure.

21.16.6 System Architecture

1. Hiway Based Distributed Control

The building block concept of electronic control design, with distributed process controlled boxes linked together by a communication network, was first introduced by Honeywell in 1975. Today, the monitor and control boxes connected to the Data Hiway as shown in Fig. 21.77 are an integral part of the TDC 3000.

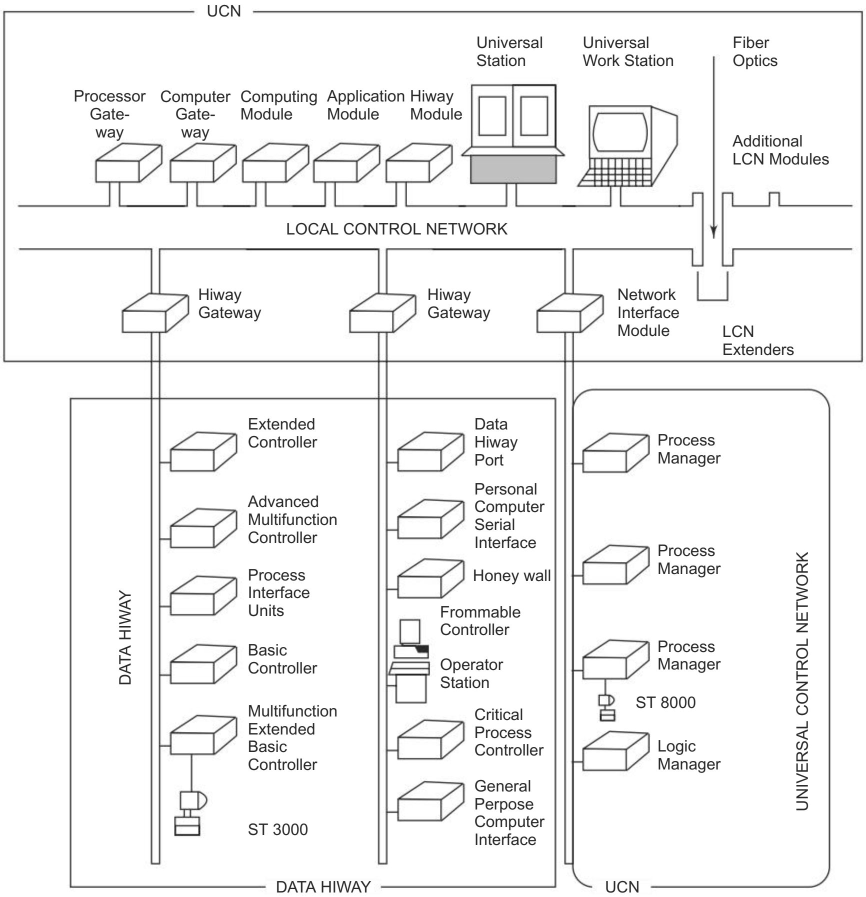


Fig. 21.77 TDC 3000 architecture.

2. Control Room Functions

By incorporating a high speed proprietary LAN called the local control network (LCN) with associated modules that accomplish specific functions, TDC 3000 offers a product that is compatible with systems already in the fields, is fully integrated and at the same time, has flexibility to continue to change and to grow along with the processes it controls. The user need purchase only those modules required to meet plant control objectives at the time, with the assurance that increased capacity and/or functionality can be achieved later by adding the additional module as required.

An LCN has the following features:

- A bit serial communication path that operates at 5 Mega bits.
- It provides communication links between all nodes of TDC 3000.
- A redundant pair of 75 ohms co-axial cables connected to each node.
- Each coaxial segment may be up to 300 m long which may be extended up to 2000 m with optic fiber cables to a remote universal station or another coaxial segment.
- It connects to up to 64 nodes. After 40 nodes it requires LCN extender.
- It may also be extended to up to 96 nodes as a special option.

3. Input output Devices (Process Interface Units (PIU)

PIU are used when a large number of process variable need to be monitroed or fed to a higher level module. They are available in low level (LLPIU), high level (HLPIU) and low energy (LEPIU) version. The levels refer to the amplitude of the signals with which the PIUs interface. The low energy PIU is configured with multiplexer boxes that can be distributed in the plant.

4. Controllers

The various controllers used in DCS (TDS3000) is listed below along with their features.

(a) *Basic Controller (CB)* This controller handles only continuous Input/Output (I/O) operations. It can accommodate input/output for eight control loops, plus eight additional monitor-only inputs.

(b) *Extended Controller (EC)* EC is used to handle continuous Input/Output and logical operations. It provides sixteen control outputs, eight outputs, sixteen digital outputs and sixteen digital inputs.

(c) *Multifunction Controller (MC)* MC controller used in DCS, handles continuos I/O operations, logical operations and sequential control for batch operations. Each MC has the capacity to control 16 analog loops and handling up to 30 analog inputs, up to 72 analog output, up to 256 digital inputs, up to 64 counter inputs.

(d) *Advanced Multifunction Controller (AMC)* AMC is the modification of MC and it has the same function as MC. It takes less space and the scanning is faster than MC.

All of the above controller types can be equipped with a digital interface to Honeywell's ST 300 smart transmitter, making it possible for the controllers to take advantage of the increased capabilities of this transmitter.

21.16.7 Other Boxes

The following are the additional boxes (units) provided in a DCS.

Data Hiway Port (DHP) Data hiway port are interface between TDC3000 and non-Honeywell process connected boxes. The list of devices included in this box are analysers, tank gauging systems, machinery monitoring systems, emergency shutdowns systems, data acquisiton systems and compressor control systems.

Critical Process Controller (CPC) The critical process controller consists of a fault tolerant logic control system and also triple redundancy.

Personal Computer Serial Interface (PCSI) This unit provides an interface to any personal or mini computer with an RS-232 serial port. Typically the PC scans selected box values which could be viewed or archived into the PC files for such applications as the data analysis, customed reports and special calculations.

General Purpose Computer Interface (GPCI) This section provides the interface between the TDC 3000 Data Hiway and any computer having general purpose I/O adapter. It has 16-bit parallel communication.

Honeywell Programmable Controller This consists of a Logic Control system (LCS) and IPC620 line of programmable controllers. It can perform logic and interlock function.

21.16.8 Man/Machine Interface Functions

Universal Station (US) US is a Universal Window through which the operator monitors and controls the process and handles process and alarm systems. It stands on the LCN and communicates with all modules on the LCN, process connected devices on the Hiway via Hiway Gateway and UCN mode, via interface module.

It provides comprehensive facilites to the process operator, process engineer and maintenance technician on the universal window.

For Process Operator The following are the functions for the process operators. They have to:

- Monitor and manipulate both continuous and discontinuous process or portions of processes
- Annuciate and handle process alarms, system alarms and operator messages.
- Display and print process history, process trends and process averages.
- Display and print journals, logs and reports.
- Monitor and change status of system equipment in the control room and in the field.
- Load other system modules with operating programs and databases from History Module, floppy diskette or cartridges.

For Process Engineers The process engineer do the following.

- Build system and process database, graphic displays and reports.
- Perpare edit and compile CL programs.
- Loading operating programs and database from or to history modules and floppy diskettes.
- Load standard Honeywell software.
- Configurations of network.

For Maintenance Technicians The maintenance technicians have the responsibility to

- Diagnose problem in any module in the system.
- Display and print troubleshooting information.

Universal Workstation (UWS)

- This workstation is made available for process engineering and maintenance function and for monitoring by the process supervisor.
- It has the capability of virtually all the functions as the Universal Station but is designed for an office environment.

Storage Function The followings are the storage functions:

- Mass storage media.
- History of process alarms, operator changes, operator messages. system status changes, system errors and system maintenance recommendations, also continuous process history to support logs and trends.
- Check point data for maintaining up to date box and module setting in the event the device is taken out of service.
- Global database can be assessed by any mode on the LCN.
- Storage of system software and application.

21.16.9 Second Level Control

This control consists of an application module or AM.

Application Module (AM) AM has the following features:

- It provides supervisory level control.
- It has access to broader scope of information.
- It can also access any process connected box in the system.
- It provides programming for custom algorithms through powerful process oriented lalnguage CL.

21.16.10 Third Level of Control and Information Management Computing Module

This module consists of the following features:

- It has an advanced control programs beyond capabilities of standards algorithms and control language (CL).
- It can perform simultaneous calculations such as multivariable process control.

- It has access to any process connected box.
- It provides an environment where the user can develop, debug and execute programs to perform such function as process optimization, advanced custom report generation, long term data storage, scheduling and plant management.

21.16.11 Microprocessor-based Control System: Distributed Control Systems

Modern plants nowadays are very complex, large scale systems with a very high degree of automation. The designs and the operation of these plants are determined by the criteria of economy availability and safety. For example, in thermal power plants, greater unit capacity with higher temperatures and higher pressures are being adopted.

Thermal power plants are required to be extremely fuel efficient because of the increasing fuel costs. Designs of instrumentation and control systems are being oriented to achieve these goals. Today microprocessor-based digital process control systems provide reliable and satisfactory control. Distributed Control Systems (DCS) are characterized by

(a) Microprocessor-based self contained functional hardware modules operating in parallel.
(b) Data buses for fast data transmission.
(c) Shared video display units for operator communication and monitoring.

The DCS provides physical means of achieving optional economy, maximum reliability and safety and minimum environmental impact. Software, that is, concepts and algorithms are available to achieve advanced power plant control schemes. The benefits of improved control are:

(a) **Improved efficiency**
   (i) by better combustion
   (ii) by reduction of losses during start-up or cycling and
   (iii) improved thermodynamic efficiency by higher mean final steam and temperature.
(b) **Longer lifetime** of components (especially critical parts) due to reduced transients stresses.
(c) **Reduced maintenance** on actuation due to smoother action.

In addition to the economic benefits of improved control, following advantages are gained.

(i) Lower cost per function (software function modules).
(ii) Less electronic hardware (less number of cabinets required).
(iii) Lower cabling and installation cost (multiplexing, data buses).
(iv) Lower costs for documentation (self documentation).
(v) Reduced maintenance costs (self diagnosis, less spare parts).
(vi) Shorter design and commissioning phase (flexibility, CAD).
(vii) Easy changes in structure, e.g. after plant changes (flexible firmware).

(viii) Higher availability, less disturbances and faults, shorter outages designed redundancy, self checking, error locating fast repair, etc.

Control system using totally distributed digital control system (DCS) is designed to achieve the following goals.

1. *Reliable Plant Operation* To diagnose system conditions and to continue plant operation in case of faults. Provision of redundancy is done, as appropriate at various levels. This can be obtained by the use of decentralized functions.
2. *Man Power Saving* This can be achieved by having a fully automated centralized operation and quick recovery from faults by centralised control and supervision.
3. *Rapid Load Response* This can be achieved by reducing steam temperature error which is due to slow response of fuel loop.
4. *Dynamic Balancing Control* This is obtained by predictive control.

21.16.12 Decentralised Functions

1. Distributed Control System Overview

The Instrumentation and Control (I & C) system is divided basically into the following four broad operating areas also shown in Fig. 21.78.

(i) Control Room
(ii) Electronic Room
(iii) Switchgear
(iv) Field

(i) Control Room All data required for process control is concentrated, conditioned and made available in the control room, so that clear information on the current status is available for operations at all times. The operator can intervene in the process using controls if necessary, e.g. switch a unit ON or OFF, or can take a closed loop control on manual and modulate the activator 'raise' or 'lower'. These actions can be carried out using:

(a) process operation keyboard together with CRT monitors.
(b) Conventional auto manual control files from backup panel.

The CRT monitors display the current process status, process faults and I & C system faults. The above data is organized in such a manner that it is easily analysed by the operator and corrective action is taken, if required. With centralized monitoring using CRTs and using printers for logging operation, the requirement of the control room operations is reduced and load on the operators is also reduced.

(ii) Electronic Room All the data acquired from the field (process) is processed, the required displays are generated and commands are issued from the electronic cubicles. The electronic cubicles are located in the electronic room. The electronics can be broadly divided into the following functions:

(a) Input/Output (I/O) section
(b) Control units
(c) Operator communication system

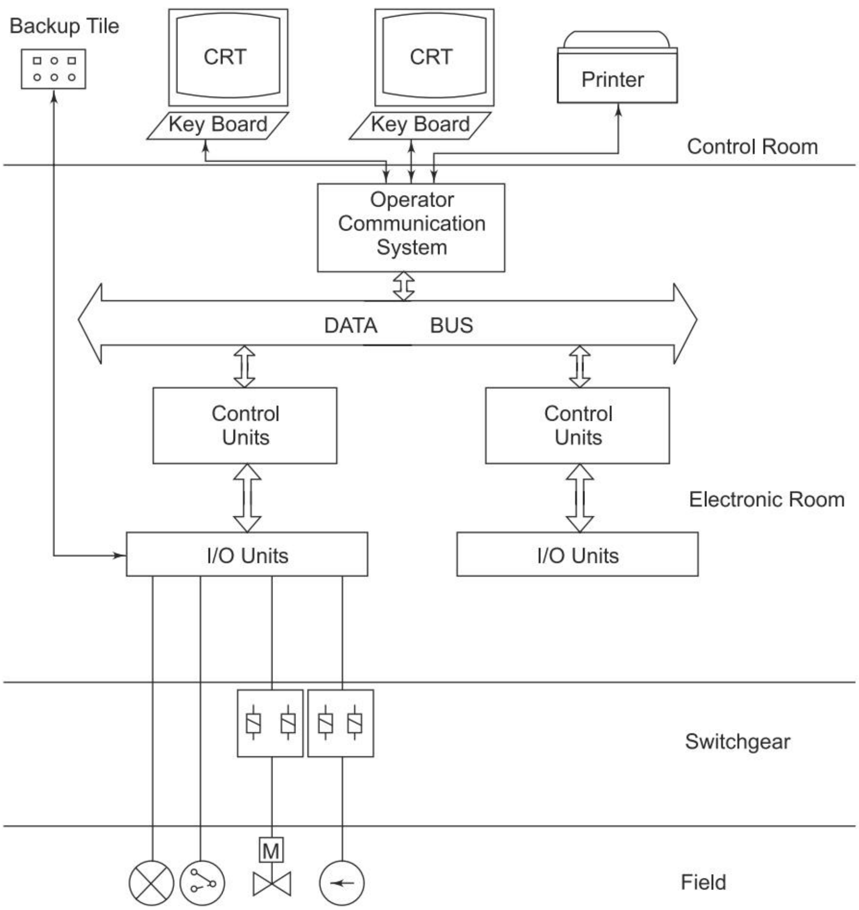


Fig. 21.78 Distributed control system overview

(a) **Input/Output section** houses the field input/output connections, carries out signal conditioning and houses the drive control modules.

(b) **Control units** carry out the open loop control functions, closed loop control functions, interlock, and protections. Control units are functionally distributed and redundancy is provided at appropriate level to avoid any process disturbance in case of single fault. Also, manual control can be carried out at drive level through back-up.

(c) **Operator communication** system carries out the function of display generation, generation of logs and reports and also controls the communication between control room equipment, like CRTs keyboards, printers, etc. and the control units. Redundancy is provided at this level to ensure that control room information is not lost in case of fault at this level.

All communications between Input and output (I/O) units, control units, operator communication systems are over system data bus. The panel to panel wiring and the wiring inside the cubicle is done using data bus with plug-in type connectors.

(iii) Switchgear All actuators, valves, pumps, fan, etc. are provided with respective switchgear. Each switchgear is provided with two numbers interposing relays for ON/OFF (open-close) operation. The interposing relays receive commands from drive control modules.

(iv) Field All transducers, actuators, pumps, fans, etc. are located in the field (process). All transducers, actuators feed back are connected to the Input/Output units using single pair/multipair instrumentation cables.

Based on the method of distribution of electronic cubicles, the DCS can be divided into two categories:

(a) Functionally distributed
(b) Geographically distributed

(a) In a functionally distributed system, all the electronic cubicles are located in one electronic room, and all the field inputs and outputs are wired directly to these cubicles. The control processors are functionally distributed to carry out different tasks.

(b) In a geographically distributed system, the field Inputs/Outputs are located in different rooms at various locations in the plant. The cubicles located in the field may have control units. The different cubicles located geographically at different locations in the plant communicate with each other and with the centrally located electronic cubicles over the system data bus.

All the electronic rooms located at various locations in the plant are required to be made dust-proof and air-conditioned. If geographic distribution is used, the dual redundant buses are routed through different routes to ensure reliability against physical damage.

2. Instrumentation and Control System Implementation (I & C)

Because of the differences in the characteristics among the various functions the functions are distributed hierarchically and assigned appropriate hardware as shown in Figs 21.79 and 21.80, respectively.

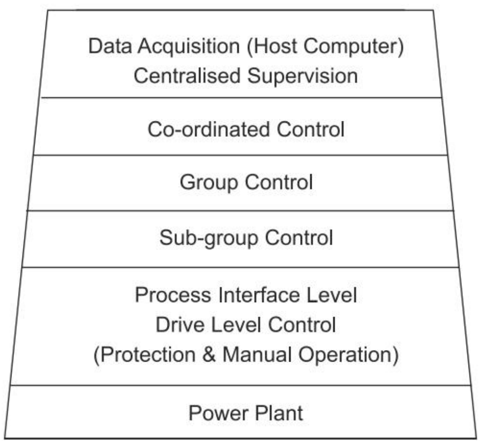


Fig. 21.79 Instrumentation and control system implementation

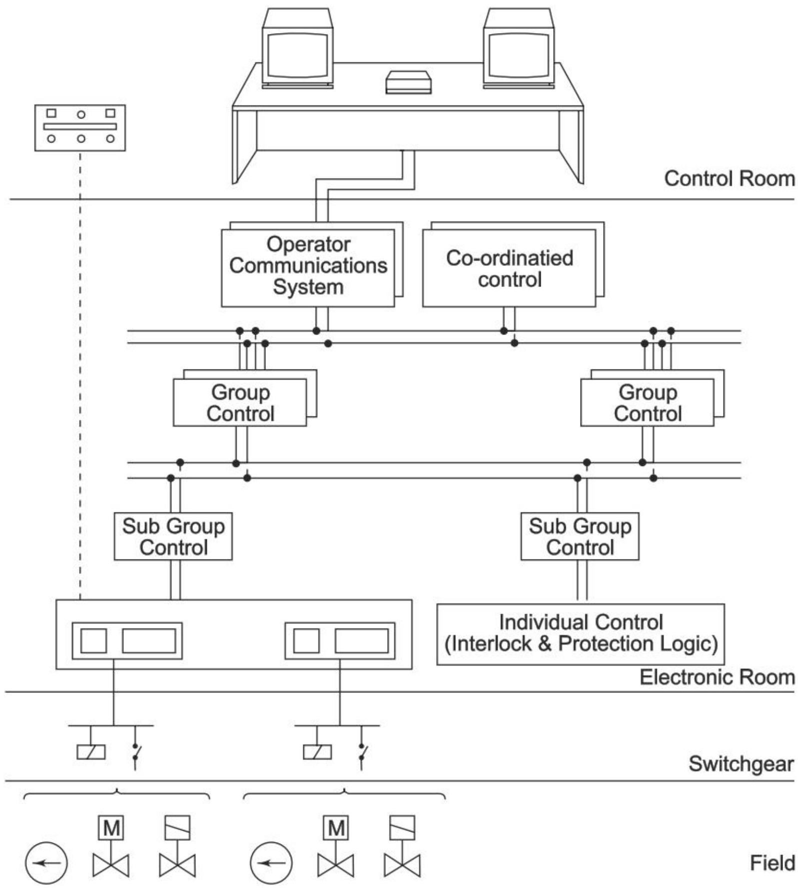


Fig. 21.80 Distributed microprocessor-based control system-(System configuration)

(i) Data Acquisition System (Host Computer) The I & C systems till about 1980's were based on hardwired electronics. With the hardwired system the plant controls were carried out using hardwired control modules, push buttons, indicators, recorders, annuciation etc. Data Acquisition System (DAS)was engineered to perform the following functions using independent computer.

(a) Data acquisition and processing
(b) Alarm monitoring
(c) Process information display
(d) Start up and shut down guidance
(e) Performance monitoring
(f) Report Generation

The DAS had 100% plant information to carry out assigned functions, however, no control function was assigned to it. The Instrumentation and Control (I&C) and DAS systems were so designed that even if DAS, due to some reasons was not available, still the plant operation would not be affected.

The microprocessor based I&C system perform the following functions, many of which were earlier performed by the DAS.

(a) Scanning and processing analog, digital and pulse inputs
(b) Closed loop controls
(c) Open loop controls
(d) Process and control information on command
(e) Alarm monitoring and reporting

In view of this, a smaller computer is provided to do the following functions:

(a) Generation of logs and reports
(b) Performance calculations
(c) Sequence of events recording
(d) Plant management information system

The computer above is designated as the Host Computer. It periodically acquires the processed plants data over the plant bus. This data is further processed in the Host Computer for performing the function given above.

(ii) Coordinated Control Coordinated control level is the highest level of control in the power plant. This issues commands for the automatic start up or shut down of the plant. This also generates set points for sub-ordinate modulating controllers or slave controllers.

For example, coordinated masters control issues load run back commands on loss of an auxillary such as boiler feed pump.

(iii) Group Control The Group control is superimposed on the sub-group control. The individual functional area of the plant are divided into groups. Each group consists of several sub groups.

For example, if a group is Feed Water then the sub groups are:

- Boiler feed pump A
- Boiler feed pump B
- Boiler feed pump C
- De-aerator

The group control decides, when, how many and which sub-groups are to be put into operation or shut down. The group controls are controlled by coordinated control. If a fault occurs in a sub-group, the group control switches over to a standby sub-group.

(iv) Sub-group The sub group control is superimposed on the individual control drive and sub-loop control. The sub-group control receives its commands either from the control station or from the superimposed group control. The sub-group control has two operation programs operation and standby.

Process signals and plant feedbacks provide information on the status of the sub-group. Step sequence control or logic control is used for the sub-group control.

(v) Process Interface Level The process Interface level establishes the signal connection between the automation system and the sensors in the process

(Transducers, feedback equipment) and the final control elements (actuator, switchgear).

The modules for signal conditioning, individual control drives and modulating controllers are therefore assigned to the process interface level.

(vi) Signal Conditioning Signal conditioning consists of the power supply to the binary and analog sensors, simulation, monitoring, conditioning and distribution of the signals from the sensors.

(vii) Protective Interlocks The protective interlocks have the function of protecting the plant auxillaries from abnormal conditions. A logic operation is carried out on binary signals with the objective to:

(a) detect faults
(b) initiate counter measures (protective command)
(c) prevent switching operations (no release signals)

Drives are switched on and off using protective commands. A drive is prevented from being switched on or off if the release criteria is missing. The protective interlocks are assigned the highest priority.

viii) Open Loop Control Drive The open loop control drive carries out open loop control functions for drives such as motors, actuators, and solenoid valves. It consists of the processing of control room commands, triggering of the drive, acquisition and processing of feedback from the drive and the switchgear. Also included are the monitoring of the drive for running time and status discrepancies, monitoring of the feedbacks and output of the status and fault signals.

(ix) Modulating Control Drive The modulating control consists of the positioning of the associated control drive, processing of control station commands, acquisition of feedback from the drive and process signals. Also included are the acquisition of analog position feedbacks, monitoring of the feedbacks, generation of the control deviation from the actual value and set point, processing of binary commands from higher level controls.

The process interface level constitutes an independent level within the automation system, subordinate to the open loop and closed loop controls. The operator can continue operation of the plant with manual control at drive interface level, even under fault at higher level unit.

3. Redundancy

To achieve high degree of reliability and availability, redundancy is provided at higher levels, i.e. at the centralized supervision, co-ordinated control and group control level. One number drive control module is provided per drive, advantage of this is that due to any fault at drive control level, only one drive is disabled.

In case of failure of one controller the back up controller takes over the control without any disturbance to the process. The controller modules are checked by using self diagnostics programs. The communication buses are also dual redundant.

During normal operation one CRT is assigned alarm functions while other CRTs take care of display and control. In case of failure of one of the CRT, the other takes over the functions of the failed CRT, in addition to its designated functions. In these conditions, the top portion of the active CRT screen is reserved for alarm display, while the remaining area of the screen is used for other displays. All cubicles, in the system have redundant power supply units.

Redundancy is also provided at the transducer level. For critical analog measurements, three transmitters are provided with median out of three signals used for control, e.g. furnance draft measurement. For critical binary signals there are three switches, with two out of three signals is used, e.g. Burner management system uses 2 out of 3 signals to ensure high reliability and high availability.

4. DCS Functional Requirements

(i) System Availability Requirements The total system availability is above 99.5% as calculated from MTBF and MTTR of all equipment/devices which are interconnected to form the system.

(ii) System Load The DCS is so designed that the system load is less than 60% with all the programs running simultaneously. This ensures proper response during plant disturbances.

(iii) System Response The following minimum response requirements are ensured

(a) Closed loop control is performed with a minimum cycle time of 100 ms.
(b) Controller response (time interval between input change and updated controller output) under worst loading conditions is less than one second.
(c) Command from the console to activate field devices is executed within one second.
(d) Alarms are reported within one second.
(e) The total display page appears on CRT within three seconds of operator's request.

(iv) Environmental Requirements Even though DCS are guaranteed to work at 50°C and relative humidity of 95%, it is preferable to locate the electronic cubicles in air conditioned atmosphere to achieve better operating performance.

21.16.13 Centralised Control and Supervision

Information processing nowadays is playing a very important role mostly in the control of thermal power plants. Mimic diagrams, bar charts and video trend graphs displayed on colour CRTs are now replacing conventional indicators, meters and recorders with the logging functions being taken over by printers and with the CRT displays designed in a manner to do fault analysis for the operator, the operator's role in the control room is reduced and so is the burden.

Man-machine Communication

The man-machine communication in the DCS is through colour CRTs and keyboards. Printers are provided for hardcopies of logs reports and alarm messages. The control room consists of a operator's console, backup panel, electronic cubicles and printers. The central operating center has the facility to control the main plant areas and associated auxiliaries. Independent control room are provided to control auxillary plants such as DM water plant, coal handling plant, Ash handling plant, and Electrostatic precipitator.

Critical data from the above plants is transmitted to main control room for information to the control room operators.

The control room operators control the plant through various displays available on the CRTs. The displays are created with static display and dynamic display. The dynamic display is built with static display as background. The dynamic information of these displays is updated generally once in every three seconds.

The display can be grouped under the following four categories:

(i) Standardised displays
(ii) Free configurable graphic displays (mimic displays)
(iii) Alarm display
(iv) Curve display

(i) Standardised Displays

Standardised Displays are based on manufacturers standard way of organizing the display. These are organized hierarchically in three levels, shown in Fig. 21.81, such as

(a) Overview display level
(b) Group display level
(c) Loop display level

Overview Summary

The functional areas of the power stations, i.e. sections of the plant (boiler, turbine, DM plant, Fuel gas desulphurisation, etc.) are grouped together in the overview summary. In the overview summary there are a maximum of 12 text areas. To select an overview display a number 1–12 must be input via the operating keyboard. In the summary display behind each functional area name, three display characters for disturbance may be displayed such as

- C for disturbance in the closed loop control system
- O for disturbance in the open loop control system
- M for disturbance in the measurement system

(a) Overview Display Level Each functional area name in the summary display is assigned to an 'overview display'. Overview display can be structured into 24 groups, for example,

01 – FD fans,
02 – ID fans,
03 – PA fans,
04 – COAL MILLS, ...
05 – FUEL OIL PUMP ... etc.

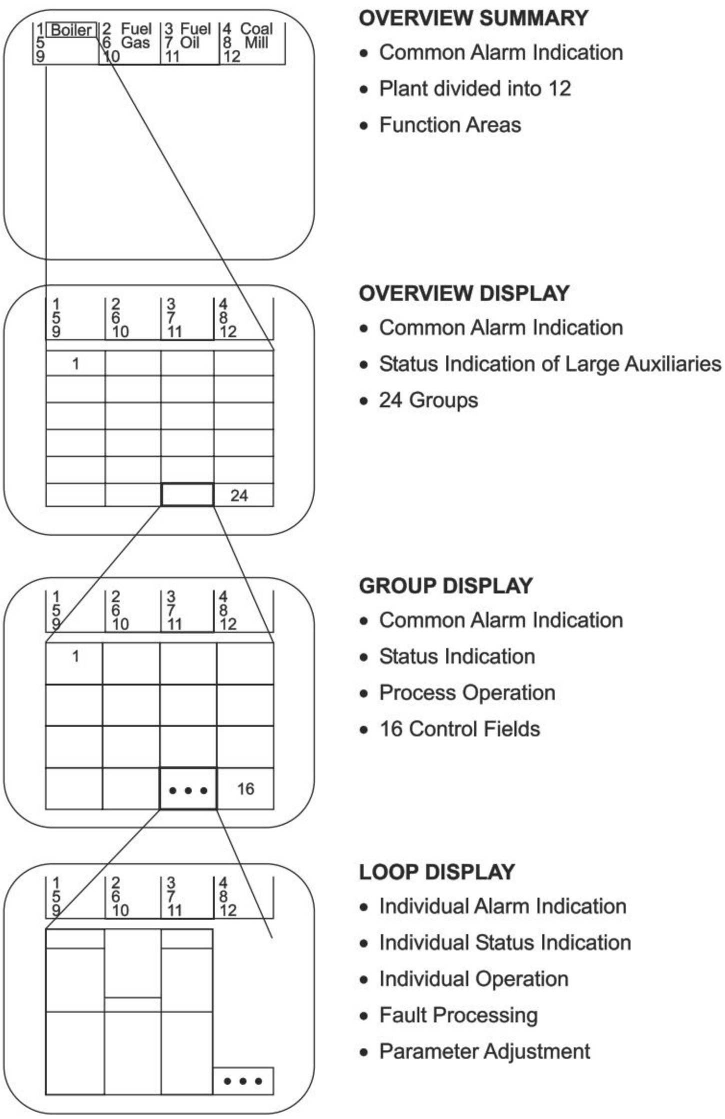


Fig. 21.81 Standard display hierarchy

To select a group display, a number corresponding to the group display (01–24) is selected via the keyboard. Each group in the overview display contains text describing the group. Behind each group name three characters—**C**, **O**, or **M**—may be displayed to indicate the disturbances, as already mentioned.

The group displays give an overview of the condition of a plant area. The operating condition of the main auxiliaries (i.e. FD fans, BFP's) is shown with differently coloured symbols.

For better understanding mimic displays may be included at this level.

(b) Group Display Level For every group in the overview display a group display is assigned. The group display level corresponds with the process operating level. At this level manual/auto station fields and indicators are displayed. The group display can have a maximum of 16 manual/auto stations or indicators. Mimic displays can be included at this level.

(c) Loop Display Level To each operating field of a group display, a loop display is assigned. The loop displays contain standard text (disturbance, status indications) for each type of operating field.

(ii) Free Configurable Graphic Displays

The display organization in free configurable graphic display shown in Fig. 21.82 is divided hierarchically into

(a) Plant or overview
(b) Graphic area display
(c) Graphic group display
(d) Sub-ordinate graphic group displays

Plant Overview In the plant overview, up to 12 area blocks (function areas in plain language) can be displayed. The plant overview is constantly displayed. The associated area display can be selected using the area numbers. All graphic displays can be selected via the display hierarchy.

(iii) Alarm Displays

Each alarm line contains the following information:

(a) Alarm identification and tag no
(b) Alarm class (alarm, warning, fault)
(c) Designation
(d) Setting
(e) Signal status (±)
(f) Alarm time

All alarms are displayed in the chronological order of their appearance. Each page contains 24 alarm lines. Current page is displayed and 64 old pages are stored.

(iv) Curve Displays

There are following four different types of display:

(a) ***Curve display*** In this type up to eight curves in each display can be shown as a function of time.
(b) ***Characteristics curve display*** In this type of display, the working point of the equipment is displayed on the static characteristics curves generated.
(c) ***Long time archive*** For a long term archive the measured value is stored and can be displayed in a 24 hour window.
(d) ***High speed recorder*** The measured values stored in the long term archive can be displayed in the high speed recorder, in an 8 minutes window, as currently entering data.

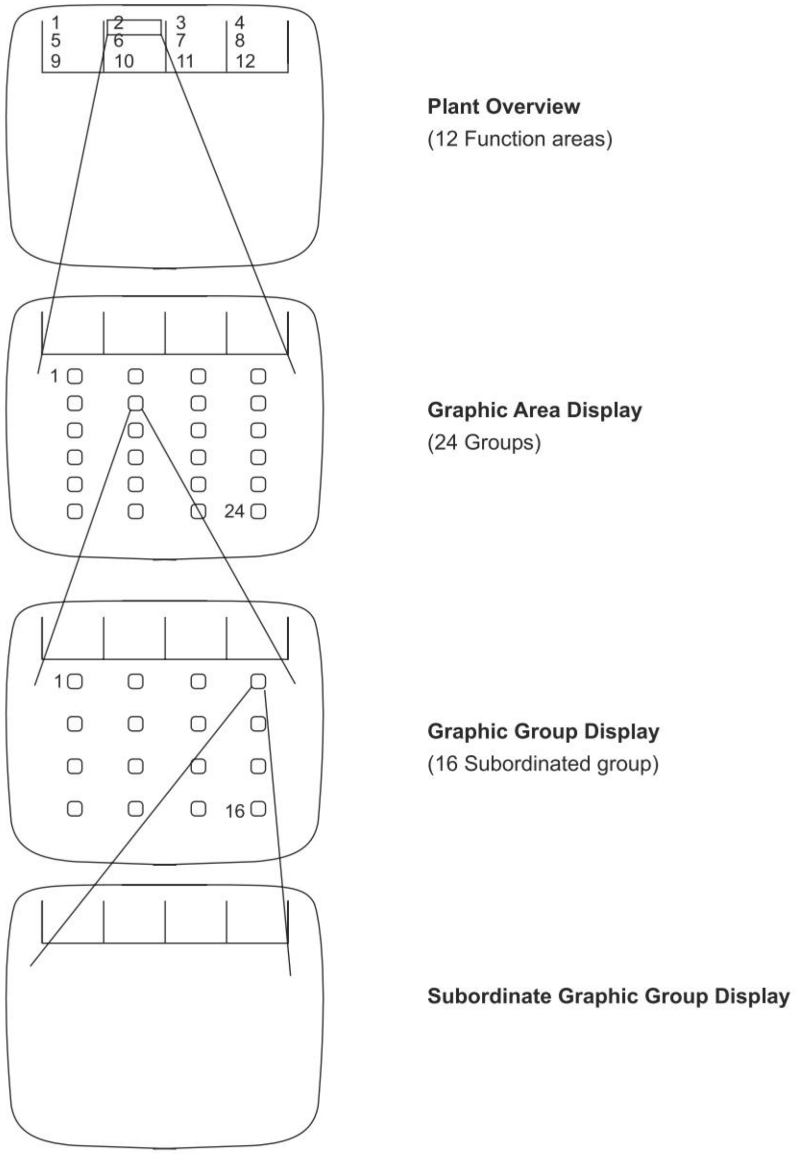


Fig. 21.82 Graphic display hierarchy

Review Questions

1. Define a control system.
2. Explain a basic controller.
3. Define errors.
4. Explain the concept ON–OFF control.
5. Explain proportional control.
6. Define offsets. How is this offset compensated?
7. Explain with diagram a basic controller configuration.

8. List classification of controller.
9. State the advantages and disadvantages of electronic controllers.
10. Define an electronic controller.
11. Explain, with diagram, electronic controller used as
    (i) Proportional control
    (ii) Integral control
    (iii) Derivative control
12. Explain, with circuit diagram, the following
    (i) PI control
    (ii) PD control
13. Explain with diagram the working of 3 term controller using Op Amps.
14. List features of an OpAmp.
15. Explain, with the help of a diagram, operation of a temperature controller.
16. Explain, with the help of a block diagram, the operation of a Digital Cotroller.
17. List the features of a Digital controller.
18. Explain, with the help of a block diagram, the operation of a Computer Supervisory control.
19. Differentiate between digital controller and computer supervisory control.
20. Explain the term UART.
21. Explain with help of block diagram the operation of a Cascade Process using Digital Controller.
22. Define PLC.
23. Explain with diagram the PLC structure.
24. Explain the basic operation of a PLC.
25. Define scan time and response time of a PLC.
26. Define a relay.
27. Explain relay diagram.
28. List and explain various types of timer and counter used in PLC.
29. Define a ladder diagram.
30. Compare relay wire diagram with ladder diagram.
31. List various inputs and outputs of a PLC.
32. Explain a program sequence executed by a PLC
33. List the various section of a PLC.
34. State the function of each sections used in PLC.
35. Explain fixed I/O PLCs.
36. Explain modular I/O PLCs.
37. Explain the principle of operations of a PLC.
38. Explain processor memory organisation.
39. Explain image table.
40. Explain program scan.
41. List the PLC hardware components.
42. Explain I/O section.
43. Explain discrete I/O module.
44. Explain with diagram the operation of a typical ac interface input module.
45. Explain with diagram the operation of a typical ac interface output module.
46. Explain the working of an analog I/O module.
47. List various I/O specifications.

Further Reading

1. Gray Dunning, *Introduction to PLC*, Delmar-Thomson Learning.
2. Ian. G. Warnock, *Programmable Controllers–operation and Application*, Prentice Hall.
3. Curtis Johnson, *Process Control Instrumentation Technology*, Prentice–Hall of India.
4. www. plc. net.

Answers to Objective Type Questions

CHAPTER 1

1. (d) 2. (a) 3. (c) 4. (a) 5. (d) 6. (d) 7. (b)
8. (a) 9. (a) 10. (b) 11. (c) 12. (a) 13. (a) 14. (a)
15. (a)

CHAPTER 2

1. (b) 2. (b) 3. (a) 4. (d) 5. (b) 6. (b) 7. (a)

CHAPTER 3

1. (b) 2. (b) 3. (b) 4. (d) 5. (b) 6. (c) 7. (b)
8. (b)

CHAPTER 4

1. (c) 2. (b) 3. (d) 4. (a) 5. (c) 6. (b) 7. (a)
8. (a) 9. (a) 10. (c) 11. (d) 12. (c) 13. (a) 14. (a)
15. (b) 16. (c)

CHAPTER 5

1. (a) 2. (b) 3. (a) 4. (c) 5. (a)

CHAPTER 6

1. (b) 2. (c) 3. (c) 4. (a) 5. (a)

CHAPTER 7

1. (d) 2. (b) 3. (b) 4. (a) 5. (d) 6. (b) 7. (a)
8. (c) 9. (c) 10. (b)

CHAPTER 8

1. (a)	2. (a)	3. (a)	4. (b)	5. (c)	6. (d)	7. (b)
8. (c)	9. (c)	10. (b)	11. (a)	12. (a)	13. (b)	14. (d)
15. (b)	16. (c)	17. (a)	18. (c)	19. (a)	20. (b)	

CHAPTER 9

1. (c)	2. (c)	3. (b)	4. (a)	5. (c)	6. (b)	7. (c)
8. (d)	9. (c)	10. (a)	11. (a)	12. (b)		

CHAPTER 10

1. (b)	2. (d)	3. (a)	4. (d)	5. (a)	6. (b)	7. (a)
8. (b)	9. (a)	10. (c)	11. (d)	12. (a)	13. (c)	14. (d)

CHAPTER 11

1. (c)	2. (d)	3. (d)	4. (c)	5. (a)	6. (a)	7. (c)
8. (b)	9. (b)	10. (d)	11. (a)	12. (d)	13. (a)	

CHAPTER 12

1. (a)	2. (a)	3. (b)	4. (d)	5. (c)	6. (d)	7. (d)
8. (d)	9. (c)	10. (b)	11. (b)	12. (a)		

CHAPTER 13

1. (b)	2. (a)	3. (a)	4. (c)	5. (c)	6. (b)	7. (c)
8. (d)	9. (d)	10. (a)	11. (a)	12. (b)	13. (b)	14. (a)
15. (c)	16. (d)					

CHAPTER 14

1. (b)	2. (b)	3. (c)	4. (b)	5. (a)	6. (c)	7. (c)
8. (a)	9. (b)	10. (b)				

CHAPTER 15

1. (d)	2. (b)	3. (d)	4. (b)	5. (a)	6. (b)	7. (a)
8. (b)	9. (a)	10. (b)	11. (d)	12. (c)	13. (d)	14. (b)
15. (a)	16. (b)					

CHAPTER 16

1. (a)	2. (a)	3. (b)	4. (b)	5. (c)	6. (c)	7. (d)
8. (b)	9. (c)	10. (a)	11. (c)	12. (a)		

CHAPTER 17

1. (a)	2. (c)	3. (c)	4. (a)	5. (b)	6. (a)	7. (b)
8. (a)	9. (a)	10. (c)	11. (d)	12. (c)	13. (a)	14. (b)

CHAPTER 18

1. (b)	2. (a)	3. (a)	4. (a)	5. (b)	6. (c)	7. (a)
8. (a)	9. (a)					

CHAPTER 19

1. (c)	2. (c)	3. (c)	4. (b)	5. (c)

CHAPTER 20

1. (b)	2. (a)	3. (a)	4. (c)	5. (b)	6. (c)	7. (d)
8. (d)						

Index

E

F

M

N

O

U

V

W

X

Y

Z