ELECTRONIC INSTRUMENTATION AND MEASUREMENT TECHNIQUES

PRENTICE-HALL OF INDIA PRIVATE LTD., *New Delhi*

ELECTRONIC INSTRUMENTATION AND MEASUREMENT TECHNIQUES

William David Cooper

Algonquin College of
Applied Arts and Technology
Ottawa, Ontario, Canada

Prentice-Hall of India Private Limited

New Delhi-110001

1958

PRENTICE-HALL INTERNATIONAL, INC., Englewood Cliffs.
PRENTICE-HALL OF INDIA PRIVATE LIMITED, New Delhi.
PRENTICE-HALL INTERNATIONAL, INC., London.
PRENTICE-HALL OF AUSTRALIA. PTY. LTD., Sydney.
PRENTICE-HALL OF CANADA LTD., Toronto.
PRENTICE-HALL OF JAPAN, INC., Tokyo.
PRENTICE-HALL OF SOUTHEAST ASIA (PTE.) LTD., Singapore.
WHITEHALL BOOKS LIMITED, Wellington, New Zealand.

Printed by G. D. Makhija at the India Offset Press, New Delhi and Published by Prentice-Hall of India Private Limited, M-97, Connaught Circus, New Delhi-110001.

PREFACE

Today's technician or technologist has the primary responsibility of using and maintaining electronic test equipment for research, design, testing, and troubleshooting. The areas requiring the use of test instruments are expanding very rapidly. In the past, electronic instruments were mainly used to determine the value of a physical quantity, such as voltage, current, resistance, frequency, and so on. The use of appropriate transducers now allows the measurement of many other physical parameters. In the industrial instrumentation area, for example, the control of production processes depends almost entirely on transducers as the primary input element of a measurement and control system. In medical electronics, transducers again play an important role for translating a patient's physiological variables into representative electrical signals which are then conditioned, displayed, and measured.

The increasing expansion of electronic instrumentation and measurement techniques is often the result of demands made by advances in other technological areas. Space engineering is a prime example of such a development. The relatively short period during which a satellite is boosted into orbit is controlled and monitored by a vast array of electronic test equipment, with information being collected, stored, and analyzed by high speed data acquisition systems.

Intelligent use of test equipment not only requires a good knowledge of the component, circuit, or system under measurement, but also an understanding of the capabilities and limitations of the test instrument itself. To make any meaningful and valid measurement, whether it involves the determination of the value of a single passive component or the analysis of the dynamic behavior of an entire system, the technologist must be able to find the answers to a few basic questions:

(a) What is the character and the range of the quantity or property to be measured?

(b) Which instrument, or combination of instruments, is capable of measuring this quantity?

(c) How can the measurement problem best be approached and which techniques may be used to perform the measurement?

(d) What degree of accuracy is required in the final test result and which factors affect the accuracy?

This textbook attempts to provide some of the answers to measurement problems by (a) developing the reader's understanding of the measurement principles involved in the determination of basic electrical parameters; (b) by developing his appreciation of the limitations and the capabilities of test equipment and measurement practices, in terms of the validity and accuracy of the result of his measurement; and (c) by discussing representative examples of electronic test instruments in the major areas of their application.

The text is not intended to describe and analyze every variation of instrument design, operation, or application. A selection of measurement topics and instruments had to be made; many test instruments, however, may include some of the basic components or concepts discussed in the text.

The standard instruments for the measurement of electrical parameters are discussed in the first 10 chapters, which include a brief treatment of the factors affecting the accuracy of the measurement. In many instances these basic instruments are discussed in considerable detail: their circuits are analyzed and the function of each component in the circuit is explained. Wherever possible, illustrative problems are integrated in the text to reinforce a description or an analysis.

The more advanced, laboratory-type instruments that are covered in the latter part of the book, are often complex in terms of the number and variety of circuits that make up the instrument. Many of these circuits, however, are standard and the student should either be familiar with the characteristics and operation of these basic circuits or he should be able to find the appropriate references for further study. The advanced instruments, therefore, are treated by the method of block diagram analysis.

While only some applications of electronic instruments in practical measurement situations are given, the number of applications for any given instrument is often limited only by the imagination and resourcefulness of the user. In a laboratory situation, a particular measurement setup may be found in the instrument manual supplied by the manufacturer. Sources of additional and very plentiful information are the technical and professional journals, manufacturer's application notes, and product engineering data.

Encouragement to proceed with the development of this book and substantial help in the preparation of the final manuscript was received from many quarters. The author wishes to extend his thanks to Mr. W. A. B. Saunders, principal of the Northern Alberta Institute of Technology at Edmonton, Alberta, for his interest and support early in the project; many valuable suggestions relating to subject matter and organization of the earlier chapters were made by the staff members of the Electronics Department of this institution. The author is deeply grateful to Dr. Irving L. Kosow, Dean of Evening Session at Staten Island Community College of the City University of New York, for his gentle guidance and expert advice during the preparation of the final manuscript; his meticulous review of both the initial and final drafts and the many detailed comments and suggestions for improvement have greatly influenced the shape of the final product. Acknowledgment is also extended to Mr. Matthew I. Fox and his staff at Prentice-Hall, for their enthusiasm, faith, and patience during the preparation of the book. The author wishes to thank the many manufacturers of electronic measuring equipment, who have so generously given their assistance in the form of photographs, circuit diagrams, and reprints of their company publications; their contributions are acknowledged throughout the text.

Writing a textbook is a very time consuming undertaking and can only be completed successfully by an almost total immersion in the subject matter. The author's wife Helen has provided this necessary "climate" through her constant encouragement and continued moral support, which is deeply appreciated.

W. D. COOPER

Ottawa, Ontario

CONTENTS

ELECTRONIC INSTRUMENTATION AND MEASUREMENT TECHNIQUES

ONE

MEASUREMENT AND ERROR

1-1 Introduction

Measurement generally involves using an *instrument* as a physical means of determining a quantity or variable. The instrument serves as an extension of human faculties and in many cases enables a man to determine the value of an unknown quantity which his unaided human faculties could not measure. An instrument, then, may be defined as *a device for determining the value or magnitude of a quantity or variable*. The *electronic* instrument, as its name implies, is based on electrical or electronic principles for its measurement function. An electronic instrument may be a relatively uncomplicated device of simple construction, for example, a basic dc current meter (see Chapter 4). As technology expands, however, the demand for more elaborate and more accurate instruments increases and produces new developments in instrument design and application. To use these instruments intelligently, one needs to understand their operating principles and to appraise their suitability for the intended application.

Measurement work employs a number of terms which should be defined here.

Instrument: a device for determining the value or magnitude of a quantity or variable.
Accuracy: closeness with which an instrument reading approaches the true value of the variable being measured.
Precision: a measure of the reproducibility of the measurements; i.e.,

given a fixed value of a variable, precision is a measure of the degree to which successive measurements differ from one another.
Sensitivity: the ratio of output signal or response of the instrument to a change of input or measured variable.
Resolution: the smallest change in measured value to which the instrument will respond.
Error: deviation from the true value of the measured variable.

Several techniques may be used to minimize the effects of errors. For example, in making precision measurements, it is advisable to record a series of observations rather than rely on one observation. Alternate methods of measurement, as well as use of different instruments to perform the same experiment, provide a good technique for increasing accuracy. Although these techniques tend to increase the *precision* of measurement by reducing environmental or random error, they cannot account for instrumental error.*

This chapter introduces the student to different types of errors in measurement and to the methods generally used to express errors, in terms of the most reliable value of the measured variable.

1-2 Accuracy and Precision

Accuracy refers to the degree of closeness or conformity to the true value of the quantity under measurement. *Precision* refers to the degree of agreement within a group of measurements or instruments.

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

Precision is composed of two characteristics: *conformity* and the number of *significant figures* to which a measurement may be made. Consider, for example, that a resistor, whose true resistance is 1,384,572 Ω, is measured by an ohmmeter. The ohmmeter consistently and repeatedly indicates the true value. But can the observer "read" this value from the scale? His estimates from the scale reading consistently yield a value of 1.4 megohms (MΩ). This is as

*Melville B. Stout, *Basic Electrical Measurements*, 2nd ed. (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960), pp. 21–26.

close to the true value as he can read the scale by estimation. Although there are no deviations from the observed value, the error created by the limitation of the scale reading is a *precision* error.

The preceding example involving a consistent reading of 1.4 MΩ illustrates that conformity is a necessary, but not sufficient, condition for precision because of the lack of significant figures obtained. Similarly, precision is a necessary, but not sufficient, condition for accuracy.

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

In critical work, good practice dictates an independent set of measurements, using different instruments or different measurement techniques, not subject to the same systematic errors. The observer must take those steps insuring proper instrumental function and calibration against a known standard so that no outside influence affects the accuracy of his measurements.

1-3 Significant Figures

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

For example, if a resistor is specified as having a resistance of 106 Ω, its resistance should be closer to 106 Ω than to 107 Ω or 105 Ω. If the value of the resistor is described as 106.0 Ω, it means that its resistance is closer to 106.0 Ω than it is to 106.1 Ω or 105.9 Ω. In 106 Ω there are three significant figures, in 106.0 Ω, there are four. The latter, with more significant figures, expresses a measurement of greater precision than the former.

Often, however, the total number of digits may not represent measurement precision. Frequently, large numbers with zeros before a decimal point are used for approximate populations or amounts of money. For example, the population of a city is reported in six figures as 380,000. This may imply that the true value of the population lies between 379,999 and 380,001, which is six significant figures. What is meant, however, is that the population is closer to 380,000 than to 370,000 or 390,000. Since the population can be reported only to two significant figures, how can large numbers be expressed?

A more technically correct notation uses *powers of ten,* 38×10^4 or 3.8×10^5. This indicates that the population figure is only accurate to two significant figures. Uncertainty caused by zeros to the *left* of the decimal point is therefore usually resolved by *scientific notation* using powers of ten. Reference to the velocity of light as 186,000 mi/s, for example, would cause no misunderstanding to anyone with a technical background. But 1.86×10^5 mi/s leaves no confusion.

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

When a number of independent measurements are taken in an effort to obtain the best possible answer (closest to the true value), the result is usually expressed as the arithmetic *mean* of all the readings, with the range of possible error as the *largest deviation* from that mean. This is illustrated in Example 1-1.

Example 1-1: A set of independent voltage measurements taken by four observers was recorded as 117.02 V, 117.11 V, 117.08 V and 117.03 V. Calculate (a) the average voltage, (b) the range of error.

SOLUTION:

(a)
$$E_{av} = \frac{E_1 + E_2 + E_3 + E_4}{N}$$

$$= \frac{117.02 + 117.11 + 117.08 + 117.03}{4} = 117.06 \text{ V}$$

(b)
$$\text{Range} = E_{max} - E_{av} = 117.11 - 117.06 = 0.05 \text{ V}$$

but also
$$E_{av} - E_{min} = 117.06 - 117.02 = 0.04 \text{ V}$$

The average range of error therefore equals

$$\frac{0.05 + 0.04}{2} = \pm 0.045 = \pm 0.05 \text{ V}$$

When two or more measurements with different degrees of accuracy are added, *the result is only as accurate as the least accurate measurement*. Suppose that two resistances are *added* in series as in Example 1-2.

Example 1-2: Two resistors, R_1 and R_2, are connected in series. Individual resistance measurements, using a Wheatstone bridge, give $R_1 = 18.7\ \Omega$ and $R_2 = 3.624\ \Omega$. Calculate the total resistance to the appropriate number of significant figures.

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

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

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

The doubtful figures are written in *italics* to indicate that, in the addition of R_1 and R_2, the last three digits of the sum are doubtful figures. There is no value whatsoever in retaining the last two digits (the *2* and the *4*) because one of the resistances is accurate only to three significant figures or tenths of an ohm. The result should therefore also be reduced to three significant figures or the nearest tenth, i.e., $22.\mathit{3}\ \Omega$.

The number of significant figures in *multiplication* may increase rapidly, but again only the appropriate figures are retained in the answer, as shown in Example 1-3.

Example 1-3: In calculating voltage drop, a current of 3.18 A is recorded in a resistance of 35.68 Ω. Calculate the voltage drop across the resistor to the appropriate number of significant figures.

SOLUTION: $E = IR = (35.6\mathit{8}) \times (3.1\mathit{8}) = 11\mathit{3.4624} = 113$ V
Since there are three significant figures involved in the multiplication, the answer can be written only to a maximum of three significant figures.

In Example 1-3, the current, I, has three significant figures and R has four; and the result of the multiplication has only three significant figures. This illustrates that the answer cannot be known to an accuracy greater than the *least* poorly defined of the factors. Note also that if extra digits accumulate in the answer, they should be discarded or rounded off. In the usual practice, if the (least significant) digit in the first place to be discarded is less than five, it and the following digits are dropped from the answer. This was done in Example 1-3. If the digit in the first place to be discarded is five or greater, the previous digit is increased by one. For three-digit precision, therefore, 113.46 should be rounded off to 113; and 113.74 to 114.

Addition of figures with a range of doubt is illustrated in Example 1-4.

Example 1-4: Add 826 ± 5 to 628 ± 3.

SOLUTION: $N_1 = 826 \pm 5$ ($= \pm 0.605\%$)

$N_2 = 628 \pm 3$ ($= \pm 0.477\%$)

Sum $= 1{,}454 \pm 8$ ($= \pm 0.55\%$)

Note in Example 1-4 that the doubtful parts are *added,* since the ± sign means that one number may be high and the other low. The worst possible combination of range of doubt should be taken in the answer. The percentage doubt in the original figure N_1 and N_2 does not differ greatly from the percentage doubt in the final result.

If the same two numbers are *subtracted,* as in Example 1-5, there is an interesting comparison between addition and subtraction with respect to the range of doubt.

Example 1-5: Subtract 628 ± 3 from 826 ± 5 and express the range of doubt in the answer as a percentage.

SOLUTION: $N_1 = 826 \pm 5\ (= \pm 0.605\%)$

$N_2 = 628 \pm 3\ (= \pm 0.477\%)$

$\text{Difference} = 198 \pm 8\ (= \pm 4.04\%)$

Again, in Example 1-5, the doubtful parts are added for the same reasons as in Example 1-4. Comparing the results of addition and subtraction of the same numbers in Examples 1-4 and 1-5, note that the precision of the results, when expressed in *percentages,* differs greatly. The final result after subtraction shows a large increase in percentage doubt compared to the percentage doubt after addition. The percentage doubt increases even more when the difference between the numbers is relatively small. Consider the case illustrated in Example 1-6.

Example 1-6: Subtract 437 ± 4 from 462 ± 4 and express the range of doubt in the answer as a percentage.

SOLUTION: $N_1 = 462 \pm 4\ (= \pm 0.87\%)$

$N_2 = 437 \pm 4\ (= \pm 0.92\%)$

$\text{Difference} = 25 \pm 8\ (= \pm 32\%)$

Example 1-6 illustrates clearly that one should avoid measurement techniques depending on subtraction of experimental results because the ratio of error to final result may be greatly increased.

1-4 Types of Errors

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

Gross errors: largely, human errors, among them misreading of instruments, incorrect adjustment and improper application of instruments, and computational mistakes.
Systematic errors: shortcomings of the instruments, such as defective or worn parts, and effects of the environment on the equipment or the user.
Random errors: those due to causes which cannot be directly established because of random variations in the parameter or the system of measurement.

Each of the foregoing classes of errors will be discussed briefly and some methods will be suggested for their reduction or elimination.

a. Gross Errors. This class of errors mainly covers *human* mistakes in reading or using instruments and in recording and computing measurement results. As long as human beings are involved, some gross errors will inevitably be committed. Although complete elimination of gross errors is probably impossible, one should try to anticipate and correct them. Some gross errors are easily detected; others may be very elusive. One common gross error, frequently committed by beginners in measurement work, involves improper use of an instrument. In general, indicating instruments change conditions to some extent when connected into a complete circuit, so that the measured quantity is altered by the method employed. For example, a well-calibrated voltmeter may give a misleading voltage reading when connected across two points in a high resistance circuit (Example 1-7). The same voltmeter, when connected in a low-resistance circuit, may give a more dependable reading (Example 1-8). These examples illustrate that the voltmeter has a "loading effect" on the circuit, altering the original situation by the measurement process.

Example 1-7: A voltmeter, having a sensitivity of 1,000 Ω/V, reads 100 V on its 150-V scale when connected across an unknown resistor in series with a milliammeter.

When the milliameter reads 5 mA, calculate (a) apparent resistance of the unknown resistor, (b) actual resistance of the unknown resistor, (c) error due to the loading effect of the voltmeter.

SOLUTION: (a) The total circuit resistance equals

$$R_T = \frac{V_T}{I_T} = \frac{100\ \text{V}}{5\ \text{mA}} = 20\ \text{k}\Omega$$

Neglecting the resistance of the milliameter, the value of the unknown resistor is $R_X = 20\ \text{k}\Omega$.

(b) The voltmeter resistance equals

$$R_T = 1000\,\frac{\Omega}{\text{V}} \times 150\ \text{V} = 150\ \text{k}\Omega$$

Since the voltmeter is in parallel with the unknown resistance, we can write

$$R_X = \frac{R_T R_V}{R_V - R_T} = \frac{20 \times 150}{130} = 23.05\ \text{k}\Omega$$

(c) $$\%\ \text{error} = \frac{\text{actual} - \text{apparent}}{\text{actual}} \times 100\% = \frac{23.05 - 20}{23.05} \times 100\%$$

$$= 13.23\%$$

Example 1-8: Repeat Example 1-7 if the milliammeter reads 800 mA and the voltmeter reads 40 V on its 150-V scale.

SOLUTION:

(a) $$R_T = \frac{V_T}{I_T} = \frac{40\ \text{V}}{0.8\ \text{A}} = 50\ \Omega$$

(b) $$R_V = 1{,}000\,\frac{\Omega}{\text{V}} \times 150\ \text{V} = 150\ \text{k}\Omega$$

$$R_X = \frac{R_T R_V}{R_V - R_T} = \frac{50 \times 150}{149.95} = 50.1\ \Omega$$

(c) $$\%\ \text{error} = \frac{50.1 - 50}{50.1} \times 100\% = 0.2\%$$

Errors caused by this loading effect of the meter can be avoided by using it intelligently. For example, do not use a low-resistance voltmeter to measure the grid voltage in a vacuum tube amplifier. In this particular measurement, a high-input impedance voltmeter (or vacuum tube voltmeter) is required.

A large number of gross errors may be attributed to carelessness or bad habits, such as improper reading of an instrument, recording the result differently from the actual reading taken, or adjusting the instrument incorrectly. Consider the case where a multirange voltmeter uses a single set of scale markings with different number designations for the various voltage ranges. It is quite easy to use a scale which does not correspond to the setting of the range control of the voltmeter. A gross error may also occur when the instrument is not set to zero before the measurement is taken; then all the readings are off.

Errors of this type cannot be treated mathematically. They can be avoided only by taking great care in reading and recording the measurement data. Good practice requires making more than one reading of the same quantity, preferably by a different observer. Never place complete dependence on one reading but take at least three separate readings, preferably under conditions in which instruments are switched off-on.

b. Systematic Errors. This type of error is usually divided into two different categories: (1) instrumental errors, defined as shortcomings of the instrument; (2) environmental errors, due to external conditions affecting the measurement.

Instrumental errors are errors inherent in measuring instruments because of their mechanical structure, for example, in the d'Arsonval movement friction in bearings of various moving components may cause incorrect readings. Irregular spring tension, stretching of the spring, or reduction in tension due to improper handling or overloading of the instrument, will result in errors. Other instrumental errors are calibration errors, causing the instrument to read high or low along its entire scale. (Failure to set the instrument to zero before making a measurement has a similar effect.)

There are many kinds of instrumental errors, depending on the type of instrument used. The experimenter should always take precautions to insure that the instrument he is using is operating properly and does not contribute excessive errors for the purpose at hand. Faults in instruments may be detected by checking for erratic behavior and stability and reproducibility of results. A quick and easy way to check an instrument is to compare it to another with the same characteristics or to one that is known to be more accurate.

Instrumental errors may be avoided by (a) selecting a suitable instrument for the particular measurement application; (b) applying correction factors after determining the amount of instrumental error; (c) calibrating the instrument against a standard.

Environmental errors are due to conditions external to the measuring device, including conditions in the area surrounding the instrument, for example, the effects of changes in temperature, humidity, barometric pressure, or of magnetic or electrostatic fields. Thus, a change in ambient temperature at which the instrument is used causes a change in the elastic properties of the spring in a moving-coil mechanism and so affects the reading of the instrument. Corrective measures to reduce these effects include air conditioning, hermetically sealing certain components in the instrument, use of magnetic shields, and the like.

Systematic errors can also be subdivided into *static* or *dynamic* errors. Static errors are caused by limitations of the measuring device or the physical laws governing its behavior. A static error is introduced in a micrometer when excessive pressure is applied in torquing the shaft. Dynamic errors are caused by the instrument's not responding fast enough to follow the changes in a measured variable.

c. Random Errors. These errors are due to unknown causes and occur even when all systematic errors have been accounted for. In well-designed experiments, few random errors usually occur, but they become important in high-accuracy work. Suppose a voltage is being monitored by a voltmeter which is read at half-hour intervals. Although the instrument is operated under ideal environmental conditions and has been accurately calibrated before the measurement, it will be found that the readings vary slightly over the period of observation. This variation cannot be corrected by any method of calibration or other known method of control and it cannot be explained without minute investigation. The only way to offset these errors is by increasing the number of readings and using statistical means in order to obtain the best approximation of the true value of the quantity under measurement.

1-5 Statistical Analysis

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

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

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_n}{n} = \frac{\sum x}{n} \tag{1-1}$$

where $\bar{x}$ = arithmetic mean

x_1, x_2, x_n = readings taken

n = number of readings

(Example 1-1 showed how the arithmetic mean is used.)

b. Deviations. Deviation is the departure of a given reading from the arithmetic mean of the group of readings. If the deviation of the first reading,

x_1, is called d_1, and that of the second reading, x_2, is called d_2, and so on, then the deviations from the mean can be expressed as

$$d_1 = x_1 - \bar{x} \qquad d_2 = x_2 - \bar{x} \qquad d_n = x_n - \bar{x} \tag{1-2}$$

Note that the deviation from the mean may have a positive or a negative value and that the algebraic sum of all the deviations must be zero.

Example 1-9 illustrates the computation of deviations:

Example 1-9: A set of independent current measurements were taken by six observers and were recorded as 12.8 mA, 12.2 mA, 12.5 mA, 13.1 mA, 12.9 mA, and 12.4 mA. Calculate (a) the arithmetic mean, (b) the deviations from the mean.

SOLUTION:

(a) Using Eq. (1-1), the arithmetic mean equals

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

(b) Using Eq. (1-2), the deviations are

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

Note that the algebraic sum of all the deviations equals zero.

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

$$D = \frac{|d_1| + |d_2| + |d_3| + \cdots + |d_n|}{n} = \frac{\sum |d|}{n} \tag{1-3}$$

Example 1-10 shows how average deviation is calculated.

Example 1-10: Calculate the average deviation for the data given in Example 1-9.

SOLUTION:

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

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

$$\sigma = \sqrt{\frac{d_1^2 + d_2^2 + d_3^2 + \ldots + d_n^2}{n}} = \sqrt{\frac{\sum d_i^2}{n}} \tag{1-4}$$

In practice, of course, the possible number of observations is finite. The standard deviation of a *finite* number of data is given by

$$\sigma = \sqrt{\frac{d_1^2 + d_2^2 + d_3^2 + \ldots + d_n^2}{n-1}} = \sqrt{\frac{\sum d_i^2}{n-1}} \tag{1-5}$$

[Equation (1-5) will be used in Example 1-11.]

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

$$\text{variance } (V) = \text{mean square deviation} = \sigma^2$$

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

1-6. Probability of Errors and the Gaussian Error Curve

One method of presenting test results is in the form of a *histogram* or block diagram. The technique is illustrated in Fig. 1-1, representing the data given in Table 1-1 which shows a set of fifty readings of a voltage measurement. The nominal value of the voltage is 100 V and the data are taken and recorded to the nearest 0.1 V. The histogram of Fig. 1-1 represents these data

TABLE 1-1

TABULATION OF VOLTAGE READINGS

Voltage Reading (volts)	Number of Readings
99.7	1
99.8	4
99.9	12
100.0	19
100.1	10
100.2	3
100.3	1
Total	50

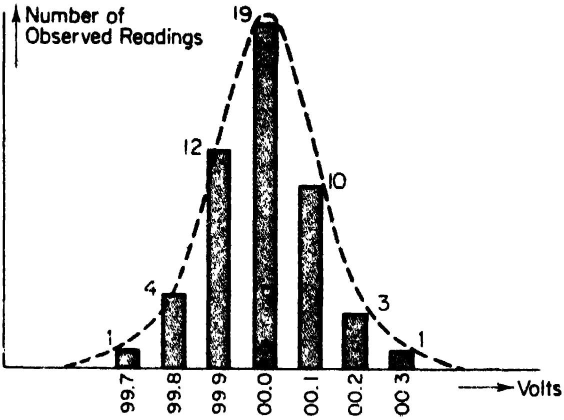


Figure 1-1
Histogram showing the frequency of occurrence of the 50 voltage readings of Table 1-1. The dashed curve represents the limiting case of the histogram when a large number of readings at small increments are taken.

where the ordinates indicate the number of observed readings of a particular value. At the central value of 100 V is a large group of readings, nineteen, with other values placed almost symmetrically on either side. If more readings were taken at smaller incremental steps, say, two hundred readings at 0.05-V intervals, the general form of the histogram would be about the same but the steps would be smaller and the curve smoother. With more and more data taken at smaller and smaller increments, the histogram would finally change into a smooth curve, as is indicated by the dashed line in Fig. 1-1. The smooth curve is used in most analytical studies of collected data.

The Normal or Gaussian law of error is the basis for the study of random effects. The equation for the Normal law may be written as

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2w^2} \tag{1-6}$$

where h = a constant
w = the magnitude of the deviation from the mean
y = the probability of occurrence of deviation w

Equation (1-6) leads to a curve of the type shown in Fig. 1-2.

Another, more convenient form of the equation describing the Gaussian curve uses the standard deviation σ and is given by

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-w^2/2\sigma^2} \tag{1-7}$$

The form of Eq. (1-7) is particularly useful because σ is usually the known quantity of interest. In Fig. 1-2, the deviations w from the mean value, w_0, are divided in terms of σ units, so that $w = 1\sigma, 2\sigma, 3\sigma$, etc. The Gaussian curve of Fig. 1-2 may represent the error distribution for any set of data obeying the Normal law and may be used as such once the value of σ has been determined.

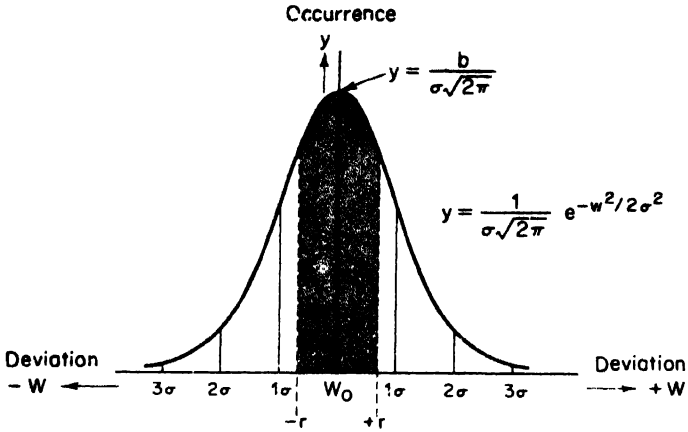


Figure 1-2
Curve for the normal law. The shaded portion indicates the region of probable error, where $r = \pm 0.6745\,\sigma$.

The sharper and narrower the error distribution curve of Fig. 1-2, the more definitely an observer may state that the most probable value of the *true* reading is the *mean* reading. In this symmetrical type of distribution, the most probable value is synonymous with the mean value w_0. The Normal curve of Fig. 1-2 may be regarded as the limiting form of the histogram of Fig. 1-1, where the most probable value of the true voltage is the mean value of 100 V.

A study of the Normal distribution curve leads to the following qualitative statements:*

(a) All observations include small disturbing effects, called *random errors*.
(b) Random errors may be both positive and negative.
(c) There is an equal possibility of positive and negative random errors.

The possibilities as to the form of the error distribution curve can then be stated as follows:

(a) There will be a strong central tendency; i.e., small random errors are more probable than large ones.
(b) Large errors are very improbable.
(c) Because of equal probability of plus and minus component errors,

*Stout, *op. cit.*, pp. 38*ff.*

the curve of probability of a given error will be symmetrical about the zero value.

The standard deviation σ, defined in Eq. (1-4) as the root-mean-square value of all the deviations, is a quantity of the same units as the observed quantity and its mean. The significance of the standard deviation is shown in the following discussion:

The area under the Gaussian probability curve of Fig. 1-2, between the limits $+\infty$ and $-\infty$, represents the entire number of observations. The area under the curve between the $+\sigma$ and $-\sigma$ limits represents the cases that differ from the mean by no more than the standard deviation. Integration of the area under the curve within the $\pm\sigma$ limits gives the total number of cases within these limits. For normally dispersed data, following the Gaussian distribution, approximately 68 per cent of all the cases lie between the limits of $+1\sigma$ and -1σ from the mean. Corresponding values of other deviations, expressed in terms of σ, are given in Table 1-2.

TABLE 1-2

AREA UNDER THE PROBABILITY CURVE

Deviation ($\pm$) (σ)	Fraction of Total Area Included
0.675	0.5000
1.0	0.6828
2.0	0.9546
3.0	0.9972

If, for example, a large number of nominally 100-Ω resistors is measured and the mean value is found to be 100.00 Ω, with a standard deviation (S.D.) of 0.20 Ω, we know that on the average 68 per cent (or roughly two-thirds) of all the resistors have values which lie between limits of $\pm 0.20\ \Omega$ of the mean. There is then approximately a 2/1 chance that any resistor, selected from the lot at random, will lie within these limits. If larger odds are required, the deviation may be extended to a limit of $\pm 2\sigma$, in this case $\pm 0.40\ \Omega$. According to Table 1-2, this now includes 95 per cent of all the cases, giving 19/1 odds that any resistor selected at random lies within $\pm 0.40\ \Omega$ of the mean value of 100.00 Ω.

Table 1-2 also shows that half of the cases are included in the deviation limits of $\pm 0.6745\ \sigma$. The quantity r is called the *probable error* and is defined as

$$\text{probable error } r = \pm 0.6745\ \sigma \tag{1-8}$$

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

An illustration of the calculation of random errors is given in Example 1-11.

Example 1-11: Ten measurements of the resistance of a resistor gave 101.2, 101.7, 101.3, 101.0, 101.5, 101.3, 101.2, 101.4, 101.3, and 101.1 Ω.

Assume that only random errors are present. Calculate (a) the arithmetic mean, (b) the standard deviation of the readings, (c) the average error in per cent, (d) the probable error.

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

x	d	d^2
101.2	−0.1	0.01
101.7	0.4	0.16
101.3	0.0	0.00
101.0	−0.3	0.09
101.5	0.2	0.04
101.3	0.0	0.00
101.2	−0.1	0.01
101.4	0.1	0.01
101.3	0.0	0.00
101.1	−0.2	0.04
$\sum x = 1{,}013.0$	$\sum \lvert d \rvert = 1.4$	$\sum d^2 = 0.36$

(a) Arithmetic mean, $\bar{x} = \dfrac{\sum x}{n} = \dfrac{1{,}013.0}{10} = 101.3\ \Omega$

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

(c) Average error, $D = \dfrac{\sum d}{n} = \dfrac{1.4}{10} = 0.14\ \Omega$

Percentage error $= \dfrac{D}{\bar{x}} = \dfrac{0.14}{101.3} \times 100\% = 0.138\%$

(d) Probable error $= 0.6745\sigma = 0.6745 \times 0.2 = 0.1349\ \Omega$

1-7 Limiting Errors

In most indicating instruments, the accuracy is guaranteed to a certain percentage of full-scale reading. Circuit components (such as capacitors, resistors, etc.) are guaranteed within a certain percentage of their rated value. The limits of these deviations from the specified values are known as *limiting errors* or *guarantee errors*. For example, if the resistance of a resistor is given as 500 Ω $\pm$ 10 per cent, this means that the manufacturer guarantees that the resistance falls between the limits 450 Ω and 550 Ω. The maker is not specifying a standard deviation or a probable error, but promises that the error is no greater than the limits set.

Example 1-12: A 0–150-V voltmeter has a guaranteed accuracy of 1 per cent full-scale reading. The voltage measured by this instrument is 83 V. Calculate the limiting error in per cent.

SOLUTION: The magnitude of the limiting error is

$$0.01 \times 150\ \text{V} = 1.5\ \text{V}$$

The percentage error at a meter indication of 83 V is

$$\frac{1.5}{83} \times 100 \text{ per cent} = 1.81 \text{ per cent.}$$

It is important to note in Example 1-12 that a meter is guaranteed to have an accuracy of better than 1 per cent of the full-scale reading, but when the meter reads 83 V the limiting error increases to 1.81 per cent. Correspondingly, when a smaller voltage is measured, the limiting error will increase further. If the meter reads 60 V, the per cent limiting error is $1.5/60 \times 100 = 2.5$ per cent; if the meter reads 30 V, the limiting error is $1.5/30 \times 100 = 5$ per cent. The increase in per cent limiting error, as smaller voltages are measured, occurs because the magnitude of the limiting error is a fixed quantity based on the full-scale reading of the meter. The voltages measured may range from 0 V to 150 V. Example 1-12 shows the importance of taking measurements *as close to full scale as possible.*

Measurements or computations, *combining* guarantee errors, are often made. Example 1-13 illustrates such a computation.

Example 1-13: Three decade boxes, each guaranteed to ± 0.1 per cent are used in a Wheatstone bridge to measure the resistance of an unknown resistor R_x. Calculate the limits on R_x, imposed by the decade boxes.

SOLUTION: The equation for bridge balance shows that R_x can be determined in terms of the resistances of the three decade boxes and $R_x = R_1R_2/R_3$, where R_1, R_2, and R_3 are the resistances of the decade boxes, guaranteed to ± 0.1 per cent. One must recognize that the two terms in the numerator may both be positive to the full limit of 0.1 per cent and the denominator may be negative to the full 0.1 per cent, giving a resultant error of 0.3 per cent. The guarantee error is thus obtained by taking the *direct sum* of all the possible errors, adopting the algebraic signs which give the worst possible combination.

As a further example, consider computing the power dissipation in a resistor, using the relationship $P = I^2R$, as shown in Example 1-14.

Example 1-14: The current passing through a resistor of $100 \pm 0.2\ \Omega$ is 2.00 ± 0.01 A. Calculate the limiting error in the computed value of power dissipation, using the relationship $P = I^2R$.

SOLUTION: Expressing the guaranteed limits of both current and resistance in percentages instead of units, we obtain

$$I = 2.00 \pm 0.01 \text{ A} = 2.00 \text{ A} \pm 0.5\%$$

$$R = 100 \pm 0.2\ \Omega = 100\ \Omega \pm 0.2\%$$

Using the worst possible combination of errors, the limiting error in the power dissipation is ($P = I^2R$):

$$(2 \times 0.5\%) + 0.2\% = 1.2\%.$$

Power dissipation should then be written as follows:

$$P = I^2R = (2.00)^2 \times 100 = 400 \text{ W} \pm 1.2\% = 400 \pm 4.8 \text{ W}$$

References

1. Bartholomew, Davis, *Electrical Measurements and Instrumentation,* chaps. 1, 2. Boston: Allyn and Bacon, Inc., 1963.

2. Frank, Ernest, *Electrical Measurement Analysis,* chap. 14. New York: McGraw-Hill Book Company, Inc., 1959.

3. Stout, Melville B., *Basic Electrical Measurements,* 2nd ed., chap. 2. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

4. Young, Hugh D., *Statistical Treatment of Experimental Data.* New York: McGraw-Hill Book Company, Inc., 1962.

Questions

1. What is the difference between accuracy and precision?

2. List four sources of possible errors in instruments.

3. What are the three general classes of errors?

4. Define (a) instrumental error, (b) limiting error, (c) calibration error, (d) environmental error, (e) random error, (f) probable error.

Problems

1. A 0–100-V voltmeter has two hundred scale divisions which can be read to ½ division. Determine the resolution of the meter in volts.

2. A digital voltmeter has a read-out range from 0 to 9,999 counts. Determine the resolution of the instrument in volts when the full-scale reading is 9.999 V.

3. State the number of significant figures in each of the following: (a) 542, (b) 0.65, (c) 27.25, (d) 0.00005, (e) 40×10^6, (f) 20,000.

4. Four resistors are placed in series. The values of the resistors are 28.4 Ω, 4.25 Ω, 56.605 Ω, 0.75 Ω, with an uncertainty of one unit in the last digit of each number. Calculate the total series resistance, giving only significant figures in the answer.

5. A voltage drop of 112.5 V is measured across a resistor passing a current of 1.62 A. Calculate the power dissipation of the resistor, giving only significant figures in the answer.

6. A voltmeter, having a sensitivity of 10 kΩ/V, reads 75 V on its 100-V scale when connected across an unknown resistor. When the current through the resistor is 1.5 mA, calculate (a) the apparent resistance of the unknown resistor, (b) the actual resistance of the unknown resistor, (c) the percentage error due to the loading effect of the voltmeter.

7. The voltage across a resistor is 200 V, with a probable error of ± 2 per cent, and the resistance is 42 Ω with a probable error of ± 1.5 per cent. Calculate (a) the power dissipated in the resistor, (b) the percentage error in the answer.

8. The following values were obtained from the measurements of the value of a resistor: 147.2 Ω, 147.4 Ω, 147.9 Ω, 148.1 Ω, 147.1 Ω, 147.5 Ω, 147.6 Ω, 147.4 Ω, 147.6 Ω, and 147.5 Ω. Calculate (a) the arithmetic mean, (b) the average deviation, (c) the standard deviation, (d) the probable error of the average of the ten readings.

9. Six determinations of a quantity, as entered on the data sheet and presented to you for analysis, are 12.35, 12.71, 12.48, 10.24, 12.63, and 12.58. Examine the data and on the basis of your conclusions, calculate (a) the arithmetic mean, (b) the standard deviation, (c) the probable error in per cent of the average of the readings.

10. Two resistors have the following ratings:

$$R_1 = 36\,\Omega \pm 5\% \quad \text{and} \quad R_2 = 75\,\Omega \pm 5\%.$$

Calculate (a) the magnitude of error in each resistor, (b) the limiting error in ohms and in per cent when the resistors are connected in series, (c) the limiting error in ohms and in per cent when the resistors are connected in parallel.

11. The resistance of an unknown resistor is determined by the Wheatstone bridge method. The solution for the unknown resistance is stated as $R_x = R_1 R_2 / R_3$,

where $R_1 = 500\ \Omega \pm 1\%$
$R_2 = 615\ \Omega \pm 1\%$
$R_3 = 100\ \Omega \pm 0.5\%$

Calculate (a) the nominal value of the unknown resistor, (b) the limiting error in ohms of the unknown resistor, (c) the limiting error in per cent of the unknown resistor.

12. A resistor is measured by the voltmeter-ammeter method. The voltmeter reading is 123.4 V on the 250-V scale and the ammeter reading is 283.5 mA on the 500-mA scale. Both meters are guaranteed to be accurate within ± 1 per cent of full-scale reading. Calculate (a) the indicated value of the resistance, (b) the limits within which you can guarantee the result.

13. In a dc circuit, the voltage across a component is 64.3 V and the current is 2.53 A. Both current and voltage are given with an uncertainty of one unit in the last place. Calculate the power dissipation to the appropriate number of significant figures.

14. A power transformer was tested to determine losses and efficiency. The input power was measured as 3,650 W and the delivered output power was 3,385 W, with each reading in doubt by ± 10 W. Calculate (a) the percentage uncertainty in the losses of the transformer, (b) the percentage uncertainty in the efficiency of the transformer, as determined by the difference in input and output power readings.

15. The power factor and phase angle in a circuit carrying a sinusoidal current are determined by measurements of current, voltage, and power. The current is read as 2.50 A on a 5-A ammeter, the voltage as 115 V on a 250-V voltmeter and the power as 220 W on a 500-W wattmeter. The ammeter and voltmeter are guaranteed accurate to within $\pm$ 0.5 per cent of full-scale indication and the wattmeter to within ± 1 per cent of full-scale reading. Calculate (a) the percentage accuracy to which the power factor can be guaranteed, (b) the possible error in the phase angle.

16. The arms of a Wheatstone bridge, marked in order around the bridge, are B, A, X, and R. The three known arms have the following constants:

$$A = 840\ \Omega\ (\text{S.D.} = 1\ \Omega)$$
$$B = 90\ \Omega\ (\text{S.D.} = 0.5\ \Omega)$$
$$R = 250\ \Omega\ (\text{S.D.} = 1\ \Omega)$$

Calculate (a) the probable value of X, (b) the standard deviation of X.

TWO

SYSTEMS OF UNITS OF MEASUREMENT

2-1 Fundamental and Derived Units

To specify and perform calculations with physical quantities, they must be defined both in *kind* and *magnitude*. The standard measure of each kind of physical quantity is the *unit*; the number of times the unit occurs in any given amount of the same quantity is the *number of measure*. For example, when we speak of a distance of 100 meters, we know that the meter is the unit of length and that the number of units of length is one hundred. The physical quantity, *length*, is therefore defined by the unit, *meter*. Without the unit, the number of measure has no physical meaning.

In science and engineering, two kinds of units are used: *fundamental units* and *derived units*. The fundamental units in mechanics are measures of *length, mass,* and *time*. The sizes of the fundamental units, whether foot or meter, pound or kilogram, second or hour, are quite arbitrary and can be selected to fit a certain set of circumstances. Since length, mass, and time are fundamental to most other physical quantities besides those in mechanics, they are called the *primary* fundamental units. Measures of certain physical quantities in the thermal, electrical, and illumination disciplines are also represented by fundamental units. These units are used only where these particular classes are involved, and they may therefore be defined as *auxiliary* fundamental units.

All other units which can be expressed in terms of the fundamental units are called *derived* units. Every derived unit originates from some physical law defining that unit. For example, the area, A, of a rectangle is proportional to its length (l) and breadth (b), or $A = lb$. If the meter has been chosen as the unit of length, then the area of a rectangle of 3 m by 4 m is 12 m^2. Note that the numbers of measure are multiplied ($3 \times 4 = 12$) as well as the units ($m \times m = m^2$). The derived unit for area (A) is then the square meter (m^2).

A derived unit is recognized by its ***dimensions***, which can be defined as the complete algebraic formula for the derived unit. The dimensional ***symbols*** for the fundamental units of length, mass, and time are L, M, and T, respectively. The dimensional symbol for the derived unit of area is L^2 and that for volume, L^3. The dimensional symbol for the unit of force is LMT^{-2}, which follows from the defining equation for force. The dimensional formulas of the derived units are particularly useful for converting units from one system to another, as is shown in Sec. 2-6.

For convenience, some derived units have been given new names. For example, the derived unit of force in the SI units is called the newton (N), instead of the dimensionally correct name kg m/s^2.

2-2 Systems of Units

In 1790, the French government issued a directive to the French Academy of Sciences to study and to submit proposals for a single system of weights and measures to replace all other existing systems. The French scientists decided, as a first principle, that a ***universal system*** of weights and measures should not depend on man-made reference standards, but instead be based on permanent measures provided by nature. As the ***unit of length***, therefore, they chose the ***meter***, defined as the ten-millionth part of the distance from the pole to the equator along the meridian passing through Paris. As the ***unit of mass*** they chose the mass of a cubic centimeter of distilled water at 4°C and normal atmospheric pressure (760 mm Hg) and gave it the name ***gram***. As the third unit, the ***unit of time,*** they decided to retain the traditional second, simply defining it as 1/86,400 of the mean solar day.

As a second principle, they decided that all other units should be ***derived*** from the aforementioned three ***fundamental units*** of length, mass, and time. Next—the third principle—they proposed that all multiples and submultiples of basic units be in the ***decimal system,*** and they devised the system of prefixes in use today. (Table 2-1 lists the decimal multiples and submultiples.)

The proposals of the French Academy were approved and introduced as the ***metric system*** of units in France in 1795. The metric system aroused considerable interest elsewhere and finally, in 1875, seventeen countries signed the so-called Metre Convention, making the metric system of units the legal system. Britain and the United States, although signatories of the convention, recog-

TABLE 2-1
DECIMAL MULTIPLES AND SUBMULTIPLES

Name	Symbol	Equivalent
tera	T	10^{12}
giga	G	10^{9}
mega	M	10^{6}
kilo	k	10^{3}
hecto	h	10^{2}
deca	da	10
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

nized its legality only in international transactions but did not accept the metric system for their own domestic use.

Britain, in the meantime, had been working on a system of electrical units, and the British Association for the Advancement of Science decided on the centimeter and the gram as the fundamental units of length and mass. From this developed the *centimeter-gram-second* or *CGS absolute system* of units, used by physicists all over the world. Complications arose when the CGS system was extended to electric and magnetic measurements because of the need to introduce at least one more unit in the system. In fact, two parallel systems were established. In the CGS *electrostatic system,* the unit of electric charge was derived from the centimeter, gram, and second by assigning the value 1 to the permittivity of free space in Coulomb's law for the force between electric charges. In the CGS *electromagnetic system,* the basic units are the same and the unit of magnetic pole strength is derived from them by assigning the value 1 to the permeability of free space in the inverse square formula for the force between magnetic poles.

The *derived* units for electric current and electric potential in the electromagnetic system, the ampere and the volt, are used in practical measurements. These two units, and the corresponding ones, such as the coulomb, ohm, henry, farad, etc., were incorporated in a third system, called the *practical system.* Further simplification in the establishment of a truly universal system came as the result of pioneer work by the Italian engineer Giorgi, who pointed out that the practical units of current, voltage, energy, and power, used by electrical engineers, were compatible with the meter-kilogram-second system. He suggested that the metric system be expanded into a *coherent* system of units by including the practical electrical units. The Giorgi system, adopted by many countries in 1935, came to be known as the MKSA system of units, where the ampere was selected as the fourth basic unit.

A more comprehensive system was adopted in 1954 and designated in 1960 by international agreement as the Système International d'Unités (SI). In the SI system, six basic units are used, namely, the meter, kilogram, second, and ampere of the MKSA system and, in addition, the degree Kelvin and the candela as the units of temperature and luminous intensity, respectively. The SI units are replacing other systems in science and technology; they have been adopted as the legal units in France, and will become obligatory in other metric countries.

The six basic quantities and units of measurement, with their unit symbols, in the SI system of units, are listed in Table 2-2.

TABLE 2-2

BASIC SI QUANTITIES, UNITS, AND SYMBOLS

Quantity	Unit	Symbol
Length	meter	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	degree Kelvin	°K
Luminous intensity	candela	cd

2-3 Electric and Magnetic Units

Before listing the SI units (sometimes called the *International* MKS system of units), a brief look at the origin of the electrical and magnetic units seems appropriate. The practical electrical and magnetic units with which we are familiar, such as the volt, ampere, ohm, henry, etc., were first derived in the CGS systems of units.

The *CGS electrostatic system* (CGSe) is based on Coulomb's experimentally derived law for the force between two electric charges. Coulomb's law states that

$$F = k\frac{Q_1Q_2}{r^2} \tag{2-1}$$

where F = the force between the two charges, expressed in CGSe units of force (g cm/s^2 = dyne)

k = a proportionality constant

$Q_{1,2}$ = electric charges, expressed in (derived) CGSe units of electric charge (statcoulomb)

r = the separation between the charges, expressed in the fundamental CGSe unit of length (centimeter)

Coulomb also found that the proportionality factor k depended on the medium, varying inversely as its permittivity ϵ. (Faraday called permittivity the *dielectric constant.*) Coulomb's law then takes the form

$$F = \frac{Q_1 Q_2}{\epsilon r^2} \tag{2-2}$$

Since ϵ is a numerical value depending only upon the medium, a value of 1 was assigned to the permittivity of free space, ϵ_0, thereby defining ϵ_0 as the *fourth fundamental unit* of the CGSe system. Coulomb's law then allowed the unit of electric charge Q to be determined in terms of these four fundamental units by the relation

$$\text{dyne} = \frac{\text{g cm}}{\text{s}^2} = \frac{Q^2}{(\epsilon_0 = 1)\ \text{cm}^2}$$

and therefore, dimensionally,

$$Q = \text{cm}^{3/2}\,\text{g}^{1/2}\,\text{s}^{-1} \tag{2-3}$$

The CGSe unit of electric charge was given the name *statcoulomb.*

The derived unit of electric charge in the CGSe system of units allowed other electrical units to be determined by their defining equations. For example, *electric current* (symbol I) is defined as the rate of flow of electric charge and is expressed as

$$I = \frac{Q}{t} \qquad \text{(statcoulomb/sec)} \tag{2-4}$$

The unit for electric current in the CGSe system was given the name *statampere.* Electric *fieldstrength, E, potential difference, V,* and *capacitance, C,* can similarly be derived from their defining equations.

The basis of the *CGS electromagnetic system* of units (CGSm) is Coulomb's experimentally determined law for the force between two magnetic poles, which states

$$F = k\frac{m_1 m_2}{r^2} \tag{2-5}$$

The proportionality factor, k, was found to depend on the medium in which the poles were placed, varying inversely with the magnetic *permeability* μ of the medium. The factor k was assigned the value 1 for the permeability of free space, μ_0, so that $k = 1/\mu_0 = 1$. This established the permeability of free space, μ_0, as the *fourth fundamental unit* of the CGSm system. The derived electromagnetic unit of pole strength was then defined in terms of these four fundamental units by the relation:

$$\text{dyne} = \frac{\text{g cm}}{\text{s}^2} = \frac{m^2}{(\mu_0 = 1)\ \text{cm}^2}$$

and therefore, dimensionally,

$$m = \text{cm}^{3/2}\,\text{g}^{1/2}\,\text{s}^{-1} \tag{2-6}$$

The derived unit of magnetic polestrength in the CGSm system led to the determination of other magnetic units, again by their defining equations. *Magnetic flux density* (symbol B), for example, is defined as the magnetic force per unit polestrength, where both force and polestrength are derived units in the CGSm system. Dimensionally, B is found to be equal to $cm^{-1/2}g^{1/2}s^{-1}$ (dyne-second/abcoulomb-centimeter) and is given the name *gauss*. Similarly, other magnetic units can be derived from defining equations and we find that the unit for *magnetic flux* (symbol Φ) is given the name *maxwell*; the unit for *magnetic fieldstrength* (symbol H), the name *oersted*; and the unit for *magnetic potential difference or magnetomotive force* (symbol U), the name *gilbert*.

The two CGS systems were linked together by Faraday's discovery that a moving magnet could induce an electric current in a conductor, and conversely, that electricity in motion could produce magnetic effects. Ampere's law of the magnetic field relates electric current (I) to magnetic fieldstrength (H),* quantitatively connecting the magnetic units in the CGSm system to the electric units in the CGSe system. The dimensions of the two systems did not quite agree and numerical conversion factors were introduced. The two systems finally formed one *practical system of electrical units* which was officially adopted by the International Electrical Congress.

These practical electrical units, derived from the CGSm system, were later defined in terms of so-called international units. It was thought at the time (1908), that the establishment of the practical units from the definitions of the CGS system would be too difficult for most laboratories and it was therefore decided (unfortunately) to define the practical units in a way which would make it fairly simple to establish them. The *ampere*, therefore, was defined in terms of the rate of deposition of silver from a silver nitrate solution by passing a current through that solution and the *ohm* as the resistance of a specified column of mercury. These units and those derived from them were called *international units*. As measurement techniques improved, it was found that small differences existed between the CGSm derived practical units and the international units, which were then specified as follows:

1 int. ohm = 1.00049 Ω (practical CGSm unit)
1 int. ampere = 0.99985 A
1 int. volt = 1.00034 V
1 int. coulomb = 0.99985 C
1 int. farad = 0.99951 F
1 int. henry = 1.00049 H
1 int. watt = 1.00019 W
1 int. joule = 1.00019 J

*See a textbook on electromagnetic theory.

Particulars of the electric and magnetic units, and their defining relationships, are given in Table 2-3. Multiplication factors for conversion into SI units are given in the columns headed CGSm and CGSe.

TABLE 2-3
ELECTRIC AND MAGNETIC UNITS

Quantity and Symbol	SI Unit			Conversion Factors	
	Name and Symbol		Defining Equation	CGSm	CGSe§
Electric current, I	ampere	A	$F_z = 10^{-7} I^2 \frac{dN^*}{dz}$	10	10/c
Electromotive force, E	volt	V	p† = IE	10^{-8}	10^{-8}c
Potential, V	volt	V	p† = IV	10^{-8}	10^{-8}c
Resistance, R	ohm	Ω	R = V/I	10^{-9}	10^{-9}c
Electric charge, Q	coulomb	C	Q = It	10	10/c
Capacitance, C	farad	F	C = Q/V	10^{9}	$10^{9}/c^2$
Electric field strength, E	—	V/m	E = V/1	10^{-6}	10^{-6}c
Electric flux density, D	—	C/m²	$D = Q/1^2$‡	10^{5}	10^{5}/c
Permittivity, ε	—	F/m	ε = D/E	—	$10^{11}/4\pi c^2$
Magnetic fieldstrength, H	—	A/m	$\oint$H dl = nI	$10^{3}/4$	—
Magnetic flux, Φ	weber	Wb	E = dΦ/dt	10^{-8}	—
Magnetic flux density, B	tesla	T	$B = \Phi/1^2$‡	10^{-4}	—
Inductance, L, M	henry	H	M = Φ/I	10^{-9}	—
Permeability, μ	—	H/m	μ = B/H	$4\pi \times 10^{-7}$	—

*N denotes Neumann's integral for two linear circuits each carrying the current I. F_z is the force between the two circuits in the direction defined by coordinate z, the circuits being in a vacuum.

†p denotes power.

‡l^2 denotes area.

§c = velocity of light in free space in cm/s = 2.997925×10^{10}.

2-4 The International System (SI) of Units

The international MKSA system of units was adopted in 1960 by the Eleventh General Conference of Weights and Measures under the name *Système International d'Unités* (SI). The SI system is replacing all other systems in the metric countries and its widespread acceptance dooms other systems to eventual obsolescence.

The six fundamental SI quantities are listed in Table 2-2. The derived units are expressed in terms of these six basic units by defining equations. Some examples of defining equations are given in Table 2-3 for the electric and magnetic quantities. Table 2-4 lists, together with the fundamental quantities which are repeated in this table, the supplementary and derived units in the SI which are recommended for use by the General Confeŕence.

The first column in Table 2-4 shows the *quantities* (fundamental, supple-

TABLE 2-4

FUNDAMENTAL, SUPPLEMENTARY, AND DERIVED UNITS

Quantity	Symbol	Dimension	Unit	Unit Symbol
Fundamental				
Length	l	L	meter	m
Mass	m	M	kilogram	kg
Time	t	T	second	s
Electric current	I	I	ampere	A
Thermodynamic temperature	T	Θ	degree Kelvin	°K
Luminous intensity			candela	cd
Supplementary*				
Plane angle	α, β, γ	$[L]^{\circ}$	radian	rad
Solid angle	Ω	$[L^2]^{\circ}$	steradian	sr
Derived				
Area	A	L^2	square meter	m^2
Volume	V	L^3	cubic meter	m^3
Frequency†	f	T^{-1}	hertz	Hz (1/s)
Density	ρ	$L^{-3}M$	kilogram per cubic meter	kg/m^3
Velocity	v	LT^{-1}	meter per second	m/s
Angular velocity	ω	$[L]^{\circ}T$	radian per second	rad/s
Acceleration	a	LT^{-2}	meter per second squared	m/s^2
Angular acceleration	α	$[L]^{\circ}T^{-2}$	radian per second squared	rad/s^2
Force	F	LMT^{-2}	newton	N (kg m/s^2)
Pressure, stress	p	$L^{-1}MT^{-2}$	newton per square meter	N/m^2
Work, energy	W	L^2MT^{-2}	joule	J (N m)
Power	P	L^2MT^{-3}	watt	W (J/s)
Quantity of electricity	Q	TI	coulomb	C (A s)
Potential difference, electromotive force	V	$L^2MT^{-3}I^{-1}$	volt	V (W/A)
Electric fieldstrength	E, ϵ	$LMT^{-3}I^{-1}$	volt per meter	V/m
Electric resistance	R	$L^2MT^{-3}I^2$	ohm	Ω (V/A)
Electric capacitance	C	$L^{-2}M^{-1}T^4I^2$	farad	F (A s/V)
Magnetic flux	Φ	$L^2MT^{-2}I^{-1}$	weber	Wb (v s)
Magnetic fieldstrength	H	$L^{-1}I$	ampere per meter	A/m
Magnetic flux density†	B	$MT^{-2}I^{-1}$	tesla	T (Wb/m^2)
Inductance	L	$L^2MT^{-2}I^2$	henry	H (V s/A)
Magnetomotive force	U	I	ampere	A
Luminous flux			lumen	lm (cd sr)
Luminance			candela per square meter	cd/m^2
Illumination			lux	lx (lm/m^2)

*The Eleventh General Conference designated these units as *supplementary*, although it could be argued that they are derived units.

†In some countries, frequency is expressed not in Hz but in the equivalent unit, cycle per second (c/s), and magnetic flux density, not in T, but in the equivalent weber per square meter (Wb/m^2).

mentary, and derived). The second column gives the *equation symbol* for each quantity. The third column lists the *dimension* of each derived unit in terms of the six fundamental dimensions. The fourth column gives the name of each *unit*; the fifth, the unit *symbol*. The *unit* symbol should not be confused with the *equation* symbol; i.e., the equation symbol for resistance is R, but the unit abbreviation (symbol) for ohm is Ω.

2-5 The English and Other Units

The English system of units uses the *foot*(ft), the *pound-mass* (lb), and the *second* (s) as the three fundamental units of length, mass, and time, respectively. Although the measures of length and weight are legacies of the Roman occupation of Britain and therefore rather poorly defined, the *inch* (defined as one-twelfth of the foot), has since been fixed at *exactly* 25.4 mm. Similarly, the measure for the pound (lb) has been determined as *exactly* 0.45359237 kg. These two figures allow all units in the English system to be converted into SI units.

Starting with the fundamental units, foot, pound, and second, the mechanical units may be derived simply by substitution into the dimensional equations of Table 2-4. For example, the unit of density will be expressed in lb/ft^3 and the unit of acceleration in ft/s^2. The derived unit of force in the ft-lb-s system is called the *poundal* and is the force required to accelerate 1 pound-mass at the rate of 1 ft/s^2. As a result, the unit for work or energy becomes the foot-poundal (ft pdl).

Various other systems have been devised and were used in various parts of the world. The *MTS* (meter-tonne-second) system, was especially designed for engineering purposes in France and provided a replica of the CGS system except that the length and mass units (meter and tonne, respectively) were more suitable in practical engineering applications. *Gravitational* systems define the second fundamental unit as the *weight* of a mass measure; i.e., as the force by which that mass is attracted to the earth by gravity. In contrast to the

TABLE 2-5

BRITISH INTO METRIC CONVERSION

English Unit		Symbol	Metric Equivalent	Reciprocal
Length	1 foot	ft	30.48 cm	0.0328084
	1 inch	in.	25.4 mm	0.0393701
Area	1 square foot	ft^2	9.29030×10^2 cm^2	0.0107639×10^{-2}
	1 square inch	in.^{2}	6.4516×10^2 mm^2	0.155000×10^{-2}
Volume	1 cubic foot	ft^3	0.0283168 m^3	35.3147
Mass	1 pound (avdp)	lb	0.45359237 kg	2.20462
Density	1 pound per cubic foot	lb/ft^3	16.0185 kg/m^3	0.062428
Velocity	1 foot per second	ft/s	0.3048 m/s	3.28084
Force	1 poundal	pdl	0.138255 N	7.23301
Work, energy	1 foot-poundal	ft pdl	0.0421401 J	23.7304
Power	1 horsepower	hp	745.7 W	0.00134102
Temperature	degree F	t°F	$5(t - 32)/9$°C	

gravitational systems, the so-called absolute systems, as the CGS and SI, use the mass measure as the second fundamental unit, but its value is independent of gravitational attraction.

Since English measures are still extensively used, both in Britain and on the North American continent, conversion into the SI becomes necessary if we wish to work in that system. Table 2-5 lists some of the more common conversion factors for British into metric units.

2-6 Conversion of Units

It is often necessary to convert physical quantities from one system of units into another. Section 2-1 stated that a physical quantity is expressed in both unit and number of measure: it is the unit which must be converted, not the number of measure. Dimensional equations are very convenient for converting the numerical value of a dimensional quantity, when the units are transformed from one system to the other. The technique requires a knowledge of the numerical relation between the fundamental units and some dexterity in the manipulation of multiples and submultiples of the units.

The method used in converting from one system into the other is illustrated by a number of examples of progressively increasing difficulty.

Example 2-1: The floor area of an office building is 5,000 m^2. Calculate the floor area in ft^2.

SOLUTION: To convert the unit m^2 into the new unit ft^2, we must know the relation between them. In Table 2-5 the metric equivalent of 1 ft is 30.48 cm, or 1 ft = 0.3048 m. Therefore,

$$A = 5{,}000 \text{ m}^2 \times \left(\frac{1 \text{ ft}}{0.3048 \text{ m}}\right)^2 = 53{,}800 \text{ ft}^2$$

Example 2-2: The floor area of a classroom measures 30 ft by 24 ft. Calculate the floor area in m^2.

SOLUTION: Again consulting Table 2-5 we find that the reciprocal of the ft-to-cm conversion is 0.0328084. Therefore,

$$1 \text{ cm} = 0.0328 \text{ ft} \quad \text{or} \quad 1 \text{ m} = 3.28 \text{ ft}.$$

$$A = 30 \text{ ft} \times 24 \text{ ft} = 720 \text{ ft}^2$$

$$720 \text{ ft}^2 = 720 \text{ ft}^2 \times \left(\frac{1 \text{ m}}{3.28 \text{ ft}}\right)^2 = 67.3 \text{ m}^2$$

Example 2-3: A flux density in the CGS system is expressed as 20 maxwells/cm^2. Calculate the flux density in lines/in^2. (NOTE: 1 maxwell = 1 line)

SOLUTION:

$$B = \frac{20 \text{ maxwells}}{\text{cm}^2} \times \left(\frac{2.54 \text{ cm}}{\text{in.}}\right)^2 \times \frac{1 \text{ line}}{1 \text{ maxwell}} = 129 \text{ lines/in}^2.$$

Example 2-4: The velocity of light in free space is given as 2.997925×10^8 m/s. Express the velocity of light in km/hr.

SOLUTION:

$$c = 2.997925 \times 10^8 \frac{\text{m}}{\text{s}} \times \frac{1 \text{ km}}{10^3 \text{ m}} \times \frac{3.6 \times 10^3 \text{s}}{1 \text{ hr}} = 10.79 \times 10^8 \text{ km/hr}$$

Example 2-5: Express the density of water, 62.5 lb/ft^3, in (a) lb/in.^{3}, (b) g/cm^3.

SOLUTION:

(a) $\frac{62.5 \text{ lb}}{\text{ft}^3} \times \left(\frac{1 \text{ ft}}{12 \text{ in.}}\right)^3 = 3.62 \times 10^{-2} \text{ lb/in.}^3$

(b) $3.62 \times 10^{-2} \frac{\text{lb}}{\text{in.}^3} \times \frac{453.6 \text{ g}}{1 \text{ lb}} \times \left(\frac{1 \text{ in.}}{2.54 \text{ cm}}\right)^3 = 1 \text{ g/cm}^3$

Example 2-6: The speed limit on a European highway is 60 km/hr. Calculate the limit in (a) mi/hr, (b) ft/s.

SOLUTION:

(a) $\frac{60 \text{ km}}{\text{hr}} \times \frac{10^3 \text{ m}}{1 \text{ km}} \times \frac{10^2 \text{ cm}}{1 \text{ m}} \times \frac{1 \text{ in.}}{2.54 \text{ cm}} \times \frac{1 \text{ ft}}{12 \text{ in.}} \times \frac{1 \text{ mi}}{5{,}280 \text{ ft}} = 37.4 \text{ mi/hr}$

(b) $\frac{37.4 \text{ mi}}{\text{hr}} \times \frac{5{,}280 \text{ ft}}{1 \text{ mi}} \times \frac{1 \text{ hr}}{3.6 \times 10^3 \text{s}} = 54.9 \text{ ft/s}$

References

1. Hvistendahl, H. S., *Engineering Units and Physical Quantities*. London: Macmillan and Co., Ltd., 1964.

2. Kaye, G. W. C., and T. H. Laby, *Tables of Physical and Chemical Constants*, 13th ed. London: Longmans, Green and Co., Ltd., 1966.

Problems

1. Express the following in Hz, using powers of ten:

(a) 1500 Hz	(d) 0.5 MHz
(b) 20 kHz	(e) 50 MHz
(c) 1800 kHz	(f) 1.2 GHz

2. Express the following in V, using powers of ten:

(a) 24 mV (d) 1.2 MV
(b) 540 μV (e) 16 nV
(c) 4.4 kV (f) 0.4 mV

3. Express the following in A, using powers of ten:
(a) 23.5 mA (d) 72 nA
(b) 45 μA (e) 620 μA
(c) 0.25 mA (f) 74.6 nA

4. Express the following in μA, using powers of ten:
(a) 0.00036 A (d) 25 pA
(b) 0.027 A (e) 2.5 A
(c) 0.250 mA (f) 1.275 mA

5. Calculate the height in cm of a man 5 ft 11 in. tall.

6. Calculate the mass in kg of 1 yd^3 of iron when the density of iron is 7.86 g/cm^3.

7. Calculate the conversion factor to change mi/hr to ft/s.

8. An electrically charged body has an excess of 10^{15} electrons. Calculate its charge in C.

9. A train covers a distance of 220 mi in 2 hr and 45 min. Calculate the average velocity of the train in m/s.

10. Two electric charges are separated by a distance of 1 m. If one charge is +10 C and the other charge −6 C, calculate the force of attraction between the charges in N and in lb. (Assume that the charges are placed in a vacuum.)

11. The practical unit of electrical energy is the kWh. The unit of energy in the SI is the joule (J). Calculate the number of joules in 1 kWh.

12. A crane lifts a 100-kg mass a height of 20 m in 5 s. Calculate (a) the work done by the crane, in SI units, (b) the increase of potential energy of the mass, in SI units, (c) the power, or rate of doing the work, in SI units.

13. Calculate the voltage of a battery if a charge of 3×10^{-4} C residing on the positive battery terminal possesses 6×10^{-2} J of energy.

14. An electric charge of 0.035 C flows through a copper conductor in 5 min. Calculate the average current in mA.

15. An average current of 25 μA is passed through a wire for 30 s. Calculate the number of electrons transferred through the conductor.

16. The speed limit on a four-lane highway is 70 mi/hr. Calculate the speed limit in (a) km/hr, (b) ft/s.

17. The density of copper is 8.93 g/cm^3. Express the density in (a) kg/m^3, (b) lb/ft^3.

18. The melting point of magnesium is 650°C. Express the melting point in (a) °F, (b) °K.

THREE

STANDARDS OF MEASUREMENT

3-1 Classification of Standards

A standard of measurement is a physical *representation* of a unit of measurement. A unit is *realized* by reference to an arbitrary material standard or to natural phenomena including physical and atomic constants. For example, the fundamental unit of mass in the metric system (SI) is the *kilogram,* defined as the mass of a cubic decimeter of water at its temperature of maximum density of 4°C (see Sec. 2-2). This unit of mass is represented by a material standard: the mass of the International Prototype Kilogram, consisting of a platinum-iridium alloy cylinder. This cylinder is preserved at the International Bureau of Weights and Measures at Sèvres, near Paris and is the *material representation* of the kilogram. Similar standards have been developed for other units of measurement, including standards for the fundamental units as well as for some of the derived mechanical and electrical units.

Just as there are fundamental and derived units of measurement, we find different types of *standards of measurement,* classified by their function and application in the following categories:

(a) International standards
(b) Primary standards
(c) Secondary standards
(d) Working standards

The *international standards* are defined by international agreement. They represent certain units of measurement to the closest possible accuracy that production and measurement technology allow. International standards are periodically evaluated and checked by *absolute measurements* in terms of the fundamental units (see Table 2-2). These standards are maintained at the International Bureau of Weights and Measures and are *not* available to the ordinary user of measuring instruments for purposes of comparison or calibration.

The *primary* (basic) *standards*, are maintained by national standards laboratories in different parts of the world. The National Bureau of Standards (NBS) in Washington is responsible for maintenance of the primary standards in North America. Other national laboratories include the National Physical Laboratory (NPL) in Great Britain, and the oldest in the world, the Physikalisch-Technische Reichsanstalt in Germany. The primary standards, again *representing* the fundamental units and some of the derived mechanical and electrical units, are independently calibrated by absolute measurements at each of the national laboratories. The results of these measurements are compared against each other, leading to a world average figure for the primary standard. Primary standards are *not* available for use outside the national laboratories. One of the main functions of primary standards is the verification and calibration of secondary standards.

The *secondary standards* are the basic *reference* standards used in industrial measurement laboratories. These secondary standards are maintained by the particular involved industry and are checked locally against other reference standards in the area. The responsibility for maintenance and calibration of secondary standards rests entirely with the industrial laboratory itself. Secondary standards are generally sent to the national standards laboratories on a periodic basis for calibration and comparison against the primary standards. They are then returned to the industrial user with a *certification* of their measured value in terms of the primary standard.

The *working standards* are the principal tools of a measurement laboratory. They are used to check and calibrate general laboratory instruments for accuracy and performance or to perform comparison measurements in industrial applications. A manufacturer of precision resistances, for example, may use a *standard resistor* (a *working* standard) in the quality control department of his plant to check his testing equipment. In this case, he *verifies* that his measurement setup performs within the required limits of accuracy.

In electrical and electronic measurement, we are concerned with the electrical and magnetic standards of measurement. These are discussed in the following sections. We have seen, however, that electrical units can be traced back to the basic units of length, mass, and time (in fact, the national laboratories perform *measurements* to relate derived electrical units to fundamental units) and they deserve some investigation here.

3-2 Standards for Mass, Length, and Volume

The metric *unit of mass* was originally defined as the mass of a cubic decimeter of water at its temperature of maximum density. The *material representation* of this unit is the International Prototype Kilogram, preserved at the International Bureau of Weights and Measures near Paris. The *primary standard* of mass in North America is the United States Prototype Kilogram, preserved by the NBS to an accuracy of 1 part in 10^8 and occasionally verified against the standard at the International Bureau. *Secondary standards* of mass, kept by the industrial laboratories, generally have an accuracy of 1 ppm (part per million) and may be verified against the NBS primary standard. The commercial *working standards* are available in a wide range of values to suit almost any application. Their accuracy is in the order of 5 ppm. The working standards, in turn, are checked against the secondary laboratory standards.

The *pound* (lb), established by the Weights and Measures Act of 1963, which actually came into effect on January 31, 1964, is defined as equal to 0.45359237 kg *exactly*. All countries which retain the pound as the basic unit of measurement have now adopted the new definition, which supersedes the former imperial standard pound made of platinum.

The metric *unit of length* (the meter), initially defined as the ten-millionth part of the meridianal quadrant through Paris (see Sec. 2-2), was materially represented by the distance between two lines engraved on a platinum-iridium bar preserved at the International Bureau of Weights and Measures near Paris. In 1960 the meter was redefined more accurately in terms of an *optical standard,* namely, the orange-red radiation of a krypton atom. The internationally specified krypton-86 discharge lamp, excited and observed under well-defined conditions, emits orange light whose wavelength now constitutes the basic standard of length to 1 ppm. The meter, as the SI unit of length, is now defined as equal to 1,650,763.73 wavelengths in vacuum of the orange-red radiation of the krypton-86 atom. The optically defined standard of length represents the same fundamental unit of length as the former platinum-iridium bar, but its accuracy is an order of magnitude greater.

The *yard* is defined as 0.9144 m *exactly* (1 in. = 25.4 mm exactly) and by this definition then also depends on the krypton-86 wavelength standard. This definition of the yard supersedes the former definition in terms of the imperial standard yard. All countries which retain the yard as the basic unit of measurement have now adopted this new definition.

The most widely used industrial *working standards* of length are precision *gage blocks,* made of steel. These steel blocks have two plane parallel surfaces, a specified distance apart, with accuracy tolerances in the 0.5–0.25-micron range

(1 micron = one millionth of 1 m). The development and use of precision gage blocks, low in cost and of high accuracy, has made it possible to manufacture interchangeable industrial components in a very economical application of precision measurement.

The unit of *volume* is a derived quantity and is not represented by an international standard. The NBS however has constructed a number of primary standards of volume, calibrated in terms of the absolute dimensions of length and mass. Secondary derived standards of volume are available and may be calibrated in terms of the NBS primary standards.

3-3 Time and Frequency Standards*

Since early times men have sought a reference standard for a uniform time scale together with means to interpolate from it a small time interval. For many centuries, the time reference used was the rotation of the earth about its axis with respect to the sun. Precise astronomical observations have shown that the rotation of the earth about the sun is quite irregular, owing to secular and irregular variations in the rotational speed of the earth. Since the time scale based on this apparent *solar time* does not represent a uniform time scale, other avenues were explored. *Mean solar time* was thought to give a more accurate time scale. A mean solar day is the average of all the apparent days in the year. A *mean solar second* is then equal to 1/86,400 of the mean solar day. The mean solar second, thus defined, is still inadequate as the fundamental unit of time, since it is tied to the rotation of the earth, which is now known to be nonuniform.

The system of *universal time* (UT) is also based on the rotation of the earth about its axis. The system of universal time or mean solar time is known as UT_0 and is subject to periodic, long-term, and irregular variations. Correction of UT_0 has led to two subsequent universal time scales: UT_1 and UT_2. UT_1 recognizes the fact that the earth is subject to polar motion and the UT_1 time scale is based on the true angular rotation of the earth, corrected for *polar motion*. The UT_2 time scale is UT_1 with an additional correction for *seasonal* variations in the rotation of the earth. These variations are apparently caused by seasonal displacement of matter over the earth's surface, such as changes in the amount of ice in the polar regions as the sun moves from the southern hemisphere to the northern and back again through the year. This *cyclic* redistribution of mass acts on the earth's rotation since it produces changes in its

*Application note AN 52 (*Frequency and Time standards*), published by Hewlett-Packard, Palo Alto, Calif., describes modern methods of frequency comparisons, time scales, world-wide standards broadcasts.

moment of inertia. The *epoch*, or *instant of time*, of UT_2 can be established to an accuracy of a few milliseconds, but it is not usually distributed to this accuracy. The epoch indicated by the standard radio time signals may differ from the epoch of UT_2 by as much as 100 ms. The actual values of the differences are given in bulletins published by the national time services (NBS) and by the Bureau Internationale de l'Heure (Paris Observatory).

The search for a truly universal time unit has led astronomers to define a time unit called *Ephemeris time* (ET). ET is based on astronomical observations of the motion of the moon about the earth. Since 1956 the *ephemeris second* has been defined by the International Bureau of Weights and Measures as the fraction 1/31,556,925.9747 of the tropical year for 1900 January 0 at 12 h ET, and adopted as the *fundamental invariable unit of time*. A disadvantage of the use of the ephemeris second is that it can be determined only several years in arrears and then only indirectly, by observations of the positions of the sun and the moon. For *physical measurements*, the unit of time interval has now been defined in terms of an *atomic standard*. The universal second and the ephemeris second, however, will continue to be used for naviagtion, geodetic surveys, and celestial mechanics.

Development and refinement of *atomic resonators* has made possible control of the frequency of an oscillator and, hence, by frequency conversion, *atomic clocks*. The transition between two energy levels, E_1 and E_2, of an atom is accompanied by the emission (or absorption) of radiation having a frequency given by $h\nu = E_2 - E_1$, where h is Planck's constant. Provided that the energy states are not affected by external conditions, such as magnetic fields, the frequency ν is a *physical constant*, depending *only* on the internal structure of the atom. Since frequency is the inverse of time interval, such an atom provides a *constant time interval*. Atomic transitions of various metals were investigated, and the first atomic clock, based on the cesium atom, was put into operation in 1955. The time interval, provided by the cesium clock, is more accurate than that provided by a clock calibrated by astronomical measurements. The *atomic unit of time* was first related to UT but was later expressed in terms of ET. The International Committee of Weights and Measures has now defined the second in terms of the frequency of the cesium transition, assigning a value of 9,192,631,770 Hz to the hyperfine transition of the cesium atom unperturbed by external fields.

The *atomic definition* of the second realizes an accuracy much greater than that achieved by astronomical observations, resulting in a more uniform and much more convenient time base. Determinations of time intervals can now be made in a few minutes to greater accuracy than was possible before in astronomical measurements that took many years to complete. An atomic clock with a precision exceeding 1 μs per day is now in operation as a *primary frequency* *standard* at the NBS. An atomic time scale, designated NBS-A, is maintained with this clock.

NBS disseminates its time and frequency standards by broadcasts from several radio stations, operating at different transmission frequencies from various parts of the continental United States and Hawaii. Complete information regarding broadcast schedules and changes in station operation may be obtained upon request from the NBS.

3-4 The Electrical Standards for Current, Voltage, and Resistance

The metric system of units (SI) defines the *ampere* (the fundamental unit of electric current) as the constant current which, if maintained in two straight parallel conductors of infinite length and negligible circular cross section placed 1 m apart in a vacuum, will produce between these conductors a force equal to 2×10^{-7} newton per meter length. Early measurements of the absolute value of the ampere were made with a *current balance* which measured the force between two parallel conductors. These measurements were rather crude and the need was felt to produce a more *practical* and *reproducible standard* for the national laboratories. By international agreement, the value of the *International Ampere* was based on the electrolytic deposition of silver from a silver nitrate solution. The International Ampere was then defined as that current which deposits silver at the rate of 1.118 mg/s from a standard silver nitrate solution. Difficulties were encountered in the exact measurement of the deposited silver and slight discrepancies existed between measurements made independently by the various national standards laboratories.

In 1948 the International Ampere was superseded by the *Absolute Ampere*. The determination of the Absolute Ampere is again made by means of a current balance, which *weighs* the *force* exerted between two current-carrying coils. Improvement in the techniques of force measurement yields a value for the ampere far superior to the early measurements. The relation between the force and the current which produces the force can be calculated from fundamental electromagnetic theory concepts and reduces to a simple computation involving the geometric dimensions of the coils. The Absolute Ampere is now the *fundamental unit of electric current* in the SI, universally accepted by international agreement.

Instruments manufactured before 1948 are calibrated in terms of the International Ampere but newer instruments are using the Absolute Ampere as the basis for calibration. Since both types of instruments may be found side by side in one laboratory, the NBS has established conversion factors to relate both units. These factors are given in Sec. 2-3.

Voltage, current, and resistance are related by Ohm's law of constant proportionality ($E = IR$). The specification of any two quantities automatically sets the third. Two types of material standards form a combination which conveniently serves to maintain the ampere with high precision over long periods of time: the *standard resistor* and the *standard cell* (for voltage). Each of these is descibed below.

The absolute value of the ohm in the rationalized MKS system (SI) is defined in terms of the fundamental units of length, mass, and time. The *absolute measurement* of the ohm is carried out by the International Bureau of Weights and Measures in Sèveres and also by the national standards laboratories, which preserve a group of *primary* resistance standards. The NBS maintains a group of those primary standards (1-Ω standard resistors) which are periodically checked against each other and are occasionally verified by absolute measurements. The standard resistor is a coil of wire of some alloy like *manganin* which has a high electrical resistivity and a low temperature coefficient of resistance (almost constant temperature-resistance relation). The resistance coil is mounted in a double-walled sealed container (Fig. 3-1) to prevent changes in resistance due to moisture conditions in the atmosphere. With a set of four or five 1-Ω resistors of this type, the unit of resistance can be represented with a precision of a few parts in 10^7 over several years.

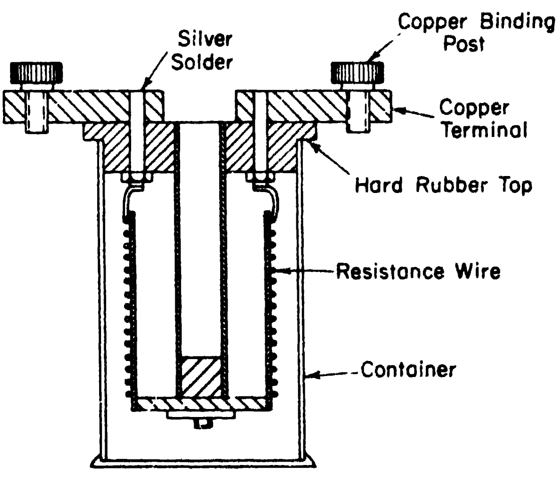


Figure 3-1
A cross-sectional view of a double-walled resistance standard (*Courtesy Hewlett Packard Co.*).

Secondary standards and *working* standards are available from some instrument manufacturers in a wide range of values, usually in multiples of 10 Ω. These standard resistors are made of alloy resistance wire, such as manganin or Evanohm. Figure 3-2 shows a photograph of a laboratory secondary standard, sometimes referred to as a *transfer resistor*. The resistance coil of the transfer resistor is supported between polyester film to reduce stresses on the

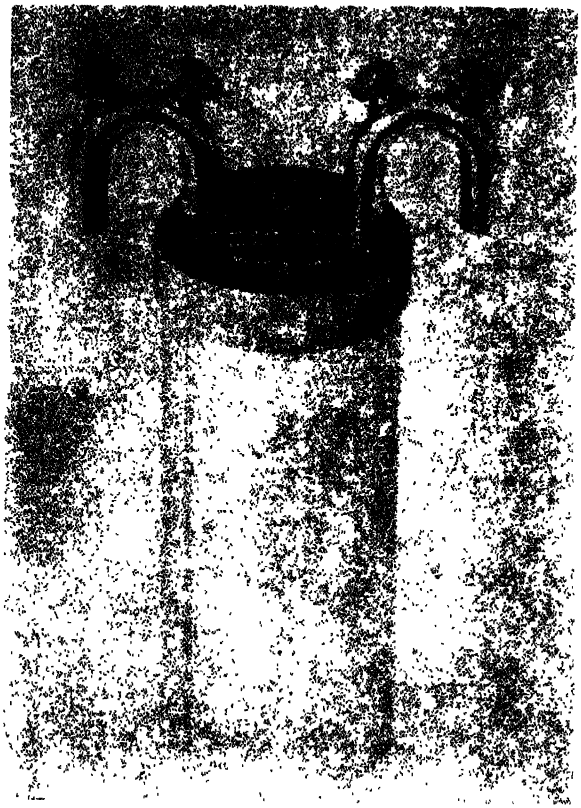


Figure 3-2
A 10-kilohm transter resistor (*Courtesy Hewlett Packard Co.*).

wire and to improve the stability of the resistor. The coil is immersed in moisture-free oil and placed in a sealed can. All the connections to the resistance coil are silver soldered and the terminal hooks are made of nickel-plated oxygen-free copper. The transfer resistor is checked for stability and temperature characteristics at its rated power and a specified operating temperature (usually 25°C). A *calibration report* accompanying the resistor, specifies its traceability to the NBS standards and also includes the α and β temperature coefficients. Although the selected resistance wire provides almost constant resistance over a fairly wide temperature range, the exact value of the resistance at any temperature can be calculated from the formula:

$$R_t = R_{25^\circ C} + \alpha(t - 25) + \beta(t - 25)^2 \tag{3-1}$$

where R_t = resistance at the ambient temperature t
$R_{25^\circ C}$ = resistance at 25°C
α and β = temperature coefficients

Temperature coefficient α is usually less than 10×10^{-6}, and coefficient β

lies between -3×10^{-7} and -6×10^{-7}. This means that a change in temperature of 10°C from the specified reference temperature of 25°C may cause a change in resistance of 30 to 60 ppm (parts per million) from the nominal value.

Transfer resistors find application in industrial, research, standards, and calibration laboratories. In typical applications, the transfer resistor may be used for resistance and ratio determinations or in the construction of ultra-linear decade dividers which can then be used for the calibration of universal ratio sets, voltboxes, and Kelvin-Varley dividers.

The *primary voltage standard,* selected by the NBS for the maintenance of the volt, is the *normal* or *saturated Weston cell.* The Weston cell has a positive electrode of mercury and a negative electrode of cadmium amalgam (10 per cent Cd). The electrolyte is a solution of cadmium sulphate. These components are placed in an H-shaped glass container, as shown in Fig. 3-3.

There are two types of Weston cells: the *saturated* cell, in which the electrolyte is saturated at all temperatures by cadmium sulphate crystals covering the electrodes; the *unsaturated* cell, in which the concentration of cadmium sulphate is such that it produces saturation at 4°C. The unsaturated cell has a negligible temperature coefficient of voltage at normal room temperatures. The saturated cell shows a voltage variation of approximately $-40\ \mu$V per 1°C rise, but it is better reproducible and more stable than the unsaturated cell.

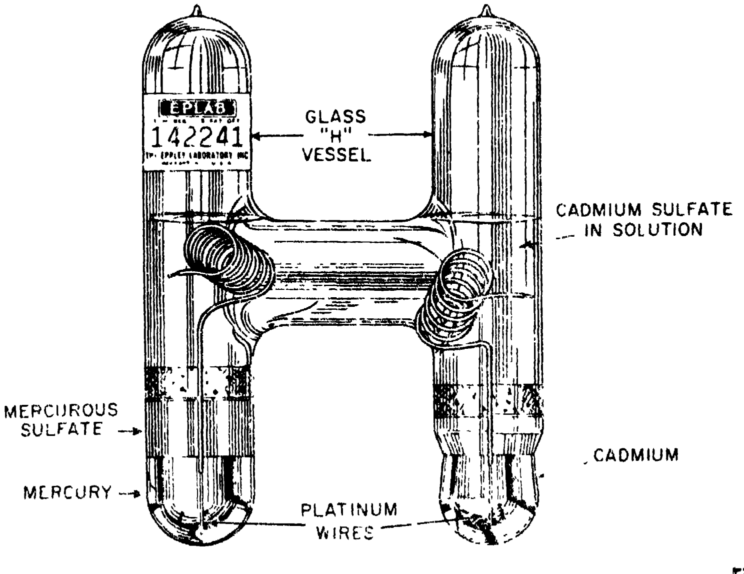


Figure 3-3

Construction details of a saturated Weston cadmium cell (*Courtesy The Eppley Laboratory, Inc.*).

National standards laboratories, such as the NBS, maintain a number of *saturated* cells as the *primary standard* for voltage. The cells are kept in an oil bath to control their temperature to within 0.01°C. The voltage of the Weston saturated cell at 20°C is 1.01858 V (absolute) and the emf at other temperatures is given by the formula,

$$e_t = e_{20°C} - 0.000046(t - 20) - 0.00000095(t - 20)^2 + 0.00000001(t - 20)^3 \tag{3-2}$$

Saturated Weston cells remain satisfactory as voltage standards for periods of ten to twenty years, provided that they are carefully treated. Their drift in voltage is in the order of 1 μV per year. Since saturated cells are temperature sensitive, they are unsuited for general laboratory use as secondary or working standards.

More rugged portable *secondary* and *working standards* are found in the *unsaturated* Weston cell. These cells are very similar in construction 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 varies less than 0.01 per cent from 10°C to 40°C. The voltage of the cell is usually indicated on

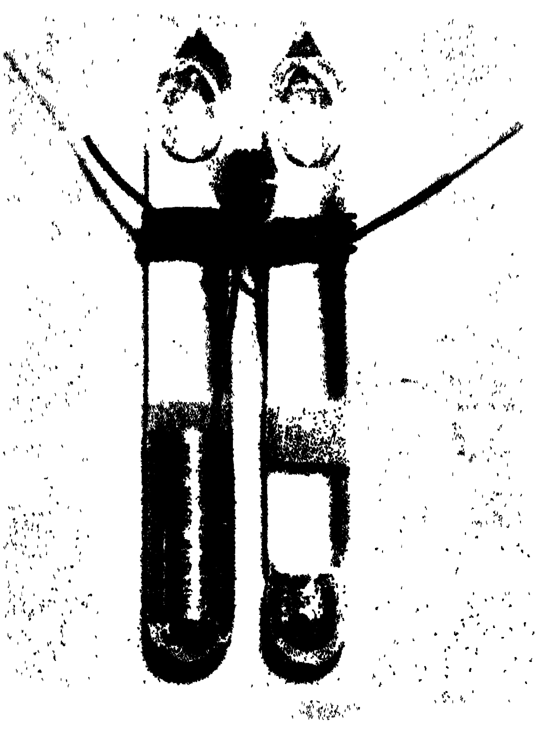


Figure 3-4
An unsaturated Weston cadmium cell: emf 1.0193 V, accuracy 0.1% (*Courtesy The Eppley Laboratory, Inc.*).

the cell housing, as shown in Fig. 3-4 (1.0193 abs. V). The internal resistance of Weston cells range from 500 Ω to 800 Ω. The current drawn from these cells, therefore, should not exceed 100 μA, because the nominal voltage is affected, owing to internal voltage drop.

Versatile *laboratory working* standards have been developed with accuracies in the order of standard cell accuracy. Figure 3-5 shows a photograph of a multipurpose laboratory voltage standard, called a *transfer standard,* based on the operation of a Zener diode as the voltage reference element. The instrument basically consists of a Zener-controlled voltage source placed in a temperature-controlled environment to improve its long-term stability, and a precision output voltage divider. The temperature-controlled oven is held to within $\pm 0.03°C$ over an ambient temperature range of 0°C to 50°C, providing an output stability in the order of 10 ppm/month. The four available outputs are (a) a 0–1000 μV source with 1-μV resolution, called (Δ); (b) a 1.000-V reference for voltbox potentiometric measurements; (c) a 1.018 + (Δ) reference for saturated cell comparisons; (d) a 1.0190 + (Δ) reference for unsaturated cell comparisons. The dc transfer standard can be used as a transfer instrument and can be moved to the piece of equipment to be calibrated, since it can easily be disconnected from the power line at one location and set up at a different location where it will recover to within ± 1 ppm in approximately 30 minutes warm-up time.

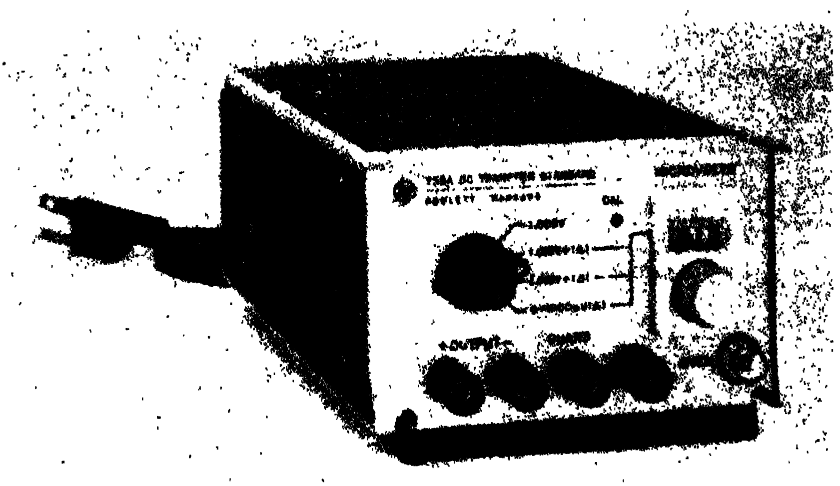


Figure 3-5
A dc transfer standard that can be used as a 1.000-V reference source, a standard cell comparison instrument, and a 0–1,000 μV dc source. (*Courtesy Hewlett Packard Co.*).

3-5 Other Electrical and Magnetic Standards

Since the unit of resistance is represented by the standard resistor and the unit of voltage by the standard Weston cell, many electrical and magnetic units may be expressed in terms of these standards. The unit of *capacitance* (the farad) can be measured with a Maxwell dc commutated bridge, where the capacitance is computed from the resistive bridge arms and the frequency of the dc commutation. This bridge is shown in Fig. 3-6. Although the exact derivation of the expression for capacitance in terms of the resistances and the frequency is rather involved, it may be seen that the capacitor could be measured by this method. Since both resistance and frequency can be determined very accurately, the value of the capacitance can be measured with great accuracy. *Standard capacitors* are usually constructed from interleaved metal plates with air as the dielectric material. The area of the plates and the distance between them must be known quite accurately, and the capacitance of the air capacitor can be determined from these basic dimensions. The NBS maintains a bank of air capacitors as standards and uses them to calibrate the secondary and working standards of measurement laboratories and industrial users.

Capacitance working standards are obtainable in a range of suitable values. Smaller values are usually air capacitors, whereas the larger capacitors use solid dielectric materials. The high dielectric constant and the very thin dielectric layer accounts for the compactness of these standards. Silver-mica capacitors make excellent working standards; they are very stable and have a very low dissipation factor (see Sec. 8-5), a very small temperature coefficient, and little or no aging effect. Mica capacitors are available in decade mounting, but decade capacitors are usually not guaranteed better than 1 per cent. Fixed standards are generally used where accuracy is important.

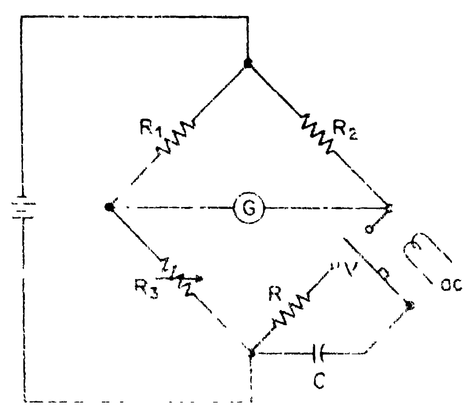


Figure 3-6
Commutated dc method for measuring capacitance. Capacitor C is alternately charged and discharged through the commutating contact and resistor R. Bridge balance is obtained by adjustment of R_3, allowing exact determination of the capacitance value in terms of bridge-arm constants and frequency of commutation.

The *primary inductance standard* is derived from the ohm and the farad, rather than from the large geometrically constructed inductors used in the determination of the absolute value of the ohm. The NBS selected a *Campbell standard* of mutual inductance as the primary standard for both mutual and self-inductance. Inductance *working standards* are commercially available in a wide range of practical values, both fixed and variable. A typical set of fixed inductance standards includes values from approximately 100 μH to 10 H, with a guaranteed accuracy of 0.1 per cent at a specified operating frequency. Variable inductors are also available. Typical mutual inductance accuracy is in the order of 2.5 per cent and inductance values range from 0 to 200 mH. Remember that *distributed capacitance* exists between the windings of these inductors, and errors they introduce must be taken into account. These considerations are usually specified with commercial equipment.

Measurements of magnetic flux generally involve the use of a *ballistic galvanometer*. The ballistic galvanometer is essentially a d'Arsonval movement, specifically designed for long period operation (20 s to 30 s) and of high sensitivity. In ballistic measurements, the coil receives a *momentary* impulse of current, which causes it to swing to one side and then return to rest in an oscillatory motion, governed by the circuit damping (see also Sec. 4-3). When the current impulse is of sufficiently short duration, the initial deflection from its rest position is directly proportional to the quantity of electrical discharge through the coil. The relative magnitude of the current impulse is then measured in terms of the *initial* angular deflection of the coil and we can write:

$$Q = K\theta \tag{3-3}$$

where $Q = \int i\,dt$ over the short time duration = charge in coulombs
K = galvanometer sensitivity, in coulomb/radian of deflection
θ = angular deflection of the coil, in radians

The sensitivity, K, depends on the damping and should be obtained experimentally by a *calibration* check conducted under the *actual* conditions of use.

Several *calibration procedures* may be adopted,* including the capacitor method, the solenoid method, and the mutual inductor method. Figure 3-7 shows the connection for calibrating a ballistic galvanometer by a mutual inductor of convenient value. The deflection θ produced by the reversal of a known primary current I is observed.

The secondary voltage, e_2, produced by a changing primary current i_1 is

$$= M\frac{di_1}{d_t} \tag{3-4}$$

where M = mutual inductance in henries

*Stout, *op. cit*, pp. 438–442.

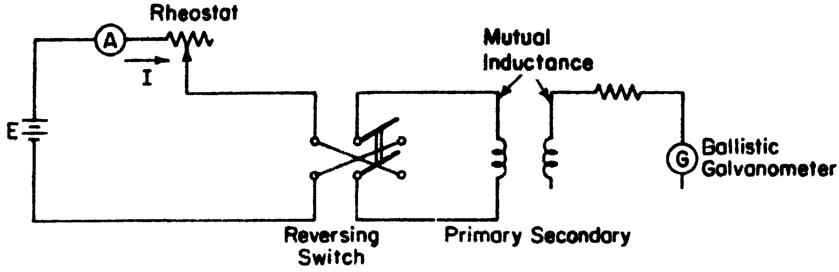


Figure 3-7
Calibration of a ballistic galvanometer by the mutual inductance method.

If R is the total resistance in the secondary circuit, the current in the galvanometer is

$$i_2 = \frac{M}{R}\frac{di_1}{dt} \tag{3-5}$$

or, rearranging,

$$i_2\,dt = \frac{M}{R}\,di_1 \tag{3-6}$$

Integration of Eq. (3-6) yields

$$\int_{t=0}^{t=t_0} i_2\,dt = \frac{M}{R}\int_{-I}^{+I} di_1 \tag{3-7}$$

and

$$Q = \frac{2MI}{R} \quad \text{(coulombs)} \tag{3-8}$$

Substitution of Eq. (3-8) into Eq. (3-3) yields a value for the galvanometer sensitivity, K, and

$$K = \frac{2MI}{R\theta} \tag{3-9}$$

Once calibrated, the ballistic galvanometer may be used to measure the flux produced by a variety of permanent magnets. The method is shown in Fig. 3-8. A *search coil,* surrounding a permanent magnet whose flux is to be measured, is connected in series with a ballistic galvanometer and a variable resistor. The variable resistor is generally adjusted to give critical damping to the galvanometer. If the permanent magnet is withdrawn *quickly* from the search coil, an impulse of current is produced and the galvanometer deflects. The instantaneous voltage induced in the galvanometer coil is

$$e = N\frac{d\phi}{dt} \quad \text{(V)} \tag{3-10}$$

where N = the number of turns in the search coil
ϕ = the flux of the permanent magnet, in webers

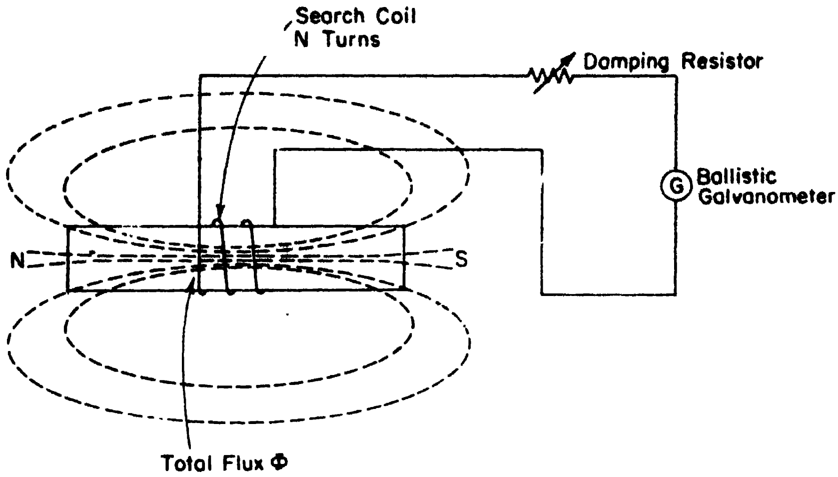


Figure 3-8
Flux measurement by ballistic galvanometer.

If R is the total resistance in the circuit, including the search coil, the galvanometer, and the series resistor, the current in the galvanometer coil is

$$i = \frac{e}{R} = \frac{N}{R}\frac{d\phi}{dt} \quad \text{(A)} \tag{3-11}$$

Integrating Eq. (3-11) over the period of the discharge,

$$\int_{t=0}^{t=t_0} i\,dt = \frac{N}{R}\int_{\phi=\Phi}^{\phi=0} d\phi. \tag{3-12}$$

and the quantity of charge through the ballistic galvanometer is

$$Q = \frac{N\Phi}{R} \quad \text{(C)} \tag{3-13}$$

where Φ = total flux of the permanent magnet

The galvanometer deflection is, from Eq. (3-3),

$$\theta = \frac{Q}{K} = \frac{N\Phi}{KR} \quad \text{(rad)} \tag{3-14}$$

Rearranging,

$$\Phi = \frac{KR\theta}{N} \quad \text{(Wb)}$$

It should be stressed here that the sensitivity factor K must be evaluated for the circuit resistance used in the measurement setup.

The measurement method, previously described, is used to measure the *standard flux* produced by a variety of permanent magnets. These permanent magnets are then maintained as *magnetic flux standards*.

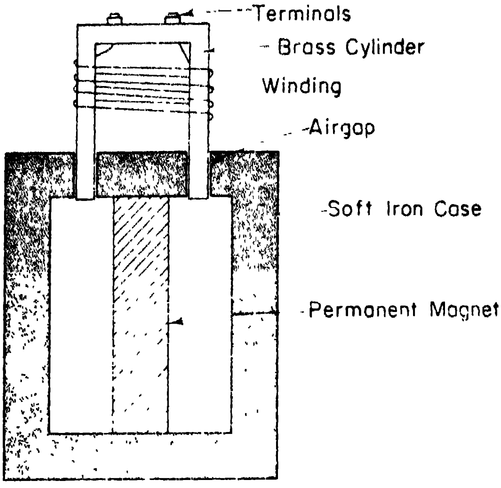


Figure 3-9
Principle of construction of the Hibbert magnetic standard.

It is often useful to have a standard flux source independent of external exciting current. The *Hibbert magnetic standard* (Fig. 3-9) is such a device. A permanent magnet is enclosed in a soft iron container which has a narrow circular air gap. A brass cylinder is suspended within the air gap. On this cylinder is wound an insulated winding of conducting material, such as copper. At the release of a catch, the brass cylinder and the winding assembly drop through the flux in the air gap. The resulting electrical current induced in the winding is proportional to the rate at which the magnetic flux was cut by the moving winding. Since the local gravitational field is the only force acting on the winding, the rate at which the flux was cut is constant. The induced current, therefore, is directly proportional to the flux in the air gap. The Hibbert standard is a *secondary* standard and must be calibrated against the mutual inductance method described earlier.

3-6 Standards of Temperature and Luminous Intensity

Thermodynamic temperature is one of the basic SI quantities whose unit is the degree Kelvin (Sec. 2-2). The thermodynamic Kelvin scale is recognized as the *fundamental scale* to which all temperatures should be referred. The temperatures on this scale are designated as °K and denoted by the symbol T. The magnitude of the degree Kelvin has been fixed by defining the thermodynamic temperature of the *triple point* of water at *exactly* 273.16°K. The triple point of water is the temperature of equilibrium between ice, liquid water, and its vapor.

Since temperature measurements on the thermodynamic scale are inherently difficult, the Seventh General Conference of Weights and Measures adopted in 1927 a *practical scale* which has been modified several times and is now called the *International Practical Scale of Temperature.* The temperatures on this scale are designated as °C (degree Celsius) and denoted by the symbol t. The Celsius scale has two *fundamental* fixed points: the boiling point of water as 100°C and the triple point of water as 0.01°C, both points established at atmospheric pressure. A number of *primary* fixed points have been established above and below the two fundamental points. These are the boiling point of oxygen (−182.97°C), the boiling point of sulphur (444.6°C), the freezing point of silver (960.8°C), and the freezing point of gold (1063°C). The numerical values of all these points are reproducible quantities at atmospheric pressure. The conversion between the Kelvin scale and the Celsius scale follows the relationship:

$$t(^\circ\mathrm{C}) = T(^\circ\mathrm{K}) - T_0 \tag{3-15}$$

where $T_0 = 273.15$ degrees

The primary *standard thermometer* is a platinum resistance thermometer of special construction so that the platinum wire is not subjected to strain. Interpolated values between the fundamental and primary fixed points on the scale are calculated by formulas based on the properties of the platinum resistance wire.

The primary *standard of luminous intensity* is a full radiator (black body or Planckian radiator) at the temperature of solidification of platinum (2,042°K approx.). The *candela* is then defined as one-sixtieth of the luminous intensity per cm^2 of the full radiator. Secondary standards of luminous intensity are special tungsten filament lamps, operated at a temperature whereby their spectral power distribution in the visible region matches that of the basic standard. These secondary standards are recalibrated against the basic standard at periodic intervals. The International Bureau of Weights and Measures manufactures national standards at intervals of three years.

References

1. Kaye, G. W. C., and T. H. Laby, *Tables of Physical and Chemical Constants,* 13th ed. London: Longmans, Green and Co., Ltd., 1966.

2. Philco Technological Center, *Electronic Precision Measurement Techniques and Experiments.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

3. Stout, Melville B., *Basic Electrical Measurements,* 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

Questions

1. Describe briefly the differences between primary and secondary standards in terms of their accuracy and their use.

2. What is meant by the "atomic time scale"? How is this time scale related to UT_2?

3. NBS radio stations WWV and WWVB transmit standard time signals which may be used in the calibration of laboratory equipment, such as clocks and counters. Briefly describe the kind of services offered by these radio stations and indicate how the transmitted signal can be traced back to the primary time standard.

4. Name several precautions which should be taken when using a standard Weston cell.

5. What is the emf of a normal Weston cell at 20°C and how much does the emf change when the cell is used at 0°C?

6. You are asked to determine the internal resistance of an unsaturated Weston cell. Describe a method which would supply the correct answer.

7. You suspect that the emf of one of the standard cells in the calibration laboratory may be in error by a fairly large amount. You would like to check this but realize that an ordinary voltmeter would draw too much current and would most certainly damage the cell. What circuit arrangement can you think of to perform this measurement?

8. A time code generator contains a precision oscillator which must be checked daily against a standard frequency transmission of radio station WWV. With the aid of a block diagram, explain how this could be done.

9. Describe briefly the construction of the primary standards for the absolute ohm and the henry.

FOUR

DIRECT-CURRENT INDICATING INSTRUMENTS

4-1 Introduction

Early measurements of direct current (dc) required a suspension galvanometer. This instrument was the forerunner of the moving-coil instrument, basic to most dc indicating movements currently used. Figure 4-1 shows the construction of a *suspension* galvanometer.

A *coil* of fine wire is suspended in a magnetic field produced by a *permanent magnet*. According to the fundamental law of electromagnetic force, the coil will rotate in the magnetic field when it carries an electric current. The fine filament suspension of the coil serves to carry current to and from it, and the elasticity of the filament sets up a moderate torque

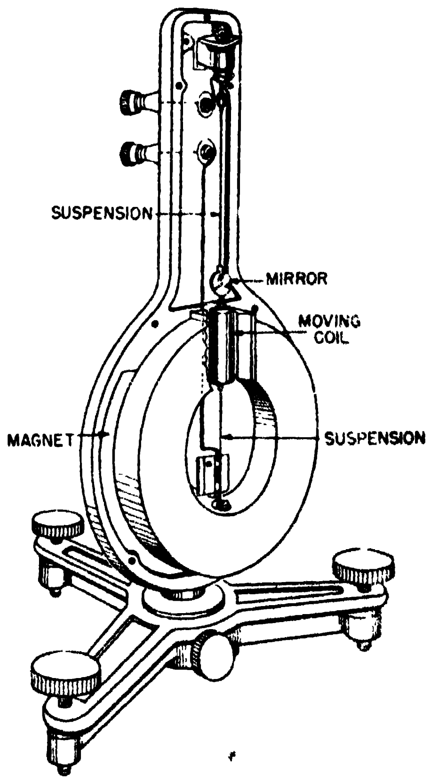


Figure 4-1
An early suspension galvanometer (*Courtesy Weston Instruments, Inc.*).

in opposition to the rotation of the coil. The coil will continue to deflect until its electromagnetic torque balances the mechanical countertorque of the suspension. The coil deflection, therefore, is a measure of the magnitude of dc current carried by the coil. A *mirror* attached to the coil deflects a beam of light, causing a (magnified) light spot to move on a *scale* at some distance from the instrument. The optical effect is that of a pointer of great length but zero mass.

With modern refinements, the suspension galvanometer is still used in certain laboratory measurements requiring extreme sensitivity, but where the delicacy of the instrument is not objectionable and portability is not required.

4-2 Torque and Deflection of the dc Current Galvanometer

Although the suspension galvanometer is neither a practical nor portable instrument, the principles which govern its operation apply equally to its more modern version, the *permanent-magnet moving-coil mechanism* (PMMC). Figure 4-2 shows the construction of the PMMC mechanism. The different parts of the instrument are identified alongside the figure.

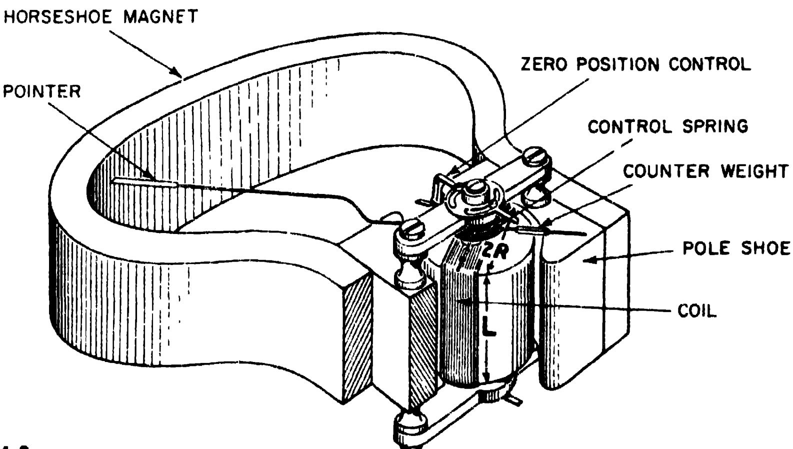


Figure 4-2
Construction details of the external magnet PMMC movement (*Courtesy Weston Instruments, Inc.*).

Here again we have a coil, suspended in the magnetic field of a permanent magnet, this time in the shape of a horseshoe. The coil is suspended so that it can rotate freely in the magnetic field. When current flows in the coil, the developed electromagnetic (EM) torque causes the coil to rotate. The EM torque is counterbalanced by the 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

$$T = B \times A \times I \times N \tag{4-1}$$

where T = torque, newton-meter (N-m)
B = flux density in the air gap, webers/square meter (Wb/m^2)
A = effective coil area, square meters (m^2)
I = current in the movable coil, amperes (A)
N = turns of wire on the coil.

Equation (4-1) shows that the developed torque is directly proportional to the flux density of the field in which the coil rotates, the current in the coil, and the coil constants (area and turns). Since both flux density and coil area are *fixed* parameters for a given instrument, the developed torque is a direct indication of the current in the coil. The pointer deflection, therefore, can be used to measure the current. Equation (4-1) also shows that the designer may vary only the value of the control torque and the number of turns on the moving coil to measure a given full-scale current. The practical coil area generally ranges from approximately 0.5 to 2.5 cm^2. Flux densities for modern instruments usually range from 1,500 to 5,000 gauss (0.15 to 0.5 Wb/m^2). Thus, a wide choice of mechanisms is available to the designer to meet many different measurement applications.

A typical panel PMMC instrument, with a $3\frac{1}{2}$-in. case, a 1-mA range, and full-scale deflection of 100 degrees of arc, would have the following characteristics:*

$$\begin{aligned} A &= 1.75 \text{ cm}^2 \\ B &= 2{,}000 \text{ G } (0.2 \text{ Wb/m}^2) \\ N &= 84 \text{ turns} \\ T &= 2.92 \times 10^{-6} \text{ N-m} \\ \text{Coil resistance} &= 88\ \Omega \\ \text{Power dissipation} &= 88\ \mu\text{W} \end{aligned}$$

The developed torque, expressed by Eq. (4-1), causes the pointer to deflect to a *steady-state* position where it is balanced by the opposing control-spring torque. The initial or *transient* current in the instrument coil will cause it to respond to this sudden change of state according to the laws which govern the response of any moving system characterized by a given mass and stiffness. The second-order differential motion-equation, describing the behavior of the instrument to a sudden current change, is

$$T = J\frac{d^2\theta}{dt^2} + D\frac{d\theta}{dt} + S\theta \tag{4-2}$$

**Data Sheets*, Weston Instruments, Inc., Newark, N.J.

where T = actuating torque of Eq. (4-1)
J = moment of inertia of the movable system
D = damping coefficient of the system
S = spring constant of the control springs
θ = angular deflection of the coil.

The first term on the right-hand side of Eq. (4-2) is the torque required to overcome the *inertia* of the moving-coil system. The second term is the torque required to overcome any *damping* (friction in the system, air-damping, magnetic damping). The third term is the torque required to overcome the restoring *countertorque* of the suspension springs.

Three specific and possible solutions of the differential equation, Eq. (4-2), may be obtained, depending on the relative magnitudes of the damping coefficient, D, the spring constant, S, and the moment of inertia, J. Equation (4-2) is satisfied by solutions of the form

$$\theta = Ke^{mt} \tag{4-3}$$

where e is the base of the natural logarithms and K and m are constants. Substituting Eq. (4-3) into Eq. (4-2) yields the characteristic equation

$$Jm^2 + Dm + S = 0 \tag{4-4}$$

Solving Eq. (4-4) by the quadratic formula gives the two roots m_1 and m_2, where

$$m_{1,2} = -\frac{D}{2J} \pm \sqrt{\frac{D^2}{4J^2} - \frac{S}{J}} \tag{4-5}$$

The *type of behavior* of the moving system can be determined from the form of the roots m_1 and m_2 of Eq. (4-5). If the quantity under the radical in Eq. (4-5) equals *one*, the two roots are real and equal, and we can write

$$m_1 = m_2 = -\frac{D}{2J} \tag{4-6}$$

This condition is fullfilled when $D^2/4J^2 = S/J$, or the effective damping constant of the system

$$D = 2\sqrt{JS} \tag{4-7}$$

Equation (4-7) represents the *critical value* of the damping constant D and is generally given by

$$D' = 2\sqrt{JS} \tag{4-8}$$

By definition, the *damping ratio*, ζ, of a movable system is the ratio of its damping constant D to the critical damping constant D'. Therefore,

$$\text{damping ratio } \zeta = \frac{D}{D'} = \frac{D}{2\sqrt{JS}} \tag{4-9}$$

The solution of the differential equation of Eq. (4-4) can be approached using the damping ratio, ζ. If $\zeta = 1$, the roots of Eq. (4-5) are real and equal and the system is *critically damped*. If $\zeta < 1$ and positive, the roots of Eq. (4-5) are complex. The transient movement of the system is then a damped sinusoid and the system is called *underdamped*. If $\zeta > 1$ and positive, the roots of Eq. (4-5) are real. The transient solution consists of two exponential terms and the system is called *overdamped*.

The *undamped natural frequency* ω_n is, by definition, the frequency of oscillation of the transient if the damping is zero. The radical in Eq. (4-5) determines the undamped natural frequency and is given by

$$\omega_n = \sqrt{\frac{S}{J}}\,\frac{\text{rad}}{\text{s}} \quad \text{or} \quad f_n = \frac{1}{2\pi}\sqrt{\frac{S}{J}} \tag{4-10}$$

and the time of one complete oscillation is

$$T_n = \text{undamped period} = 2\pi\sqrt{\frac{J}{S}} \tag{4-11}$$

The quadratic factors of Eq. (4-4) are frequently written in terms of the damping ratio, ζ, and the undamped natural frequency, ω_n. (The underdamped system is generally analyzed in terms of these two parameters.) By factoring S, the quadratic factors of the characteristic equation of Eq. (4-4) become

$$\frac{J}{S}m^2 + \frac{D}{S}m + 1 = 0 \tag{4-12}$$

Substituting Eq. (4-9) and Eq. (4-10) into Eq. (4-12) yields

$$\frac{1}{\omega_n^2}m^2 + \frac{2\zeta}{\omega_n}m + 1 = 0 \tag{4-13}$$

Multiplying Eq. (4-13) by ω_n^2 yields

$$m^2 + 2\zeta\omega_n m + \omega_n^2 = 0 \tag{4-14}$$

Equations (4-13) and (4-14) are the *standard forms* of the equation describing the motion of a galvanometer responding to a sudden change of state caused by the application of a current. For the three stated possibilities of the damping ratio, ζ, a solution may be obtained for the value of the deflecting angle θ in Eq. (4-2), in terms of the undamped natural frequency ω_n and the damping ratio ζ.

Figure 4-3 gives *nondimensional curves* of galvanometer motion. These curves are general, not limited to a given set of data, and represent a function to which the constants may be inserted to obtain a particular solution. The ordinate represents the ratio θ/θ_F, where θ is the instantaneous deflecting angle of the galvanometer and θ_F the final, steady-state deflecting angle. The abscissa represents the ratio t/T_n, where t is the time of observation and T_n the natural undamp-

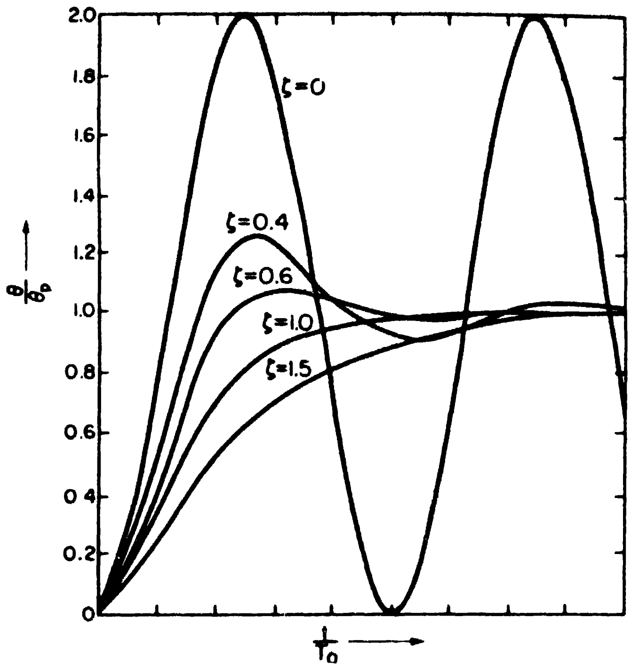


Figure 4-3
Nondimensional curves of galvanometer motion. θ/θ_p represents the ratio of the deflection angle θ and the zero-current angle θ_p (rest position). t/T_0 represents the ratio of the time t after application of the impulse and the natural period T_0 of oscillation. ζ is the ratio of the actual damping D to the critical damping D_r. Since these functions are nondimensional they apply to any galvanometer.

ed period of the galvanometer. Both coordinates are therefore *dimensionless* and the family of curves applies to any galvanometer. Each curve is plotted for a certain value of damping ratio, ζ, which is also a dimensionless quantity.

To find the motion of a certain galvanometer as a function of time, the ordinate θ/θ_p is multiplied by θ_p and the abscissa t/T_n is multiplied by T_n. In the *underdamped* case, when $\zeta < 1$, the pointer attached to the movable coil oscillates back and forth several times *above and below* its final steady-state position, before coming to rest. In the *critically damped* case, the pointer moves promptly up to its steady-state position with *no oscillation.* Critical damping, when $\zeta = 1$, is the ideal behavior for a PMMC movement. In the *overdamped* case, when $\zeta > 1$, the pointer tends to approach the steady-state position asymptotically but has the appearance of being *sluggish* and never actually reaches its "true" position (i.e., where the deflection is proportional to the current in the coil).

Ideally, as just noted, the response of the instrument should be such that the pointer travels to its true final position without any overshoot: the movement should be critically damped. In practice, however, the instrument is *slightly underdamped,* causing the pointer to overshoot a little before coming to its steady-state position. This method decreases the response time, but assures the user that the instrument has not been damaged owing to rough handling and compensates for any additional friction which may develop in time because of dust, wear, etc.

PMMC instruments are constructed to produce as little *viscous damping* (virtually frictionless) as possible and the desired degree of damping is then added. Different mechanisms are used to provide sufficient damping of the PMMC. One of the simplest methods involves the use of an aluminum vane, attached to the shaft of the moving coil, and rotating in an air chamber. The amount of clearance between the chamber walls and the air vane effectively controls the amount of damping. Some instruments use the principle of electromagnetic damping (Lenz's law), where the movable coil is wound on a light aluminum frame. The deflection of the coil (and frame) in the magnetic field of the permanent magnet sets up circulating currents in the conductive metal frame, causing a retarding torque which opposes the motion of the coil. Indeed, during shipment, the same principle is employed to protect PMMC instruments by placing a metal shorting strap across the coil terminals to reduce deflection.

A galvanometer or PMMC instrument may also be damped by connecting a resistor across the coil. When the coil is energized by a current and rotates in the magnetic field, a voltage is generated in the coil and this then circulates a current through the coil and the external resistance to produce an opposing or retarding torque. The opposing torque damps the motion of the movement. For any galvanometer, a value for the external resistor can be found that produces critical damping (Fig. 4-3). This resistance is called the *Critical Damping Resistance External* (CDRX) and is an important galvanometer constant. The dynamic damping torque produced by the CDRX depends on the total circuit resistance: the smaller the total circuit resistance, the larger the damping torque.

One way to determine the CDRX consists of observing the galvanometer swing when a current is applied or removed from the coil. Beginning with the oscillating condition, decreasing values of external resistances are tried until a value is found for which the overshoot just disappears. A determination like this is not very precise, but is adequate for most practical purposes. The value of the CDRX may be computed from known galvanometer constants. (For further information, consult the reference list at the end of this chapter.)

4-3 The Permanent Magnet Moving-Coil Mechanism (PMMC)

Figure 4-2 represents one of the basic models of the PMMC. This movement is often called the *d'Arsonval movement,* after its inventor. This design offers the largest magnet in a given space and is used where maximum flux in the air gap is required. It provides an instrument with very low power consumption and low current required for *full-scale deflection* (fsd). Figure 4-4 shows a *phantom view* of this instrument.

Inspection of the photograph of Fig. 4-4 shows the *permanent magnet* of horseshoe form, 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 cylinder. The *coil* is wound

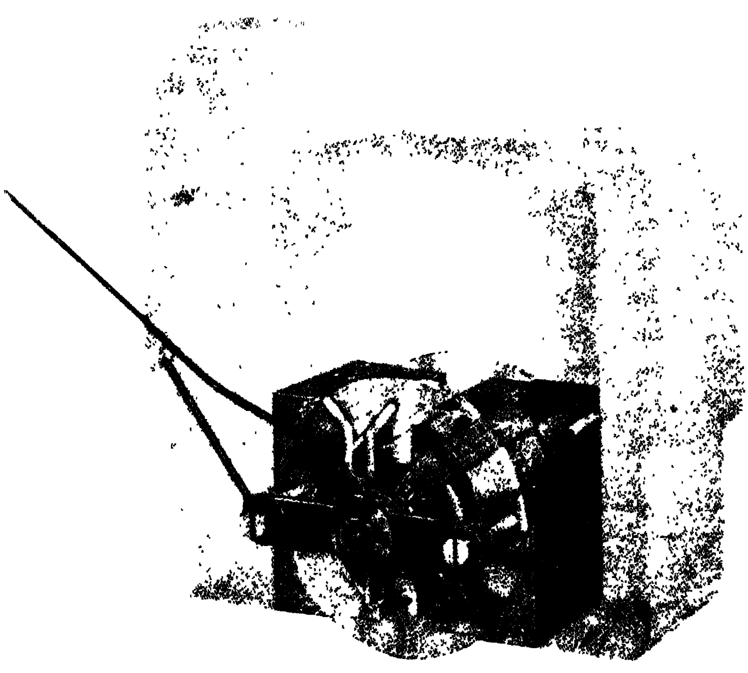


Figure 4-4
This phantom photograph of the external magnet moving-coil mechanism shows details of the coil construction, the external horseshoe magnet, and the indicating pointer (*Courtesy Weston Instruments Inc.*).

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 and therefore the current through the coil.

The Y-shaped member is the *zero adjust* control and is connected to the fixed end of the *front control-spring.* An *eccentric pin* through the instrument cover engages the Y-shaped member so that the zero position of the pointer can be adjusted from outside the case. Two phosphor-bronze *conductive springs,* normally equal in strength, provide the calibrated force opposing the moving-coil torque. Constancy of spring performance is essential to maintain instrument accuracy. The spring thickness is accurately controlled in manufacture in order to avoid permanent set of the springs. Current is conducted to and from the coil by the control springs. Construction details of a typical coil for a PMMC system are shown in Fig. 4-5.

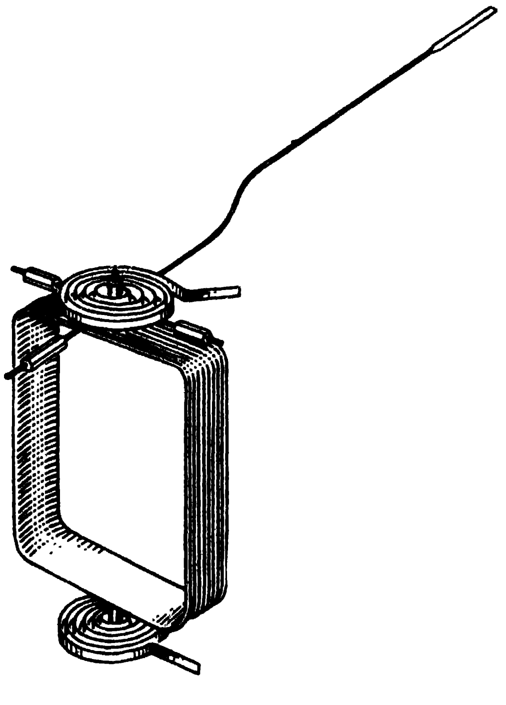
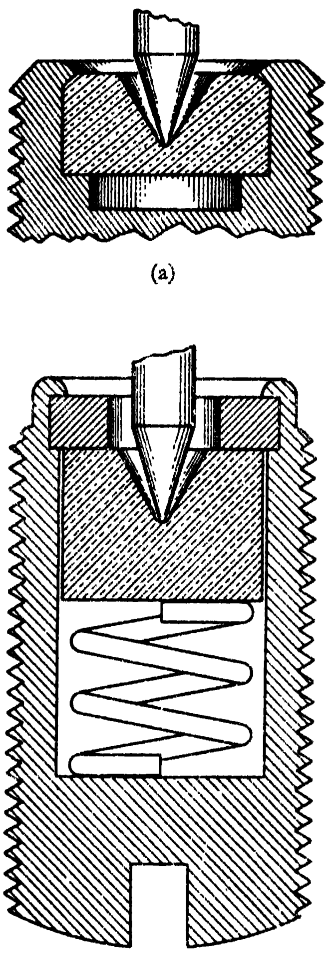


Figure 4-5
Details of a moving coil for a PMMC movement, showing the control springs and the indicator with its counterbalance weights (*Courtesy Weston Instruments Inc.*).

Most voltmeter coils are wound on metal frames to provide the required electromagnetic damping. The metal frame constitutes a short-circuited turn in a magnetic field. Most ammeter coils, however, are wound on a nonconductive frame, because the coil turns are effectively shorted out by the ammeter shunt. The coil itself, therefore, provides the EM damping.

The entire moving system is statically balanced for all deflection positions by three *balance weights,* as shown in Fig. 4-5. The pointer, springs, and pivots are assembled to the coil structure by means of pivot bases, and the entire movable-coil element is supported by jewel bearings. Different types of bearing systems are shown in Fig. 4-6.

The V-jewel, shown in Fig. 4-6(a), is almost universally used in instrument bearings. The pivot, bearing in the pit in the jewel, may have a radius at its tip from 0.003 in. to 0.0005 in., depending on the weight of the mechanism and the vibration the instrument will encounter. The radius of the pit in the jewel is slightly larger than the pivot radius, so that the contact area is circular, a fraction of a thousandth of an inch across. The V-jewel design of Fig. 4-6(a) has the least friction of any practical type of instrument bearing. Although the moving elements of instruments are designed to have the smallest possible weight, the extremely minute area of contact between pivot and jewel results in enormous stresses (force per unit area). For example, a moving element weighing 300 mg resting on an area of a circle of 0.0002 in. diameter, produces a stress of approximately 10 tons/in.^{2}. If the weight of the moving element is further increased, the contact area does not increase in proportion so that the stress is even greater. Stresses set up by relatively moderate accelerations (like jarring or dropping an instrument) may consequently cause pivot damage. Specially protected (*ruggedized*) instruments use the spring-back (incabloc) jewel bearing, whose construction is shown in Fig. 4-6(b). It is located in its normal position by the spring and is free to move axially when the shock to the mechanism becomes severe.



(a)

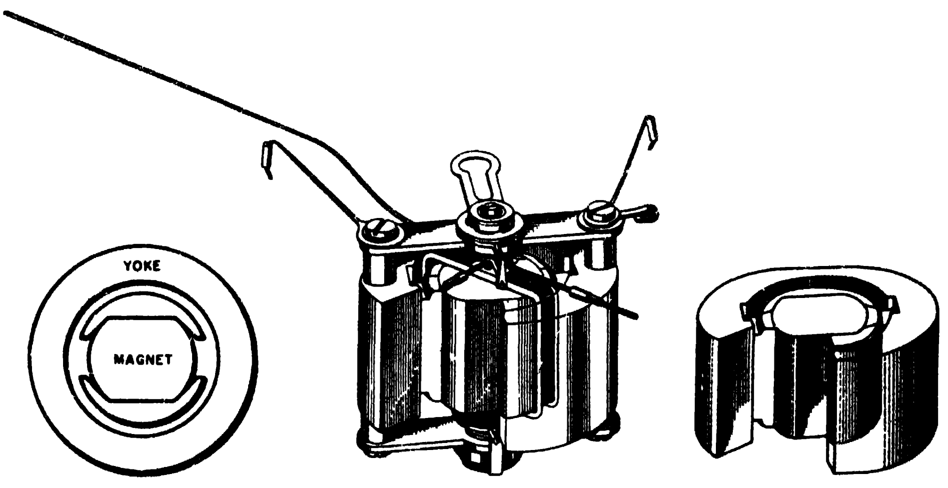
(b)

Figure 4-6
Details of instrument bearings. (a) The V-jewel bearing. (b) The spring-back jewel bearing. (*Courtesy Weston Instruments Inc.*)

The scale markings of the basic dc PMMC instrument are usually linearly spaced because the torque (and hence the pointer deflection) is directly proportional to the coil current. (See Eq. (4-1) for the developed torque.) The basic PMMC instrument is therefore a *linear-reading dc device*. The power requirements of the d'Arsonval movement are surprisingly small: typical values range from 25 μW to 200 μW. Accuracy of the instrument is generally in the order of 2 per cent of full-scale reading.

If low-frequency alternating current is applied to the movable coil, the deflection of the pointer would be up-scale for one half-cycle of the input waveform and down-scale (in the opposite direction) for the next half-cycle. At powerline frequencies (60 Hz) and above, the pointer cannot follow the rapid variations in direction and will quiver slightly around the zero mark, seeking the *average value* of the alternating current (which equals zero). The PMMC instrument is therefore unsuitable for ac measurements, unless suitable rectification is provided before application to the coil (see Sec. 5-4).

In recent years, with the development of Alnico (and other improved magnetic materials), it has become feasible to design a magnetic system in which the magnet itself serves as the core. These magnets have the obvious advantage of being relatively unaffected by external magnetic fields, eliminating the magnetic shunting effects in steel panel construction, where several meters operating side by side may affect each other's readings. The need for magnetic shielding, in the form of iron cases, is also eliminated by the *core-magnet* construction. Details of the core-magnet self-shielded movement are shown in Fig. 4-7.



(a) (b) (c)

Figure 4-7

Construction details of the core-magnet moving-coil mechanism. (a) The magnet with its poleshoes is surrounded by the yoke, which acts as a magnetic shield. (b) The assembled movement. (c) A cut-away view of the yoke, the core, and the poleshoes. (*Courtesy Weston Instruments Inc.*)

Self-shielding makes the core-magnet mechanism particularly useful in aircraft and aerospace applications, where a multiplicity of instruments must be mounted in close proximity to each other. An example of this type of mounting may be found in the *cross-pointer indicator,* where as many as five mechanisms are mounted in one case to form a unified display. Obviously, the elimination of iron cases and the corresponding weight reduction are of great advantage in aircraft and aerospace instruments.

The *suspension-type* galvanometer mechanism has been known for many years. Until recently, the device was used only in the laboratory where high sensitivities were required and the torque was extremely low (because of small currents). It was desirable in such instruments to eliminate even the low friction of pivots and jewels. The suspension galvanometer (Fig. 4-1) had to be used in the upright position, because sag in the low-torque ligaments caused the moving system to come in contact with stationary members of the mechanism in any other position. This increase in friction caused errors.

The *taut-band* instrument (Fig. 4-8) has the advantage of eliminating the friction of the jewel-pivot suspension. The ribbons are placed under sufficient tension to eliminate any sag, as was the case in the suspension galvanometer of Fig. 4-1. This tension is provided by a tension-spring, so that the instrument can be used in any position. Generally speking, taut-band suspension instru-

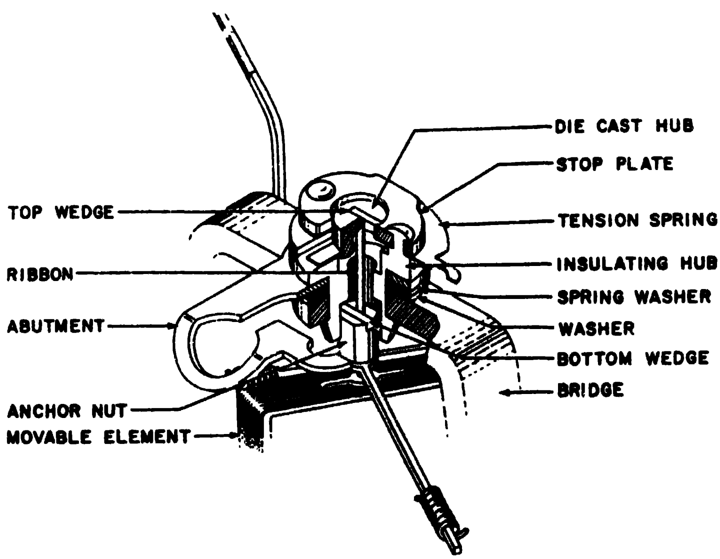


Figure 4-8
The taut-band suspension system, eliminating the friction of conventional pivot-and-jewel type suspensions. This figure shows some of the construction details, in particular the torsion ribbon with its tension spring mechanism (*Courtesy Weston Instruments Inc.*).

ments can be made with higher sensitivities than those using pivots and jewels and they can be used in almost every application which is presently served by pivoted instruments. Furthermore, taut-band instruments are relatively insensitive to shock and temperature, and are capable of withstanding greater overloads than previous types described.

The PMMC basic movement is not inherently insensitive to temperature, but it may be *temperature-compensated* by the appropriate use of series- and shunt-resistors of copper and manganin. Both the magnetic fieldstrength and spring tension decrease with an increase in temperature. The coil resistance increases with an increase in temperature. These changes (with increased temperature in the PMMC movement), tend to make the pointer read low for a given current with respect to magnetic fieldstrength and coil resistance. The spring change, conversely, tends to cause the pointer to read high with an increase in temperature. The effects are not identical, however; hence an *uncompensated meter tends to read low* by approximately 0.2 per cent per °C rise in temperature. For purposes of instrument specification, the movement is considered to be

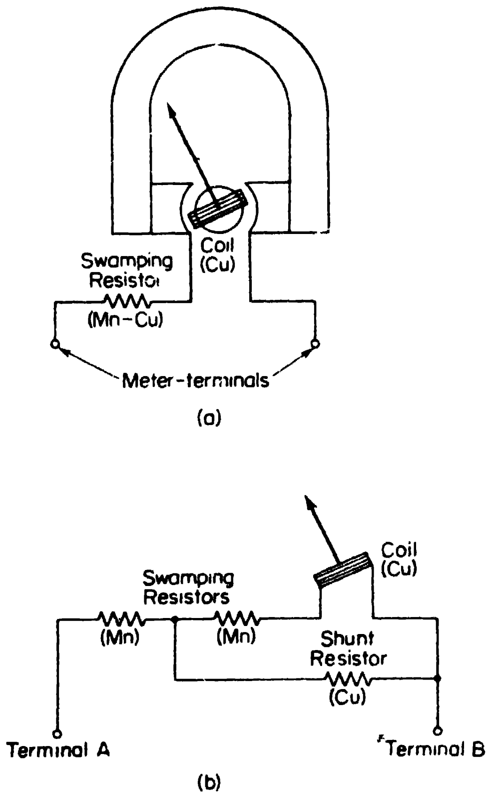


Figure 4-9
Placement of swamping resistors for temperature compensation of a meter movement. (a) Basic circuit. (b) Improved compensation.

compensated when the change in accuracy, due to a 10°C change in temperature, is not more than one-fourth of the total allowable error.*

Compensation may be accomplished by using *swamping resistors* in series with the movable coil, as shown in Fig. 4-9(a). The swamping resistor is made of manganin (which has a temperature coefficient of practically zero) combined with copper in the ratio of 20/1 to 30/1. The total resistance of coil and swamping resistor increases slightly with a rise in temperature, but only just enough to counteract the change of springs and magnet, so that the over-all temperature effect is zero.

A more complete cancellation of temperature effects is obtained with the arrangement of Fig. 4-9(b). Here the *total* circuit resistance increases slightly with a rise in temperature, owing to the presence of the copper coil and the copper shunt-resistor. For a fixed applied voltage, therefore, the total current decreases slightly with a rise in temperature. The resistance of the copper shunt-resistor increases more than the series combination of coil and manganin resistor, hence a larger fraction of the total current passes through the coil circuit. By correct proportioning of the copper and manganin parts in the circuit, complete cancellation of temperature effects may be accomplished. One *disadvantage* of the use of swamping resistors is a reduction in the full-scale sensitivity of the movement, because a higher applied voltage is necessary to sustain the full-scale current.

4-4 Definitions of Galvanometer Sensitivity

Three different kinds of sensitivity definitions are generally used in specifying the sensitivity of a galvanometer: (a) current sensitivity, (b) voltage sensitivity, (c) megohm sensitivity.

Current sensitivity may be defined as the ratio of the deflection of the galvanometer to the current producing this deflection. The current is usually expressed in μA and the deflection in millimeters. For galvanometers which do not have a scale calibrated in millimeters, the deflection may be given in scale divisions. The current sensitivity is

$$S_I = \frac{d}{I} \quad \frac{\text{mm}}{\mu\text{A}} \tag{4-15}$$

where d = deflection of the galvanometer in scale divisions or in mm
I = galvanometer current, in μA.

Voltage sensitivity may be defined as the ratio of the galvanometer deflection to the voltage producing this deflection. Therefore,

*PMMC *Data Sheets,* Weston Instruments, Inc., Newark, N.J.

$$S_V = \frac{d}{V} \quad \frac{\text{mm}}{\text{mV}} \tag{4-16}$$

where d = deflection of the galvanometer in scale divisions or in mm
V = voltage applied to the galvanometer, in mV.

It is customary to consider the galvanometer together with its critical damping resistance (CDRX), and most manufacturers specify the voltage sensitivity of a galvanometer in mm/mV across the CDRX.

Megohm sensitivity may be defined as the number of megohms required in series with the (CDRX shunted) galvanometer to produce one scale division deflection when 1 V is applied to the circuit. Since the equivalent resistance of the shunted galvanometer is negligible compared with the number of megohms in series with it, the applied current practically equals 1/R μA and produces one division deflection. Numerically, therefore, the megohm sensitivity is equal to the current sensitivity and therefore

$$S_R = \frac{d}{I} = S_I \quad \frac{\text{mm}}{\mu\text{A}} \tag{4-17}$$

where d = deflection of the galvanometer in scale divisions or in mm
I = galvanometer current, in μA.

A fourth sensitivity figure is used with ballistic galvanometers. It is called the *ballistic sensitivity* and is defined as the ratio of the maximum deflection, d_m, of a galvanometer to the quantity Q of electric charge in a single pulse which produces this deflection. Then

$$S_Q = \frac{d_m}{Q} \quad \frac{\text{mm}}{\mu\text{C}} \tag{4-18}$$

where d_m = maximum galvanometer deflection in scale divisions or in mm
Q = quantity of electricity, in μC.

Example 4-1 illustrates a procedure for testing a galvanometer.

Example 4-1: A galvanometer is tested in the circuit of Fig. 4-10

where $E = 1.5$ V
$R_1 = 1.0\ \Omega$
$R_2 = 2500\ \Omega$
R_3 is variable.

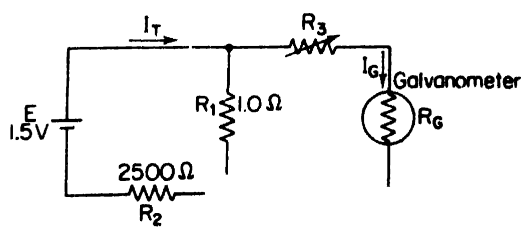


Figure 4-10
A circuit for testing a galvanometer.

With R_3 set at 450 Ω, the galvanometer deflection is 150 mm, and with R_3 set at 950 Ω, the deflection is reduced to 75 mm. Calculate: (a) the resistance of the galvanometer; (b) the current sensitivity of the galvanometer.

SOLUTION:

(a) The fraction of the total current I_T taken by the galvanometer equals

$$I_G = \frac{R_1}{R_1 + R_3 + R_G} \times I_T$$

Since for $R_3 = 450\ \Omega$, the deflection is 150 mm and for $R_3 = 950\ \Omega$, the deflection is 75 mm, the galvanometer current I_G in the second case is one-half of the galvanometer current in the first case. Therefore, we can write

$$I_{G_1} = 2I_{G_2} \quad \text{or} \quad \frac{1.0}{1.0 + 450 + R_G} = 2\frac{1.0}{1.0 + 950 + R_G}$$

and solving for R_G yields $R_G = 49\ \Omega$.

(b) Inspection of the circuit of Fig. 4-10 indicates that the total circuit resistance, R_T, is

$$R_T = R_2 + \frac{R_1(R_3 + R_G)}{R_1 + R_3 + R_G} \approx 2500\ \Omega$$

and therefore,

$$I_T = \frac{1.5\ \text{V}}{2500\ \Omega} = 0.6\ \text{mA}$$

For $R_3 = 450\ \Omega$, the galvanometer current I_G is

$$I_{G_1} = \frac{R_1}{R_1 + R_3 + R_G} I_T$$

$$= \frac{1.0}{1.0 + 450 + 49} \times 0.6\ \text{mA} = 1.2\ \mu\text{A}$$

and

$$S_I = \frac{150\ \text{mm}}{1.2\ \mu\text{A}} = 125\ \text{mm}/\mu\text{A}$$

4-5 Ammeter Shunts and Multirange Ammeters

The *basic movement* of a *dc ammeter* is a PMMC d'Arsonval galvanometer. The coil winding of a basic movement is small and light, and can carry only very small currents. It is necessary to extend the current range of dc ammeters. When high currents are to be measured, the major part of the current bypasses the basic movement through a resistance, called a *shunt*. Figure 4-11 shows the basic movement and its shunt to produce an *ammeter*.

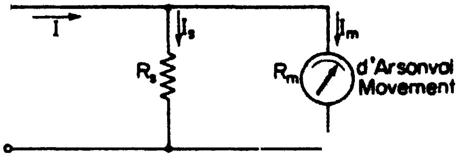


Figure 4-11
The basic dc ammeter circuit.

The resistance of the shunt can be calculated using conventional circuit analysis. See Fig. 4-11,

where R_m = internal resistance of the movement (the coil)
R_s = resistance of the shunt
I_m = full-scale deflection current of the movement
I_s = shunt current
I = full-scale current of the ammeter including the shunt.

Since the shunt resistance is in parallel with the meter movement, the voltage drops across shunt and movement must be the same and we can write

$$V_{\text{shunt}} = V_{\text{movement}}$$

or

$$I_s R_s = I_m R_m \quad \text{and} \quad R_s = \frac{I_m R_m}{I_s} \tag{4-19}$$

Since $I_s = I - I_m$, we can write

$$R_s = \frac{I_m R_m}{I - I_m} \tag{4-20}$$

For each required value of full-scale meter current we can then solve for the value of the shunt resistance required.

Example 4-2: A 1-mA meter movement with an internal resistance of 100 Ω is to be converted into a 0–100 mA ammeter. Calculate the value of the shunt resistance required.

SOLUTION: $I_s = I - I_m = 100 - 1 = 99$ mA

$$R_s = \frac{I_m R_m}{I_s} = \frac{1 \text{ mA} \times 100\ \Omega}{99 \text{ mA}} = 1.01\ \Omega$$

The shunt resistance used with a basic movement may consist of a length of constant-temperature resistance wire within the case of the instrument or it may be an external (manganin or constantan) shunt having a very low resistance. Figure 4-12 shows an external shunt. It consists of evenly spaced sheets of resistive material welded into a large block of heavy copper on each end of the sheets. The resistance material has a very low temperature coefficient, and a low thermoelectric effect exists between the resistance material and the cop-

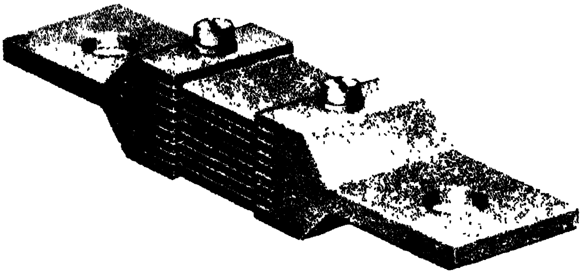


Figure 4-12
High current switchboard instrument shunt (*Courtesy Weston Instruments Inc.*).

per. External shunts of this type are normally used for measuring very large currents.

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. Figure 4-13 shows the schematic diagram of a multirange ammeter. The circuit has four shunts, R_a, R_b, R_c, and R_d, which can be placed in parallel with the movement to give four different current ranges. Switch S is a multiposition, *make-before-break* type switch, so that the movement will not be damaged, unprotected in the circuit, without a shunt as the range is changed.

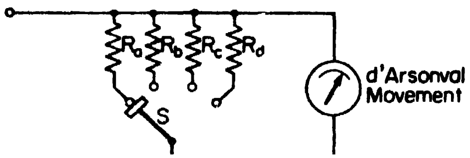


Figure 4-13
Schematic diagram of a simple multirange ammeter.

The *universal shunt* or *Ayrton shunt* of Fig. 4-14 eliminates the possibility of the meter being in the circuit without a shunt. This advantage is gained at the price of a slightly higher over-all meter resistance. The Ayrton shunt provides an excellent opportunity to apply basic network theory to a practical circuit.

Example 4-3: Design an Ayrton shunt to provide an ammeter with current ranges of 1 A, 5 A, and 10 A. A d'Arsonval movement with an internal resistance $R_m = 50\ \Omega$ and full-scale deflection current of 1 mA is used in the configuration of Fig. 4-14.

SOLUTION: *On the 1-A range*: $R_a + R_b + R_c$ are in parallel with the 50-Ω movement. Since the movement requires 1 mA for full-scale deflection, the shunt will be required to pass a current of 1 A − 1 mA = 999 mA. Using Eq. (4-19), we get

$$R_a + R_b + R_c = \frac{1 \times 50}{999} = 0.05005\ \Omega \qquad \text{(I)}$$

On the 5-A range: $R_a + R_b$ are in parallel with $R_c + R_m$ (50 Ω). In this

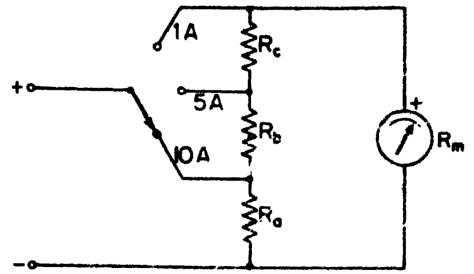


Figure 4-14
The universal shunt or Ayrton shunt.

case there will be a 1-mA current through the movement and R_c in series, and 4,999 mA through $R_a + R_b$. Again using Eq. (4-19),

$$R_a + R_b = \frac{1 \times (R_c + 50\ \Omega)}{4{,}999} \tag{II}$$

On the 10-A range: R_a now serves as the shunt and $R_b + R_c$ are in series with the movement. The current through the movement again is 1 mA and the shunt passes the remaining 9,999 mA. Using Eq. (4-19) again,

$$R_a = \frac{1 \times (R_b + R_c + 50\ \Omega)}{9{,}999} \tag{III}$$

Solving the three simultaneous equations (I, II, and III), we obtain

$$\begin{aligned}
4{,}999 \times \text{(I)}:\quad & 4{,}999\,R_a + 4{,}999\,R_b + 4{,}999\,R_c = 250.2\\
\text{(II)}:\quad & 4{,}999\,R_a + 4{,}999\,R_b - R_c = 50
\end{aligned}$$

Subtracting (II) from (I):

$$\begin{aligned}
5{,}000\,R_c &= 200.2\\
R_c &= 0.04004\ \Omega
\end{aligned}$$

Similarly,

$$\begin{aligned}
9{,}999 \times \text{(I)}:\quad & 9{,}999\,R_a + 9{,}999\,R_b + 9{,}999\,R_c = 500.45\\
\text{(III)}:\quad & 9{,}999\,R_a - R_b - R_c = 50
\end{aligned}$$

Subtracting (III) from (I),

$$10{,}000\,R_b + 10{,}000\,R_c = 450.45$$

Substituting the previously calculated value for R_c into this expression,

$$\begin{aligned}
10{,}000\,R_b &= 450.45 - 400.4\\
R_b &= 0.005005\ \Omega\\
\text{and}\quad R_a &= 0.005005\ \Omega
\end{aligned}$$

This calculation indicates that for larger currents the value of the shunt resistance may become quite small.

Direct-current ammeters are commercially available in a large number of ranges, from 20 μA to 50 A full-scale for a self-contained meter and to 500 A for a meter with external shunt. Laboratory-type precision ammeters are furnished with a calibration chart, so that the user may correct his readings for any scale errors. Some instruments have accurately drawn, hand-calibrated scales, with the scale marking corresponding to accurately known calibration currents.

Observe the following precautions when using an ammeter in measurement work:

(a) *Never* connect an ammeter *across* a source of emf. Because of its low resistance it would draw damaging high currents and destroy the delicate movement. *Always* connect an ammeter in series with a load capable of limiting the current.

(b) Observe the correct *polarity*. Reverse polarity causes the meter to deflect against the mechanical stop and this may damage the pointer.

(c) When using a multirange meter, first use the highest current range, then decrease the current range until substantial deflection is obtained. To increase accuracy of the observation (see Chapter 1), use the range which will give a reading as near to full-scale as possible.

4-6 DC Voltmeters

The addition of a series resistor, or *multiplier*, converts the basic d'Arsonval movement into a *dc voltmeter*, as shown in Fig. 4-15. The multiplier limits the current through the movement so as not to exceed the value of the full-scale-deflection current (I_{fsd}). A dc voltmeter measures the potential difference between two points in a dc circuit and is therefore connected *across* a source of emf or a circuit component. The meter terminals are generally marked "pos" and "neg," since polarity must be observed.

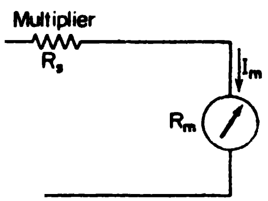


Figure 4-15
The basic dc voltmeter circuit.

The value of a multiplier, required to extend the voltage range, is calculated from Fig. 4-15,

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

For the circuit of Fig. 4-15,

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

Solving for R_s gives

$$R_s = \frac{V - I_m R_m}{I_m} = \frac{V}{I_m} - R_m \tag{4-21}$$

The multiplier is usually mounted inside the case of the voltmeter for moderate ranges up to 500 V. For higher voltages, the multiplier may be mounted separately outside the case on a pair of binding posts to avoid excessive heating inside the case.

The addition of a number of multipliers, together with a *range switch*, provides the instrument with a workable number of voltage ranges. Figure 4-16 shows a *multirange* voltmeter using a four-position switch and four multipliers, R_1, R_2, R_3, and R_4, for the voltage ranges V_1, V_2, V_3, and V_4, respectively. The values of the multipliers can be computed by using the method shown earlier or, alternatively, by the *sensitivity method.* The sensitivity method is illustrated by Example 4-5 (Sec. 4-7) where sensitivity is discussed.

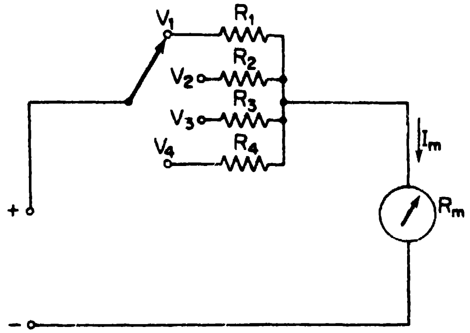


Figure 4-16
A multirange voltmeter.

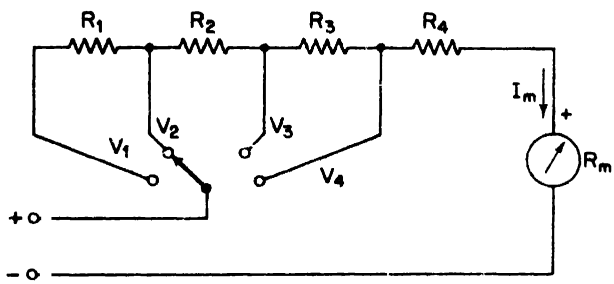


Figure 4-17
A more practical arrangement of multiplier resistors in the multirange voltmeter.

A variation of the circuit of Fig. 4-16 is shown in Fig. 4-17, where the multipliers are connected in a series string and the range selector switches the appropriate amount of resistance in series with the movement. This system has the advantage that all multipliers except the first have standard resistance values and can be obtained commercially in precision tolerances. The low-range multiplier, R_4, is the only special resistor which must be manufactured to meet the specific circuit requirements:

Example 4-4: A basic d'Arsonval movement with internal resistance, $R_m = 100\ \Omega$, and full-scale current, $I_{fsd} = 1$ mA, is to be converted into a multirange dc voltmeter with voltage ranges of 0–10 V, 0–50 V, 0–250 V, and 0–500 V.

The circuit arrangement of Fig. 4-17 is to be used for this voltmeter.

SOLUTION: For the 10-V range (the V_4 position of the range switch), the total circuit resistance is

$$R_T = \frac{10\ V}{1\ \text{mA}} = 10\ \text{k}\Omega$$

$$R_4 = R_T - R_m = 10\ \text{k}\Omega - 100\ \Omega = 9{,}900\ \Omega$$

For the 50-V range (V_3 position of range switch),

$$R_T = \frac{50\ \text{V}}{1\ \text{mA}} = 50\ \text{k}\Omega$$

$$R_3 = R_T - (R_4 + R_m) = 50\ \text{k}\Omega - 10\ \text{k}\Omega = 40\ \text{k}\Omega$$

For the 250-V range (V_2 position of range switch),

$$R_T = \frac{250\ \text{V}}{1\ \text{mA}} = 250\ \text{k}\Omega$$

$$R_2 = R_T - (R_3 + R_4 + R_m) = 250\ \text{k}\Omega - 50\ \text{k}\Omega = 200\ \text{k}\Omega$$

For the 500-V range (V_1 position of range switch),

$$R_T = \frac{500\ \text{V}}{1\ \text{mA}} = 500\ \text{k}\Omega$$

$$R_1 = R_T - (R_2 + R_3 + R_4 + R_m) = 500\ \text{k}\Omega - 250\ \text{k}\Omega = 250\ \text{k}\Omega$$

Notice in Example 4-4 that only the low-range multiplier R_4 has a nonstandard value.

4-7 Voltmeter Sensitivity and Loading Effect

In Sec. 4-6 it was shown that the full-scale deflection current I_{fsd} was reached on all voltage ranges when the corresponding full-scale voltage was applied. As shown in Example 4-4, a current of 1 mA is obtained for voltages of 10 V, 50 V, 250 V, and 500 V across the meter terminals. For each voltage range, the quotient of the total circuit resistance R_T and the range voltage V is always 1,000 Ω/V. This figure is often referred to as the ***sensitivity*** or the ***ohms-per-volt rating*** of the voltmeter. Note that the sensitivity, S, is essentially the ***reciprocal*** of the full-scale deflection current of the basic movement, or

$$S = \frac{1}{I_{fsd}}\ \frac{\Omega}{\text{V}} \tag{4-22}$$

The sensitivity S of the voltmeter may be used to advantage in the *sensitivity method* of calculating the resistance of the multiplier in a dc voltmeter. Consider the circuit of Fig. 4-17,

where S = sensitivity of the voltmeter, in Ω/V
V = the voltage range, as set by the range switch
R_m = internal resistance of the movement (plus the previous series resistors)
R_s = resistance of the multiplier.

For the circuit of Fig. 4-17,

$$R_T = S \times V$$

and

$$R_s = (S \times V) - R_m \tag{4-23}$$

Use of the sensitivity method is illustrated in Example 4-5.

Example 4-5: Repeat Example 4-4, now using the sensitivity method for calculating the multiplier resistances.

SOLUTION:

$$S = \frac{1}{I_{fsd}} = \frac{1}{0.001\ \text{A}} = 1{,}000\frac{\Omega}{V}$$

$$R_4 = (S \times V) - R_m = \frac{1{,}000\ \Omega}{\text{V}} \times 10\ \text{V} - 100\ \Omega = 9{,}900\ \Omega$$

$$R_3 = (S \times V) - R_m = \frac{1{,}000\ \Omega}{\text{V}} \times 50\ \text{V} - 10{,}000\ \Omega = 40\ \text{k}\Omega$$

$$R_2 = (S \times V) - R_m = \frac{1{,}000\ \Omega}{\text{V}} \times 250\ \text{V} - 50\ \text{k}\Omega = 200\ \text{k}\Omega$$

$$R_1 = (S \times V) - R_m = \frac{1{,}000\ \Omega}{\text{V}} \times 500\ \text{V} - 250\ \text{k}\Omega = 250\ \text{k}\Omega$$

The sensitivity of a dc voltmeter is an important factor when selecting a meter for a certain voltage measurement. A low-sensitivity meter may give correct readings when measuring voltages in low-resistance circuits but is certain to produce very unreliable readings in high-resistance circuits. A voltmeter, when connected across two points in a highly resistive circuit, acts as a shunt for that portion of the circuit and thus reduces the equivalent resistance in that portion of the circuit. The meter will then give a lower indication of the voltage drop than actually existed before the meter was connected. This effect is called the *loading effect* of an instrument and is caused principally by *low-sensitivity* instruments. The loading effect of a voltmeter is illustrated in Example 4-6.

Example 4-6: It is desired to measure the voltage across the 50-kΩ resistor in the circuit of Fig. 4-18. Two voltmeters are available for this

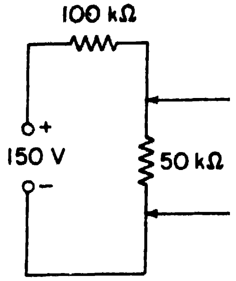


Figure 4-18
Voltmeter loading effect.

measurement: voltmeter 1 has a sensitivity of 1,000 Ω/V and voltmeter 2, 20,000 Ω/V. Both meters are used on their 50-V range. Calculate (a) the reading of each meter; (b) the error in each reading, expressed as a percentage of the true value.

SOLUTION: Inspection of the circuit indicates that the voltage across the 50-kΩ resistor is

$$\frac{50\ \text{k}\Omega}{150\ \text{k}\Omega} \times 150\ \text{V} = 50\ \text{V}$$

This is the *true* value of voltage across the 50-kΩ resistor.

(a) *Voltmeter 1* ($S = 1{,}000\ \Omega/\text{V}$) has a resistance of 50 V × 1,000 Ω/V = 50 kΩ on its 50-V range. Connecting the meter across the 50-kΩ resistor causes the equivalent parallel resistance to be decreased to 25 kΩ and the total circuit resistance to 125 kΩ. The potential difference across the combination of meter and 50-kΩ resistor is

$$V_1 = \frac{25\ \text{k}\Omega}{125\ \text{k}\Omega} \times 150\ \text{V} = 30\ \text{V}$$

Hence, the voltmeter indicates a voltage of 30 V. *Voltmeter 2* ($S = 20\ \text{k}\Omega/\text{V}$) has a resistance of 50 V × 20 kΩ/V = 1 mΩ on its 50-V range. When this meter is connected across the 50-kΩ resistor, the equivalent parallel resistance equals 47.6 kΩ. This combination produces a voltage of

$$V_2 = \frac{47.6\ \text{k}\Omega}{147.6\ \text{k}\Omega} \times 150\ \text{V} = 48.36\ \text{V}$$

which is indicated on the voltmeter.

(b) *Voltmeter 1* indicates an error of

$$\%\ \text{error} = \frac{\text{true voltage} - \text{apparent voltage}}{\text{true voltage}} \times 100\%$$

$$= \frac{50\ \text{V} - 30\ \text{V}}{50\ \text{V}} \times 100\% = 40\%$$

Voltmeter 2 indicates an error of

$$\%\ \text{error} = \frac{50\ \text{V} - 48.36\ \text{V}}{50\ \text{V}} \times 100\% = 3.28\%$$

The computation of Example 4-6 indicates that the meter with the higher sensitivity or ohms-per-volt rating gives the most *reliable* result. It is important to realize the factor of sensitivity, particularly in cases where voltage measurements are made in high-resistance circuits.

The matter of reliability and accuracy of the test result raises an interesting point. When an *insensitive, yet highly accurate,* dc voltmeter is placed across the terminals of a high resistance, the meter accurately reflects the voltage condition produced by loading. The error is a *human error* or gross error (Sec. 1-4), because the proper instrument was not selected. The meter "disturbs" the circuit, and the ideal of instrumentation, at all times, is to measure a condition without affecting it in any way. But the human investigator has the responsibility to select an instrument which is precise, reliable, and sufficiently sensitive not to disturb that which is being measured. The fault lies not with the highly accurate instrument but with the investigator, who is using it incorrectly. In fact, the sophisticated instrument user could calculate the true voltage using an insensitive yet accurate meter.

Therefore, *accuracy* is always required in instruments; *sensitivity* is needed only in special applications where loading disturbs that which is being measured. Example 4-7 illustrates how an insensitive, yet accurate instrument is used to perform an entirely valid measurement.

Example 4-7: The only voltmeter available in a laboratory has a sensitivity of 100 Ω/V and three scales, 50 V, 150 V, and 300 V. When connected in the circuit of Fig. 4-19, the meter reads 4.65 V on its lowest (50-V) scale. Calculate the value of R_x.

SOLUTION: The equivalent resistance of the voltmeter on its 50-V scale is

$$R_V = 100\,\frac{\Omega}{V} \times 50\text{ V} = 5\text{ k}\Omega$$

Let R_p = the parallel resistance of R_x and R_V.

$$R_p = \frac{V_p}{V_s} \times R_s = \frac{4.65}{95.35} \times 100\text{ k}\Omega = 4.878\text{ k}\Omega$$

Then,

$$R_x = \frac{R_p \times R_V}{R_V - R_p} = \frac{4.878\text{ k}\Omega \times 5\text{ k}\Omega}{0.122\text{ k}\Omega} = 200\text{ k}\Omega$$

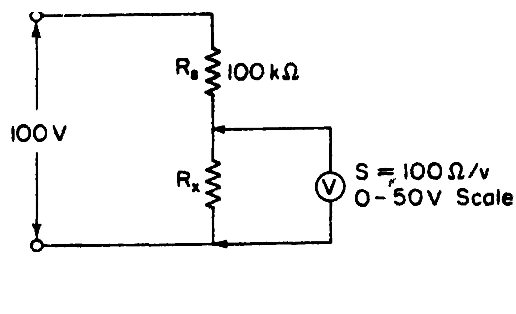


Figure 4-19
The use of an accurate but insensitive voltmeter to determine the resistance of R_x.

Note that Example 4-7 shows that when the instrument user is aware of the limitations of his instrument, he can still make allowances provided that the voltmeter is accurate.

The following general precautions should be observed when using a voltmeter:

(a) Observe the correct polarity. Wrong polarity causes the meter to deflect against the mechanical stop and this may damage the pointer.

(b) Place the voltmeter *across* the circuit or component whose voltage is to be measured.

(c) When using a multirange voltmeter, always use the highest voltage range and then decrease the range until a good up-scale reading is obtained.

(d) Always be aware of the loading effect. The effect can be minimized by using as high a voltage range (and highest sensitivity) as possible. The precision of measurement decreases if the indication is at the low end of the scale (Sec. 1-4).

4-8 The Voltmeter-Ammeter Method

A popular type of resistance measurement involves the ***voltmeter-ammeter method,*** since the instruments required are usually available in most laboratories. If the voltage V across the resistor and the current I through the resistor are measured, then the unknown resistance R_x is given by Ohm's law:

$$R_x = \frac{V}{I} \tag{4-24}$$

Equation (4-24) implies that the ammeter resistance is zero and the voltmeter resistance infinite, so that the conditions in the circuit are not disturbed.

In Fig. 4-20(a), the *true* current supplied to the load is measured by the

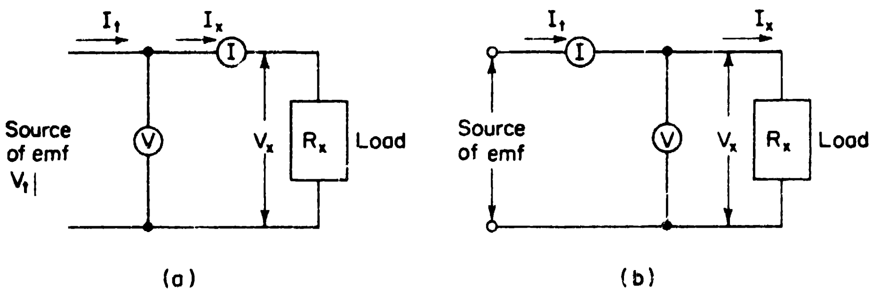


Figure 4-20
The effect of placing the meters in voltmeter-ammeter measurements.

ammeter, but the voltmeter measures the supply voltage rather than the actual load voltage. To find the true voltage across the load, subtract the voltage drop across the ammeter from the voltmeter reading. If the voltmeter is placed directly across the resistor, as in Fig. 4-20(b), it measures the true load voltage, but now the ammeter is in error by the amount of current drawn by the voltmeter. In either situation, an *error* is introduced in the measurement of R_x by this method. The correct method of connecting the voltmeter depends on the value of R_x and the voltmeter and ammeter resistances. In general, the ammeter resistance is low and the voltmeter resistance is high.

In Fig. 4-20(a), the ammeter reads the true value of the load current (I_x), and the voltmeter is measuring the supply voltage (V_t). If R_x is large compared to the internal resistance of the ammeter, the error introduced by neglecting the voltage drop across the ammeter is negligible and V_t is very close to the true load voltage V_x. The connection of Fig. 4-20(a) is therefore the best circuit when measuring *high resistance values.*

In Fig. 4-20(b), the voltmeter reads the true value of the load voltage (V_x) and the ammeter reads the supply current (I_t). If R_x is small compared to the internal resistance of the voltmeter, the current drawn by the voltmeter does not appreciably affect the total supply current and I_t is very close to the true value of the load current I_x. The connection of Fig. 4-20(b), therefore, is the best circuit when measuring *low resistance values.*

Given an unknown value of R_x, how then can we determine if the voltmeter is connected in the right place? Consider the circuit of Fig. 4-21, where the voltmeter and the ammeter may be connected in two different ways for simultaneous readings. The procedure is as follows:

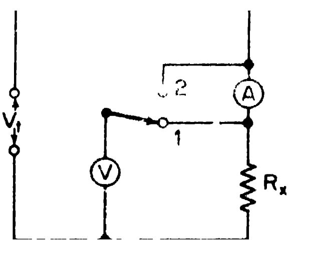


Figure 4-21
The effect of voltmeter position in a voltmeter-ammeter measurement.

(a) Connect the voltmeter across R_x with the switch in position 1 and observe the ammeter reading.

(b) Now switch the voltmeter to position 2. If the ammeter reading does not change, restore the voltmeter to position 1. The symptoms indicate a low resistance measurement. Record both current and voltage readings and calculate R_x from Eq. (4-24).

(c) If the ammeter reading decreases when the voltmeter is changed from position 1 to position 2, leave the voltmeter at position 2. The symptoms indicate a high resistance measurement. Record both current and voltage readings and calculate R_x from Eq. (4-24).

Voltage measurements in electronic circuits are generally made with multirange voltmeters or multimeters, with sensitivities in the range of 20 kΩ/V to 50 kΩ/V. In power measurements, where the current is usually large, voltmeters may have sensitivities as low as 100 Ω/V. Ammeter resistances depend on the design of the coil and are generally larger for low values of full-scale currents. A few typical values of ammeter resistance are given in Table 4-1.

TABLE 4-1

TYPICAL VALUES OF AMMETER RESISTANCE*

Full-scale Current (μA)	Resistance (ohms) Pivot and jewel	Taut-band
50	2,000–5,000	1,000–2,000
500	200–1,000	100–250
1,000	50–120	30–90
10,000	2–4	1–3

NOTE: Current ranges in excess of 30 mA are usually shunted.

**Data Sheets*, Weston Instruments, Inc., Newark, N.J.

4-9 The Series-Type Ohmmeter

The series-type ohmmeter essentially consists of a basic d'Arsonval movement connected in series with a resistance and a battery to a pair of terminals to which the unknown is connected. The current through the movement then depends on the magnitude of the unknown resistor, and the meter indication is proportional to the value of the unknown, provided that calibration problems are taken into account. Figure 4-22 shows the elements of a simple single-range series ohmmeter.

In Fig. 4-22,

R_1 = current limiting resistor
R_2 = zero adjust resistor
E = internal battery
R_m = internal resistance of the d'Arsonval movement
R_x = the unknown resistor

When the unknown resistor $R_x = 0$ (terminals A and B shorted) maximum current flows in the circuit. Under this condition, shunt resistor R_2 is adjusted until the movement indicates full-scale current (I_{fsd}). The full-scale current position of the pointer is marked "0 Ω" on the scale. Similarly, when $R_x = \infty$ (terminals A and B open), the current in the circuit drops to zero and the movement indicates zero current, which is then marked "∞." Intermediate-scale markings may be placed on the scale by connecting different known values

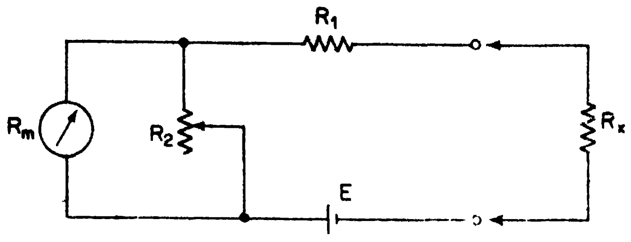


Figure 4-22
The series-type ohmmeter.

of R_x to the instrument. The accuracy of these scale markings depends on the repeating accuracy of the movement and the tolerances of the calibrating resistors.

Although the series-type ohmmeter is a popular design and is used extensively in portable instruments for general-service work, it has certain difficulties. Important among these is the internal battery whose voltage decreases gradually with time and age, so that the full-scale current drops and the meter does not read "0" when A and B are shorted. The variable shunt resistor R_2 in Fig. 4-22 provides an adjustment to counteract the effect of battery change. Without R_2, it would be possible to bring the pointer back to full scale by adjusting R_1, but this would change the calibration all along the scale. Adjustment by R_2 is a superior solution, since the parallel resistance of R_2 and the coil R_m is always low compared to R_1 and therefore the change in R_2 needed for adjustment does not change the calibration very much. The circuit of Fig. 4-22 does not compensate completely for aging of the battery, but it does a reasonably good job within the expected limits of accuracy of the instrument.

A convenient quantity to use in the design of a series-type ohmmeter is the value of R_x which causes half-scale deflection of the meter. At this position, the resistance across terminals A and B is defined as the half-scale position resistance R_h. Given the full-scale current I_{fsd} and the internal resistance of the movement R_m, the battery voltage E, and the desired value of the half-scale resistance R_h, the circuit can be analyzed; i.e., values can be found for R_1 and R_2.

The design can be approached by recognizing that, if introducing R_h reduces the meter current to $\frac{1}{2}I_{fsd}$, the unknown resistance must be equal to the total internal resistance of the ohmmeter. Therefore,

$$R_h = R_1 + \frac{R_2 R_m}{R_2 + R_m} \tag{4-25}$$

The total resistance, presented to the battery, then equals $2R_h$, and the battery current needed to supply the half-scale deflection, is

$$I_h = \frac{E}{2R_h} \tag{4-26}$$

To produce full-scale deflection, the battery current must be doubled, and therefore,

$$I_t = 2I_h = \frac{E}{R_h} \tag{4-27}$$

The shunt current through R_2 is

$$I_2 = I_t - I_{fsd} \tag{4-28}$$

The voltage across the shunt (E_{sh}) is equal to the voltage across the movement and

$$E_{sh} = E_m \quad \text{or} \quad I_2R_2 = I_{fsd}R_m$$

and

$$R_2 = \frac{I_{fsd}R_m}{I_2} \tag{4-29}$$

Substituting Eq. (4-28) into Eq. (4-29) we obtain:

$$R_2 = \frac{I_{fsd}R_m}{I_t - I_{fsd}} = \frac{I_{fsd}R_mR_h}{E - I_{fsd}R_h} \tag{4-30}$$

Solving Eq. (4-25) for R_1:

$$R_1 = R_h - \frac{R_2R_m}{R_2 + R_m} \tag{4-31}$$

Substituting Eq. (4-30) into Eq. (4-31) and solving for R_1 yields

$$R_1 = R_h - \frac{I_{fsd}R_mR_h}{E} \tag{4-32}$$

A sample calculation for the series-type ohmmeter is given in Example 4-8.

Example 4-8: The ohmmeter of Fig. 4-22 uses a 50-Ω basic movement requiring a full-scale current of 1 mA. The internal battery voltage is 3 V. The desired scale marking for half-scale deflection is 2,000 Ω. Calculate (a) the values of R_1 and R_2; (b) the maximum value of R_2 to compensate for a 10% drop in battery voltage; (c) the scale error at the half-scale mark (2,000 Ω) when R_2 is set as in (b).

SOLUTION:
(a) The total battery current at full-scale deflection is

$$I_t = \frac{E}{R_h} = \frac{3\text{ V}}{2{,}000} = 1.5\text{ mA} \tag{4-27}$$

The current through the zero-adjust resistor R_2 then is

$$I_2 = I_t - I_{fsd} = 1.5\text{ mA} - 1\text{ mA} = 0.5\text{ mA} \tag{4-28}$$

The value of the zero-adjust resistor R_2 is

$$R_2 = \frac{I_{fsd}R_m}{I_2} = \frac{1\text{ mA} \times 50\ \Omega}{0.5\text{ mA}} = 100\ \Omega \tag{4-29}$$

The parallel resistance of the movement and the shunt (R_p) is

$$R_p = \frac{R_2 R_m}{R_2 + R_m} = \frac{50 \times 100}{150} = 33.3\ \Omega$$

The value of the current-limiting resistor R_1 is

$$R_1 = R_h - R_p = 2{,}000 - 33.3 = 1{,}966.7\ \Omega$$

(b) At a 10% drop in battery voltage

$$E = 3\ \text{V} - 0.3\ \text{V} = 2.7\ \text{V}$$

The total battery current I_t then becomes

$$I_t = \frac{E}{R_h} = \frac{2.7\ \text{V}}{2{,}000\ \Omega} = 1.35\ \text{mA}$$

The shunt current I_2 is

$$I_2 = I_t - I_{fsd} = 1.35\ \text{mA} - 1\ \text{mA} = 0.35\ \text{mA}$$

and the zero-adjust resistor R_2 equals

$$R_2 = \frac{I_{fsd} R_m}{I_2} = \frac{1\ \text{mA} \times 50\ \Omega}{0.35\ \text{mA}} = 143\ \Omega$$

(c) The parallel resistance of the meter movement and the new value of R_2 becomes

$$R_p = \frac{R_2 R_m}{R_2 + R_m} = \frac{50 \times 143}{193} = 37\ \Omega$$

Since the half-scale resistance R_h is equal to the total internal circuit resistance, R_h will increase to

$$R_h = R_1 + R_p = 1{,}966.7\ \Omega + 37\ \Omega = 2{,}003.7\ \Omega$$

Therefore, the true value of the half-scale mark on the meter is 2,003.7 Ω, whereas the actual scale mark is 2,000 Ω. The percentage error is then

$$\%\ \text{error} = \frac{2{,}000 - 2{,}003.7}{2{,}003.7} \times 100\% = -0.185\%$$

The negative sign indicates that the meter reading is low.

The ohmmeter of Ex. 4-8 could be designed for other values of R_h, within limits. If $R_h = 3{,}000\ \Omega$, the battery current would be 1 mA, which is required for the full-scale deflection current. If the battery voltage would decrease owing to aging, the total battery current would fall below 1 mA and there would then be no provision for adjustment.

4-10 The Shunt-Type Ohmmeter

The circuit diagram of a ***shunt-type ohmmeter*** is shown in Fig. 4-23. It consists of a battery in series with an adjustable resistor R_1 and a basic d'Arsonval movement. The unknown resistance is connected across terminals A and

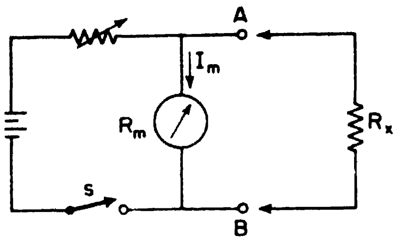


Figure 4-23
The shunt-type ohmmeter.

B, parallel with the meter. In this circuit it is necessary to have an *off-on switch* to disconnect the battery from the circuit when the instrument is not used. When the unknown resistor $R_x = 0\ \Omega$ (*A* and *B* shorted), the meter current is zero. If the unknown resistor $R_x = \infty$ (*A* and *B* open), the current finds a path only through the meter and by appropriate selection of the value of R_1, the pointer can be made to read full scale. This ohmmeter therefore has the "zero" mark at the left-hand side of the scale (no current) and the "infinite" mark at the right-hand side of the scale (full-scale deflection current).

The shunt-type ohmmeter is particularly suited to the measurement of *low-value resistors.* It is not a commonly used test instrument, but it is found in laboratories or for special low-resistance applications.

The analysis of the shunt-type ohmmeter is similar to that of the series-type ohmmeter (Sec. 4-8). In Fig. 4-23, when $R_x = \infty$, the full-scale meter current will be

$$I_{fsd} = \frac{E}{R_1 + R_m} \tag{4-33}$$

where E = the internal battery voltage
R_1 = current limiting resistor
R_m = internal resistance of the movement.

Solving for R_1, we find:

$$R_1 = \frac{E}{I_{fsd}} - R_m \tag{4-34}$$

For any value of R_x, connected across the meter terminals, the meter current decreases and is given by

$$I_m = \left\{\frac{E}{R_1 + [R_m R_x/(R_m + R_x)]}\right\} \times \frac{R_x}{R_m + R_x}$$

or

$$I_m = \frac{ER_x}{R_1 R_m + R_x(R_1 + R_m)} \tag{4-35}$$

The meter current for any value of R_x, expressed as a fraction of the full-scale current, is

$$s = \frac{I_m}{I_{fsd}} = \frac{R_x(R_1 + R_m)}{R_1(R_m + R_x) + R_m R_x}$$

or

$$s = \frac{R_x(R_1 + R_m)}{R_x(R_1 + R_m) + R_1 R_m} \tag{4-36}$$

Defining

$$\frac{R_1 R_m}{R_1 + R_m} = R_p \tag{4-37}$$

and substituting Eq. (4-37) into Eq. (4-36), we obtain

$$s = \frac{R_x}{R_x + R_p} \tag{4-38}$$

Using Eq. (4-38), the meter can be calibrated by calculating s in terms of R_x and R_p.

At half-scale reading of the meter ($I_m = 0.5\, I_{fsd}$), Eq. (4-35) reduces to

$$0.5\, I_{fsd} = \frac{ER_h}{R_1 R_m + R_h(R_1 + R_m)} \tag{4-39}$$

where R_h = external resistance causing half-scale deflection. To determine the relative scale values for a given value of R_1, the half-scale reading may be found by dividing Eq. (4-33) by Eq. (4-39) and solving for R_h,

$$R_h = \frac{R_1 R_m}{R_1 + R_m} \tag{4-40}$$

The analysis shows that the half-scale resistance is determined by the limiting resistor R_1 and the internal resistance of the movement R_m. The limiting resistance, R_1, is in turn determined by the meter resistance, R_m, and the full-scale deflection current, I_{fsd}.

To illustrate that the shunt-type ohmmeter is particularly suited to the measurement of very low resistances, consider the computation given in Example 4-9:

Example 4-9: The circuit of Fig. 4-23 uses a 10-mA basic d'Arsonval movement with an internal resistance of 5 Ω. The battery voltage E = 3 V. It is desired to modify the circuit by adding an appropriate resistor R_{sh} across the movement, so that the instrument will indicate 0.5 Ω at the midpoint on its scale. Calculate (a) the value of the shunt resistor, R_{sh}; (b) the value of the current-limiting resistor, R_1.

SOLUTION:

(a) For half-scale deflection of the movement,

$$I_m = 0.5\, I_{fsd} = 5 \text{ mA}$$

The voltage across the movement is

$$E_m = 5 \text{ mA} \times 5\ \Omega = 25 \text{ mV}$$

Since this voltage also appears across the unknown resistor, R_x, the current through R_x is

$$I_x = \frac{25 \text{ mV}}{0.5 \, \Omega} = 50 \text{ mA}$$

The current through the movement (I_m) plus the current through the shunt (I_{sh}) must be equal to the current through the unknown (I_x). Therefore,

$$I_{sh} = I_x - I_m = 50 \text{ mA} - 5 \text{ mA} = 45 \text{ mA}$$

The shunt resistor then is

$$R_{sh} = \frac{E_m}{I_{sh}} = \frac{25 \text{ mV}}{45 \text{ mA}} = \frac{5}{9} \, \Omega$$

(b) The total battery current is

$$I_t = I_m + I_{sh} + I_x = 5 \text{ mA} + 45 \text{ mA} + 50 \text{ mA} = 100 \text{ mA}$$

The voltage drop across the limiting resistor R_1 equals 3 V − 25 mV = 2.975 V. Therefore,

$$R_1 = \frac{2.975 \text{ V}}{100 \text{ mA}} = 29.75 \, \Omega$$

4-11 The Multimeter, or VOM

The ammeter, the voltmeter, and the ohmmeter all use a basic d'Arsonval movement. The difference between these instruments is the circuit in which the basic movement is used. It is therefore obvious that an instrument can be designed to perform the three measurement functions. This instrument, which contains a function switch to connect the appropriate circuits to the d'Arsonval movement, is often called a *multimeter* or *volt-ohm-milliammeter* (VOM).

A representative example of a commercial multimeter is shown in Fig. 4-24, and the circuit diagram of this meter is given in Fig. 4-25. The meter is a combination of a dc milliammeter, a dc voltmeter, an ac voltmeter, a multirange ohmmeter, and an output meter. (The circuits of the ac voltmeter and the output meter are discussed in Sec. 5-4.)

Figure 4-26 shows the circuit for the *dc voltmeter* section, where the common input terminals are used for voltage ranges of 0–1.5 V to 0–1,000 V. An external voltage jack, marked "DC 5,000 V," is used for dc voltage measurements to 5,000 V. The operation of this circuit is similar to the circuit of Fig. 4-15, which was discussed in Sec. 4-6.

The basic movement of the multimeter of Fig. 4-24 has a full-scale current of 50 μA and an internal resistance of 2,000 Ω. The values of the multipliers are given in Fig. 4-26. Notice that on the 5,000-V range, the range switch should

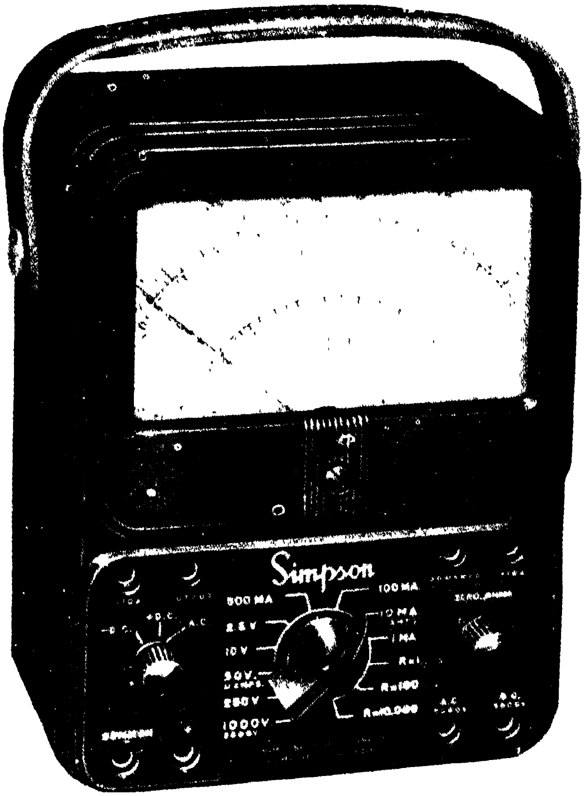


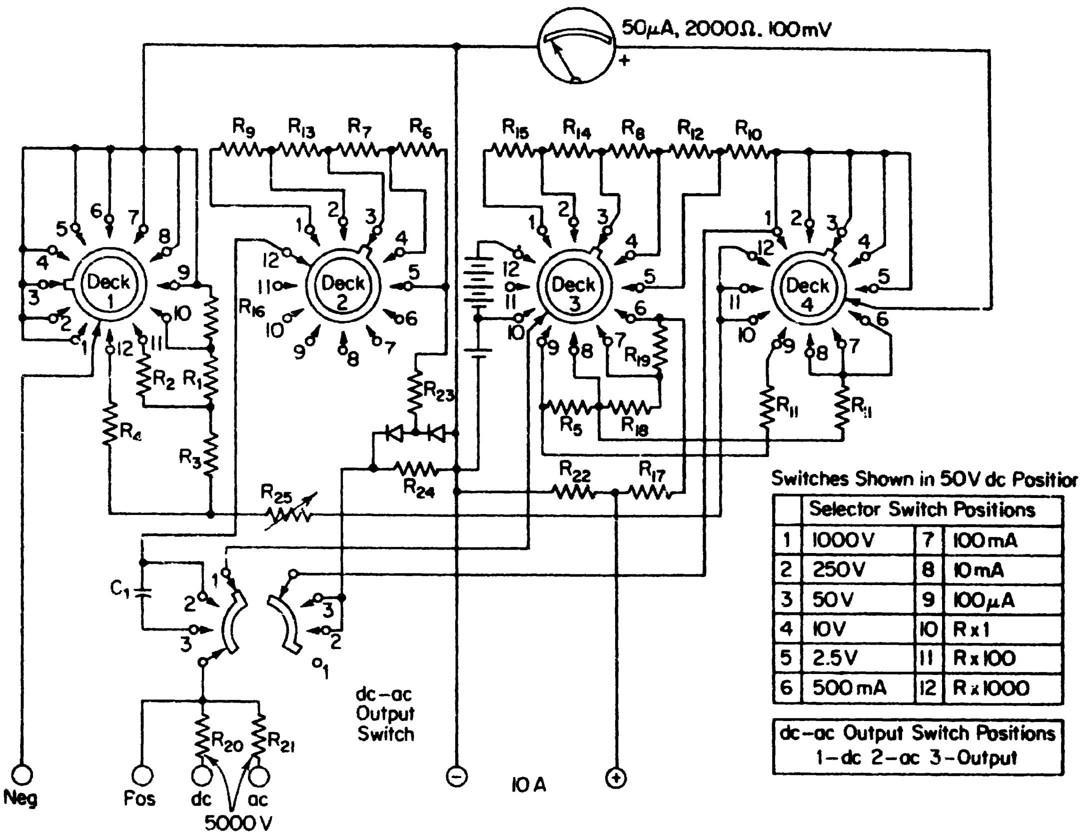
Figure 4-24
A general purpose multimeter, the Simpson Model 260 (*Courtesy Simpson Electric Company*).

be set to the 1,000-V position, but the test lead should be connected to the external jack marked "DC 5,000 V." The normal precautions for measuring voltage should be taken. Because of its fairly high sensitivity (20 kΩ/V), the instrument is suited to general-service work in the electronics field.

The circuit for measuring dc milliamperes and amperes is given in Fig. 4-27 and again the circuit is self-explanatory. The "common" (+) and "negative" (−) terminals are used for current measurements up to 500 mA and the jacks marked "+10 A" and "−10 A" are used for the 0–10-A range.

Details of the *ohmmeter section* of the VOM are shown in Fig. 4-28 on page 88. The circuit in Fig. 4-28(a) gives the ohmmeter circuit for a scale multiplication of 1. Before making any measurement, the instrument is short-circuited and the "zero adjust" control varied until the meter reads zero resistance (full-scale current). Notice that the circuit takes the form of a variation of the shunt-type ohmmeter. Scale multiplications of 100 and 10,000 are shown in Fig. 4-28(b), (c).

AC voltmeter readings are obtained by setting the "ac-dc" switch to the "ac" position. (The operation of this circuit is discussed in Sec. 5-4.)



	Selector Switch Positions		
1	1000V	7	100mA
2	250V	8	10mA
3	50V	9	100μA
4	10V	10	R x 1
5	2.5V	11	R x 100
6	500mA	12	R x 1000

dc–ac Output Switch Positions
1–dc 2–ac 3–Output

Figure 4-25 Schematic diagram of the Simpson Model 260 VOM (*Courtesy Simpson Electric Company*).

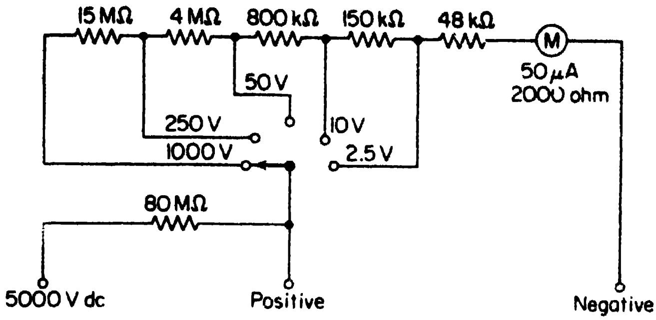


Figure 4-28
DC voltmeter section of the Simpson Model 260 multimeter (*Courtesy Simpson Electric Company*).

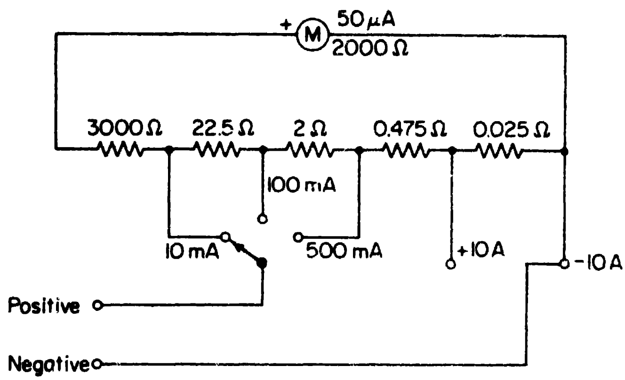
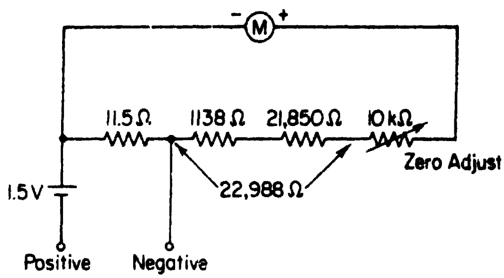


Figure 4-27
DC ammeter section of the Simpson Model 260 multimeter (*Courtesy Simpson Electric Company*).

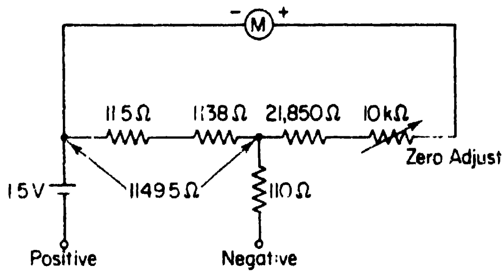
4-12 Calibration of dc Instruments

Although detailed calibration techniques are beyond the scope of this chapter, some general procedures for calibration of basic dc instruments are given.

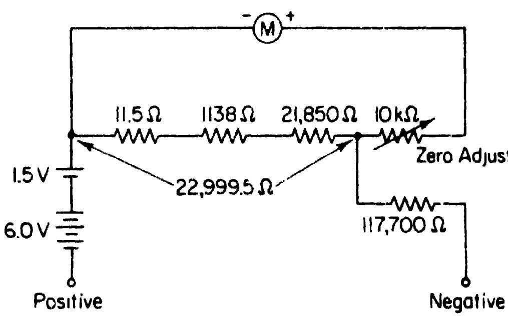
Calibration of a *dc ammeter* can most easily be carried out by the arrangement of Fig. 4-29. The value of the current through the ammeter under calibration is determined by measuring the potential difference across a standard resistor by the *potentiometer method* and then calculating the current by Ohm's law. The result of the computation is compared to the actual reading of the ammeter to be calibrated, which is also inserted in the circuit. (Voltage measurements by the potentiometer method are discussed in some detail in Sec. 6-6.)



(a) Ohmmeter Circuit R x 1 Range



(b) Ohmmeter Circuit R x 100 Range



(c) Ohmmeter Circuit R x 10,000 Range

Figure 4-28
The ohmmeter section of the Simpson Model 260 multimeter (*Courtesy Simpson Electric Company*).

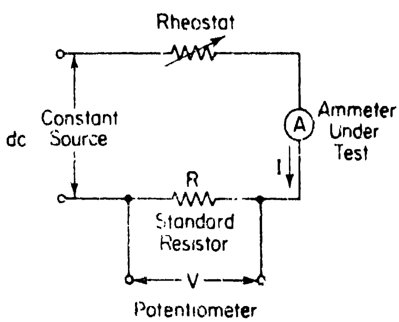


Figure 4-29
Potentiometer method of calibrating a dc ammeter.

A good source of constant current is required and is usually provided by storage cells or a precision power supply, A rheostat is placed in the circuit to control the current to any desired value, so that different points on the meter scale can be calibrated.

A simple method of calibrating a *dc voltmeter* is shown in Fig. 4-30, where the voltage across dropping resistor, R, is accurately measured with a potentiometer. The meter to be calibrated is connected across the same two points as the potentiometer and should therefore indicate the same voltage. A rheostat is placed in the circuit to control the amount of current and therefore the drop across the resistor, R, so that several points on the voltmeter scale can be calibrated. Voltmeters tested with the method of Fig. 4-30 can be calibrated with an accuracy of ± 0.01 per cent, which is well beyond the usual accuracy of a basic d'Arsonval movement.

The *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 reading of the ohmmeter. Doing this for several points on the ohmmeter scale and on several ranges, allows one to obtain an indication of the correct operation of the instrument. Precision measurements of resistance are normally carried out by a bridge method, which is discussed in some detail in Chapter 7.

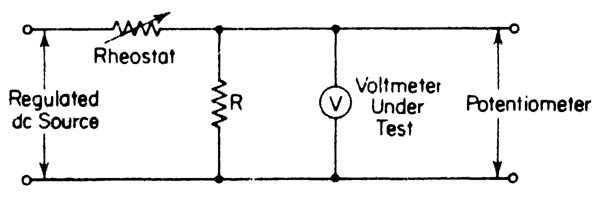


Figure 4-30 Potentiometer method of calibrating a dc voltmeter.

Problems

1. Determine the full-scale voltage indicated by a 500-μA meter movement with an internal resistance of 250 Ω, if no multiplier is used.

2. Design a multirange dc ammeter with ranges of 0–10 mA, 0–50 mA, 0–100 mA, and 0–500 mA. A basic d'Arsonval movement is used with an internal resistance $R_m = 50\ \Omega$ and a full-scale deflection current $I_{fsd} =$ 1 mA. (a) Calculate the values of the required shunts. (b) Draw the complete circuit diagram.

3. A simple shunted dc ammeter, using a basic meter movement with an internal resistance, $R_m = 1{,}800\ \Omega$, and a full-scale deflection current, $I_{fsd} = 100\ \mu$A, is connected into a circuit and gives a reading of 3.5 mA on its 5-mA range. This reading is checked by a recently calibrated dc ammeter which gives a reading of 4.1 mA. The implication is that the first ammeter has a faulty shunt on its 5-mA range. Calculate (a) the actual value of the faulty shunt; (b) the correct shunt for the 5-mA range.

4. Design an Ayrton shunt for a meter movement with internal resistance $R_m = 2{,}500\ \Omega$ and full-scale deflection current $I_{fsd} = 50\ \mu$A to provide current ranges of 50 μA, 100 μA, 500 μA, 10 mA, and 100 mA. (a) Calculate the resistances of the Ayrton shunt components. (b) Draw the schematic diagram, including the switching arrangement, for this multirange dc ammeter.

5. It is desired to convert a 50-μA dc movement with internal resistance of 1,000 Ω to a 0–2,500-V dc voltmeter. Calculate (a) the resistance of the multiplier; (b) the sensitivity of the instrument.

6. An existing 0–200-V dc voltmeter has a sensitivity of 1,000 Ω/V. Determine the value of additional series resistance required to convert this voltmeter to an instrument with a range of 0–1,000 V dc.

7. Design a multirange voltmeter with ranges of 0–5 V, 0–10 V, 0–50 V, and 0–100 V, using a 50-μA movement with internal resistance of 1,500 Ω. Calculate (a) the values of the multipliers; (b) the sensitivity of the instrument. Draw the circuit diagram of the completed design.

8. A dc microammeter with an internal resistance of 250 Ω and a full-scale deflection current of 500 μA indicates a current of 300 μA when connected into a circuit consisting of a 1.5-V dry cell and an unknown resistance. Determine the value of the unknown resistance.

9. Design a series-type ohmmeter, similar to the circuit of Fig. 4-22. The movement to be used requires 0.5 mA for full-scale deflection and has an internal resistance of 50 Ω. The internal battery has a voltage of 3.0 V.

The desired value of half-scale resistance is 3,000 Ω. Calculate (a) the value of resistors R_1 and R_2; (b) the range of values of R_2, if the battery voltage may vary from 2.7 V to 3.1 V. Use the value of R_1 as calculated in (a).

10. A series-type ohmmeter, designed to operate with a 6-V battery, has a circuit diagram as shown in Fig. 4-22. The meter movement has an internal resistance of 2,000 Ω and requires a current of 100 μA for full-scale deflection. The value of resistor R_1 is 49 kΩ. (a) Assuming that the battery voltage has fallen to 5.9 V, calculate the value of R_2 required to zero the ohmmeter. (b) Under the conditions mentioned in (a), an unknown resistor R_x is connected to the meter, causing a 60 per cent meter deflection. Calculate the value of the unknown resistor R_x.

11. The movement of the multirange voltmeter in Fig. 4-17 has a full-scale current of 50 μA and an internal resistance of 2,000 Ω. The full-scale meter reading is 150 V with the range switch set in position V_1, 50 V with the switch in position V_2, 10 V with the switch in position V_3, and 1 V with the switch in position V_4. Calculate (a) the resistance of the multipliers R_1, R_2, R_3, and R_4; (b) the sensitivity of the voltmeter.

12. A dc voltmeter is rated with a sensitivity of 10 kΩ/V and is used on its 0–150-V range to measure the voltage across the 100-kΩ resistor in Fig. 4-18. Determine the percentage error of the meter indication.

13. Design a volt-ohm-milliammeter with the following characteristics:

(a) Voltage ranges: 0–5, 0–25, 0–100, and, 0–500 V dc.
(b) Current ranges: 0–10, 0–100, 0–500, and 0–1,000 mA dc.
(c) Resistance ranges: 20 Ω, 2,000 Ω, and 200 kΩ at half scale.

The movement used in this instrument is a basic d'Arsonval mechanism with internal resistance of 1,500 Ω and full-scale current of 50 μA. (Refer to the circuit diagrams and description of the multimeter of Fig. 4-24 for information about circuit arrangements.)

14. The dc voltmeter of Fig. 4-20(b) has a sensitivity of 1,000 Ω/V and a full-scale reading of 100 V. The meter indicates 84 V as the voltage across the load. Calculate the error in measuring the power dissipation of the load by the voltmeter-ammeter method when the ammeter indicates a current of (a) 50 mA; (b) 1 A; (c) 10 A.

FIVE

ALTERNATING-CURRENT INDICATING INSTRUMENTS

5-1 Introduction

The d'Arsonval movement (Sec. 4-3) responds to the *average* or *dc value* of the current through the moving coil. If this movement carries an alternating current (ac) with positive and negative half cycles, the driving torque would be in one direction for the positive alternation and in the other for the negative alternation. 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, for example at the powerline frequency (60 Hz) or higher, the inertia of the coil is so great, that the pointer cannot follow the rapid reversals of the driving torque and hovers around the zero mark, vibrating slightly.

To measure ac on an indicating instrument, such as the d'Arsonval movement, some means must be devised to obtain a *unidirectional* torque which does not reverse each half cycle. One solution to this problem is found by reversing the direction of the field flux at exactly the same instant that the current through the movable coil reverses. The torque then stays in the same direction for both halves of the ac cycle and the meter reads up-scale for the complete cycle.

Another approach involves rectification of the ac; then the dc d'Arsonval movement may be used with the rectified current deflecting the movement. Still other methods use the heating effect of ac to produce an analog indication of its magnitude. These and other methods are discussed in this chapter.

5-2 The Electrodynamometer Movement

One of the most important ac movements is the *electrodynamometer*. It is used in constructing accurate ac voltmeters and ammeters or *standards*, not only at the powerline frequencies but also in the lower audiofrequency range. With some slight modifications, the electrodynamometer can be used as a wattmeter, a VARmeter, a power-factor meter, or a frequency meter. The electrodynamometer movement may also serve as a *transfer* instrument, since it can be calibrated on dc and then used directly on ac, establishing a direct means of equating ac and dc measurements of voltage and current. (See definitions of effective value of current and voltage, Sec. 5-2.)

Whereas the d'Arsonval movement employs a permanent magnet to provide the magnetic field in which the movable coil rotates, the electrodynamometer uses the current under measurement to produce the necessary field flux. Figure 5-1 shows the schematic arrangement of the parts of this movement.

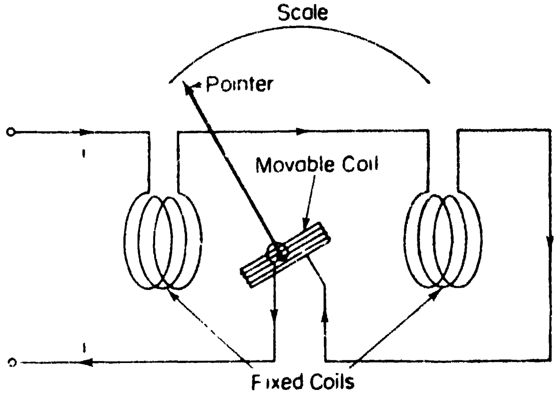


Figure 5-1
Schematic diagram of an electrodynamometer movement used as a current meter.

A fixed coil, split into two equal halves, provides the magnetic field in which the movable coil rotates. The two coil halves are connected in series with the moving coil and are fed by the current under measurement. The fixed coils are spaced far enough apart to allow passage of 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 the d'Arsonval movement construction. The complete assembly is surrounded by a laminated shield to protect the instrument from stray magnetic fields which may affect its operation. (Figure 5-2 shows a cutaway view of the electrodynamometer.)

Damping is provided by aluminum air vanes, moving in sector-shaped chambers. It can be seen that the entire movement is very solid and rigidly constructed in order to keep its mechanical dimensions stable and its calibration intact.

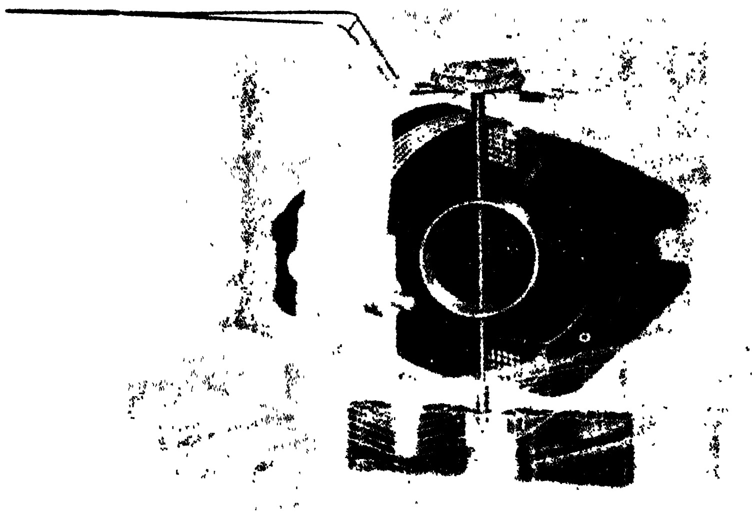


Figure 5-2
Phantom photograph of an electrodynamometer movement, showing the arrangement of fixed and movable coils. The rigidly constructed mechanism is surrounded by a laminated magnetic shield to minimize the effect of external fields on the meter indication (*Courtesy Weston Instruments Inc.*).

The operation of the instrument may be understood by returning to the expression for the torque developed by a coil suspended in a magnetic field. We previously stated, Eq. (4-1), that

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

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

If the electrodynamometer is exclusively designed for dc use, its square-law scale is easily noticed, with equal increments crowded at the very low current values and progressively spreading out at the higher current values.

For ac use, the developed torque at any instant is proportional to the *instantaneous* current *squared* (i^2). The instantaneous value of i^2 is always positive and torque pulsations are therefore produced. The movement, however, cannot follow the rapid variations of the torque and takes 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 electrodynamometer is usually calibrated in terms of the square root of the average current squared and the meter therefore reads the *rms or effective value* of the ac.

The transfer properties of the electrodynamometer become apparent when we compare the effective value of alternating current and direct current in terms of their heating effect or transfer of power. An alternating current that produces heat in a given resistance (R) at the same average rate as a direct current (I) has, by definition, a value of I amperes. The average rate of producing heat by a dc of I amperes in a resistance R, is I^2R watts. The average rate of producing heat by an ac of i amperes during one cycle in the same resistance R, is $1/T\int_0^T i^2 \mathrm{R}\, dt$. By definition, therefore,

$$I^2\mathrm{R} = \frac{1}{T}\int_0^T i^2\mathrm{R}\, dt \quad \text{and}$$

$$I = \sqrt{\frac{1}{T}\int_0^T i^2\, dt} = \sqrt{\text{average } i^2}$$

This current, I, is then called the root-mean-square (rms) or effective value of the alternating current and is often referred to as the *equivalent* dc value.

If the electrodynamometer is calibrated using 1-A dc and a mark is placed on the scale, indicating this 1-A dc value, then that ac which causes the pointer to deflect to the same mark on the scale must have an rms value of 1 A. We can therefore "transfer" a reading made with dc to its corresponding ac value and have thereby established a direct connection between ac and dc. The electrodynamometer then becomes very useful as a *calibration* laboratory instrument and is often used because of its inherent accuracy.

The electrodynamometer, however, has certain disadvantages. One of these is its high power consumption, a direct result of its construction. The current under measurement must not only pass through the movable coil, but must also provide the field flux. To get a sufficiently strong magnetic field, a high mmf is required and the source must supply a high current and power. In spite of this high power consumption, the magnetic field is very much weaker than that of a comparable d'Arsonval movement because there is no iron in the former (i.e., the entire flux path consists of air). Some instruments have been designed using special laminated steel for part of the flux path, but the presence of metal introduces calibration problems caused by frequency and waveform effects. Typical values of electrodynamometer flux density are in the range of approximately 60 gauss. This compares very unfavorably with the high flux densities (1,000–4,000 gauss) of a good d'Arsonval movement. The low flux density of the electrodynamometer immediately affects the developed torque and therefore the sensitivity of the instrument, which is typically very low.

The addition of a series resistor converts the electrodynamometer into a *voltmeter,* which again can be used to measure dc *and* ac voltages. For reasons previously mentioned, the sensitivity of the electrodynamometer voltmeter is low, approximately 10–30 Ω/V (compare this to the 20 kΩ/V of a d'Arsonval meter). The reactance and resistance of the coils also increase with increasing frequency, limiting the application of the electrodynamometer voltmeter to the lower frequency ranges. It is, however, very accurate at the powerline frequencies and is therefore often used as a *secondary* standard.

The electrodynamometer movement (even unshunted) may be regarded as an ammeter, but it becomes rather difficult to design a moving coil which can carry more than approximately 100 mA. Larger current would have to be carried to the moving coil through heavy lead-in wires, which would lose their flexibility. A shunt, when used, is usually placed across the movable coil only. The fixed coils are then made of heavy wire which can carry the large total current and it is feasible to build ammeters for currents up to 20 A. Larger values of ac currents are usually measured by using a current transformer and a standard 5-A ac ammeter (Sec. 5-12).

5-3 Moving-Iron Instruments

The various types of moving-iron instruments can be classified into *attraction-* and *repulsion-*type instruments. The latter are among the more commonly used instruments. A *radial-vane* repulsion movement is shown in diagrammatic form in Fig. 5-3.

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; the other vane is connected to the instrument shaft which can rotate freely. The current through the coil magnetizes both vanes with the same polarity, regardless of the instantaneous direction of current. The two magnetized vanes experience a repelling force and since only one vane can move, its displacement is an analog of the magnitude of coil current. The repelling force is proportional to the current squared, but the effects of frequency and hysteresis tend to produce a pointer deflection which is not linear nor has it a perfect square-law relationship.

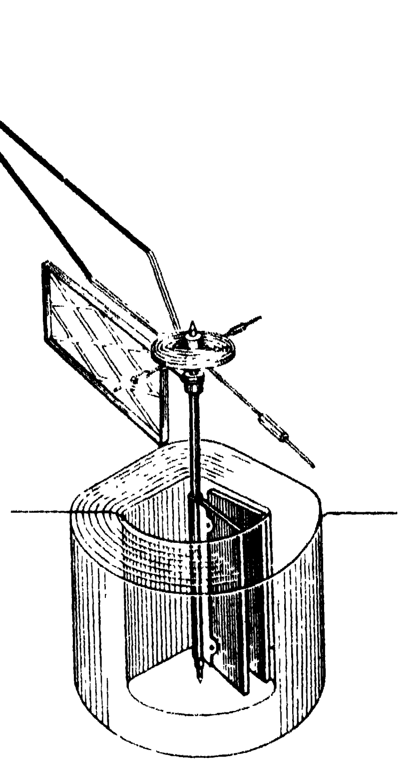


Figure 5-3
The radial-vane moving-iron mechanism. The aluminum damping vane, attached to the shaft just below the pointer, rotates in a closely fitting chamber to bring the pointer to rest quickly.

The radial-vane repulsion instrument is the most sensitive of the moving iron-vane mechanisms and has the most linear scale. A good design and high-quality magnetic vanes are required for instruments of good grade. Note the aluminum vane attached to the shaft, just under the pointer, which rotates in a closely fitting chamber to bring the pointer to rest quickly.

A variation of the radial-vane instrument is the *concentric-vane* repulsion movement, which is shown in the phantom photograph of Fig. 5-4. This instrument has two concentric vanes. One vane is rigidly attached to the coil frame; the other can rotate coaxially inside the stationary vane. Both vanes are magnetized 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 which is a function of the current in the coil. Controlled by a spring 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, it may be used on dc or ac; it is most commonly used for ac measurements.

Damping is obtained by a light aluminum *damping vane,* strengthened by flanges on all sides, rotating with small clearance in a closed air chamber. When used on ac, the actual operating torque is pulsating and this may cause vibration

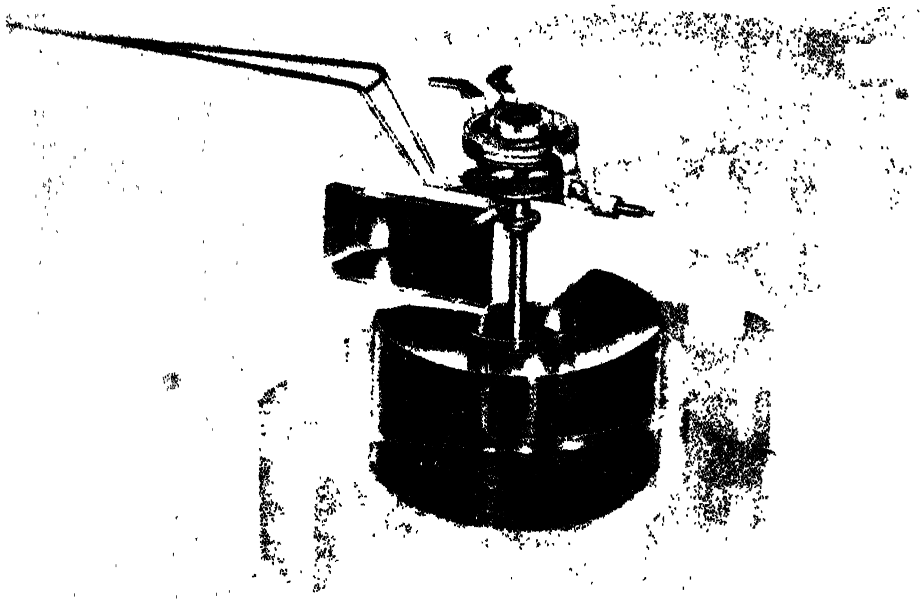


Figure 5-4
Phantom photograph of a concentric-vane moving-iron instrument. The figure shows details of the indicator with its counterweight, control spring, and damping vane. The moving vane may be seen as distinguished from the fixed vane in its brass retainer and is indicated by the lightly shaded area (*Courtesy Weston Instruments Inc.*).

of the pointer tip. The rigid trussed pointer construction (Fig. 5-4) effectively eliminates such vibration over a wide frequency range and serves to prevent bending of the pointer on heavy overloads.

The concentric-vane moving-iron instrument is only moderately sensitive and has square-law scale characteristics. It is possible to modify the shape of the vanes to secure special scale characteristics, thereby "opening the scale" where desired.

Accuracy of moving-iron instruments is limited by several factors. Firstly, the magnetization curve of the iron vanes is nonlinear. At low current values, the peak of the ac produces a greater displacement per unit current than the average value, resulting in an ac reading which may be appreciably higher than the equivalent dc reading at the lower end of the scale. Similarly, at the high end of the scale, the knee of the magnetization curve is approached, and the peak value of the ac will produce less deflection per unit current than the average value, so that the ac reading will be lower than the equivalent dc value.

Hysteresis in the iron and eddy currents in the vanes and other metal parts of the instrument further affects the accuracy of the reading. The flux density, even at full-scale values of current, is quite small, so that the instrument has a rather 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 will convert iron-vane movements into *voltmeters*; adding a shunt, similarly, will produce different *current ranges*. When the iron-vane movement is used as an ac voltmeter, the frequency increases the impedance of the instrument circuit and therefore tends to give a lower reading 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 instruments may be used within the accuracy tolerance from 25–125 Hz. Special compensating circuits may improve the performance of the meter at higher frequencies although the upper frequency limit is not easily extended beyond about 2,500 Hz. Although these instruments will respond to dc, they cannot be used as transfer instruments. Nevertheless they are very popular because they are cheap and rugged and perform adequately within their stated limitations.

5-4 Rectifier-Type ac Instruments

One obvious answer to the question of ac measurement is found by using a rectifier to convert ac into a unidirectional dc and then to use a dc movement to indicate the value of the rectified ac. This method is very attractive, because a dc movement generally has a higher sensitivity than either the electrodynamometer or the moving-iron instrument.

Rectifier-type instruments generally use a PMMC movement in combination with some chosen rectifier arrangement. The rectifier element may consist of a copper-oxide or selenium cell, a germanium or a silicon diode. Copper oxide and selenium rectifiers are rapidly becoming obsolete, because they have small inverse voltage ratings and can handle only limited amounts of current. Germanium diodes have a peak inverse voltage (PIV) on the order of 300 V and a current rating of approximately 100 mA. Low-current silicon diode rectifiers have a PIV of up to 1,000 V and a current rating on the order of 500 mA. (High-current silicon diodes are seldom used in indicating instruments but find extensive use in power-supply applications requiring output currents as high as 85 A per rectifying element.)

Rectifiers for instrument work sometimes consist of four elements in a bridge configuration, providing full-wave rectification. Figure 5-5 shows an

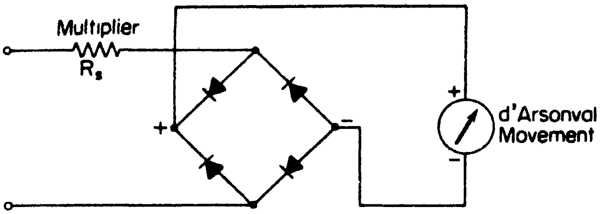


Figure 5-5
Principle of a full-wave rectifier ac voltmeter circuit.

ac voltmeter circuit consisting of a multiplier, a bridge rectifier, and a PMMC movement.

The *bridge circuit* is a full-wave rectifier, producing a pulsating unidirectional current through the meter movement over the complete cycle of the input voltage. Because of the inertia of the movable coil, the meter will indicate a steady deflection proportional to the average value of the current. For practical purposes, since currents and voltages are expressed in rms values, the meter scale is usually calibrated in terms of the *rms value* of an *alternating sine-wave* input. A *nonsinusoidal* waveform has an average value that may differ considerably from the average value of a pure sine wave (for which the meter is calibrated) and the indicated reading may be quite erroneous. The *form factor* relates the average value and the rms value of time varying voltages and currents

$$\text{form factor} = \frac{\text{effective value of the ac wave}}{\text{average value of the ac wave}}$$

For a sinusoidal waveform, the form factor is

$$\text{form factor} = \frac{E_{rms}}{E_{av}} = \frac{(\sqrt{2}/2)E_m}{(2/\pi)E_m} = 1.11 \tag{5-1}$$

Example 5-1: An experimental ac voltmeter uses the circuit of Fig. 5-5, where the PMMC movement has an internal resistance of 50 Ω and requires a dc current of 1 mA for full-scale deflection. Assuming ideal diodes (zero forward resistance and infinite reverse resistance), calculate (a) the value of the multiplier R_s to obtain full-scale meter deflection with 10-V ac (rms) applied to the input terminals, (b) the sensitivity of the ac voltmeter.

SOLUTION:
(a) For full-wave rectification,

$$E_{dc} = \frac{2}{\pi}E_m = \frac{2\sqrt{2}}{\pi}E_{rms} = 0.9\,E_{rms}$$

and $$E_{dc} = 0.9 \times 10\text{ V} = 9\text{ V}$$

The total circuit resistance, neglecting the forward diode resistance, is

$$R_t = R_s + R_m = \frac{9\text{ V}}{1\text{ mA}} = 9\text{ k}\Omega$$

$$R_s = 9{,}000\ \Omega - 50\ \Omega = 8{,}950\ \Omega$$

(b) The sensitivity or ohms-per-volt rating is

$$S = \frac{9{,}000\ \Omega}{10\text{ V}} = 900\ \Omega/\text{V}$$

Note that the voltmeter of Example 5-1 has a scale suitable only for sinusoidal ac measurements. The form factor of Eq. (5-1) is therefore also the factor by which the actual (average) dc current is multiplied to obtain the equivalent rms scale markings.

The ideal rectifier element should have zero forward and infinite reverse resistance. In practice, however, the rectifier is a nonlinear device, indicated by the characteristic curves of Fig. 5-6. At low values of forward current,

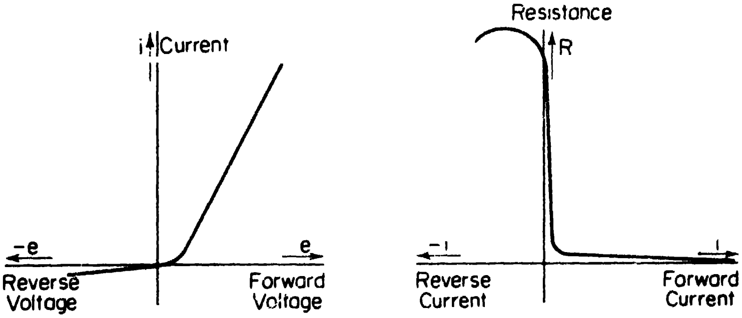


Figure 5-6
Characteristic curves of a solid state rectifier.

the rectifier operates in an extremely nonlinear part of its characteristic curve, and the resistance is large as compared to the resistance at higher current values. The lower part of the ac scale of a low-range voltmeter is therefore often crowded, and most manufacturers supply a separate low-voltage scale, calibrated especially for this purpose. The high resistance in the early part of the rectifier characteristics also sets a limit on the sensitivity which can be obtained in microammeters and voltmeters.

The resistance of the rectifying element changes with varying temperature, one of the major drawbacks of rectifier-type ac instruments. The meter accuracy is usually satisfactory under normal operating conditions at room temperature and is generally on the order of ± 5 per cent of full-scale reading for sinusoidal waveforms. At very much higher or lower temperatures, the resistance of the rectifier changes the total resistance of the measuring circuit sufficiently to cause the meter to be gravely in error. If large temperature variations are expected, the meter should be enclosed in a temperature-controlled box.

Frequency also a cts the operation of the rectifier elements. The rectifier exhibits capacitive properties and tends to bypass the higher frequencies. Meter readings may be in error by as much as 0.5 per cent decrease for every 1-kHz rise in frequency.

General rectifier-type ac voltmeters often use the arrangement shown in Fig. 5-7. Two diodes are used in this circuit, forming a full-wave rectifier with the

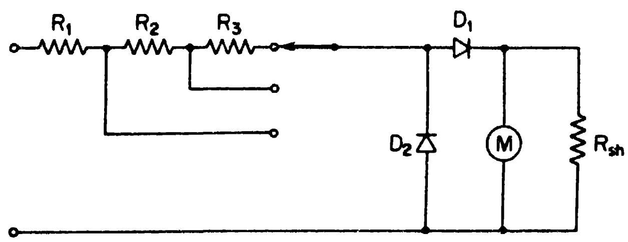


Figure 5-7
Typical ac voltmeter circuit, as found in a commercial multimeter.

movement so connected that it receives only half of the rectified current. Diode D_1 conducts during the positive half cycle of the input waveform and causes the meter to deflect according to the average value of this half cycle. The meter movement is shunted by a resistance R_{sh}, in order to draw more current through the diode D_1 and move its operating point into the linear portion of the characteristic curve. In the absence of diode D_2, the negative half cycle of the input voltage would apply a reverse voltage across diode D_1, causing a small leakage current in the reverse direction. The average value of the complete cycle would therefore be lower than it should be for half-wave rectification. Diode D_2 deals with this problem. On the negative half cycle, D_2 conducts heavily and the current through the measuring circuit, which is now in the opposite direction, bypasses the meter movement.

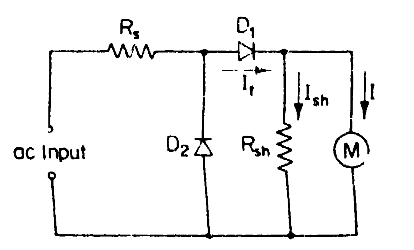


Figure 5-8
Illustrating the computation of the multiplier resistor and the ac voltmeter sensitivity.

The commercial multimeter often uses the same scale markings for both its dc and ac voltage ranges. Since the dc component of a sine wave for half-wave rectification equals 0.45 times the rms value, a problem arises immediately. In order to obtain the same deflection on corresponding dc and ac voltage ranges, the multiplier for the ac range must be lowered proportionately. The circuit of Fig. 5-8 illustrates a solution to the problem and is discussed in some detail in Example 5-2.

Example 5-2: A meter movement has an internal resistance of 100 Ω and requires 1-mA current for full-scale deflection. The shunting resistor R_{sh}, placed across the movement, has a value of 100 Ω. Diodes D_1 and D_2 each has an average forward resistance of 400 Ω and is assumed to have infinite resistance in the reverse direction. Calculate (a) value of the multiplier R_s, (b) voltmeter sensitivity on the ac range.

SOLUTION:

(a) To provide the movement with 1-mA dc for its full-scale deflection, the source must also deliver the current through the shunting resistor R_{sh}. Since R_m and R_{sh} are both 100 Ω, the total current supplied by the source must be $I_t = 2$ mA. For half-wave rectification the equivalent dc value of the rectified ac voltage will be

$$E_{dc} = 0.45\, E_{rms} = 0.45 \times 10\text{ V} = 4.5\text{ V}$$

The total resistance of the instrument circuit, R_t, then is

$$R_t = \frac{E_{dc}}{I_t} = \frac{4.5\text{ V}}{2\text{ mA}} = 2{,}250\ \Omega$$

This total resistance is made up of several parts. Since we are interested only in the resistance of the circuit during the half cycle that the movement receives current, we can eliminate the infinite resistance of diode D_2 from the circuit. Therefore,

$$R_t = R_s + R_{D_1} + \frac{R_m R_{sh}}{R_m + R_{sh}} \quad \text{and}$$

$$R_t = R_s + 400 + \frac{100 \times 100}{200} = R_s + 450\ \Omega$$

The value of the multiplier therefore is

$$R_s = 2{,}250 - 450 = 1{,}800\ \Omega.$$

(b) The sensitivity of the voltmeter on this 10-V ac range is

$$S = \frac{2{,}250\ \Omega}{10\text{ V}} = 225\ \Omega/\text{V}$$

The same movement, used in a dc voltmeter, would have given a sensitivity figure of 1,000 Ω/V.

Section 4-10 dealt with the dc circuitry of a typical multimeter, using the simplified circuit diagram of Fig. 4-25. The circuit for measuring ac volts (subtracted from Fig. 4-25), is reproduced in Fig. 5-9. Resistances R_9, R_{13}, R_7, and R_6 form a chain of multipliers for the voltage ranges of 1,000 V,

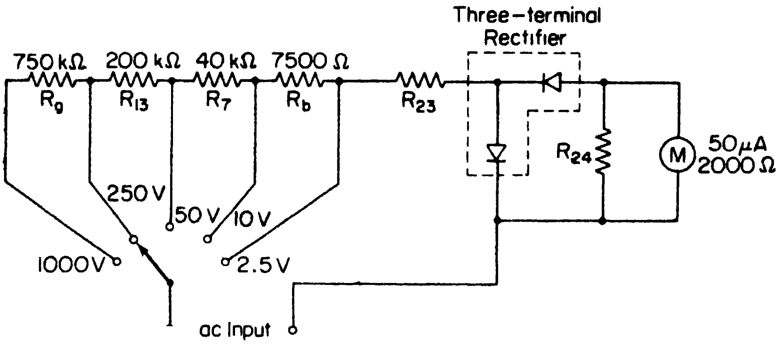


Figure 5-9
Multirange ac voltmeter circuit of the Simpson Model 260 multimeter (*Courtesy of the Simpson Electric Company*).

250 V, 50 V, and 10 V, respectively, and their values are indicated in the diagram of Fig. 5-9. On the 2.5-V ac range, resistor R_{23} acts as the multiplier and corresponds to the multiplier R_s of Example 5-2 shown in Fig. 5-8. Resistor R_{24} is the meter shunt and again acts to improve the rectifier operation. Both values are unspecified in the diagram and are factory selected. A little thought, however, will convince us that the meter-shunt resistance could be 2,000 Ω, equal to the meter resistance. If the average forward resistance of the rectifier elements is 500 Ω (a reasonable assumption), then resistance R_{23} must have a value of 1,000 Ω. This follows because the meter sensitivity on the ac ranges is given as 1,000 Ω/V; on the 2.5-V ac range, the circuit must therefore have a total resistance of 2,500 Ω. This value is made up of the sum of R_{23}, the diode forward resistance, and the combination of movement and shunt-resistance, as shown in Example 5-2.

5-5 Thermoinstruments

The historical forerunner of the thermoinstruments is the ***hot-wire mechanism***, shown schematically in Fig. 5-10. The current under measurement passes through a fine wire tightly stretched between two terminals. A second wire is attached to the fine wire at one end and, at the other, to a spring, which exerts a downward pull on the fine wire. This second wire passes over a roller to which the pointer is connected. The current under measurement causes the fine wire to heat and thus expand approximately in proportion to the heating current squared. The change in wire length drives the pointer, which indicates the magnitude of the current. Instability due to wire stretch, sluggishness in response, and lack of ambient temperature compensation have made this

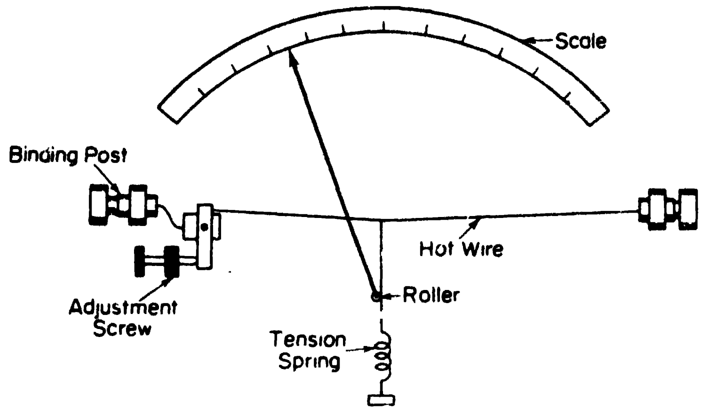


Figure 5-10
A schematic representation of the hot-wire ammeter.

mechanism commercially unsatisfactory. Hot-wire mechanisms are now obsolete, being replaced by the more sensitive, more accurate, and better-compensated combination of thermoelectric heating element and permanent-magnet moving-coil instrument.

The basic type of *thermoelement* and *PMMC movement* for the indication of ac and dc is shown in Fig. 5-11. This combination is called a *thermocouple instrument*, since its operation depends on the action of the thermocouple element. When two dissimilar metals are mutually in contact, a voltage is generated at the junction of the two dissimilar metals. This voltage rises in proportion to the temperature of the junction. In Fig. 5-11, *CE* and *DE* represent the two dissimilar metals, joined at point *E*, and are drawn as a light and a heavy line, to indicate dissimilarity. The potential difference between *C* and *D* depends on the temperature of the so-called *cold junction, E*. A rise in temperature causes an increase in the voltage and this is used to advantage in the thermocouple.

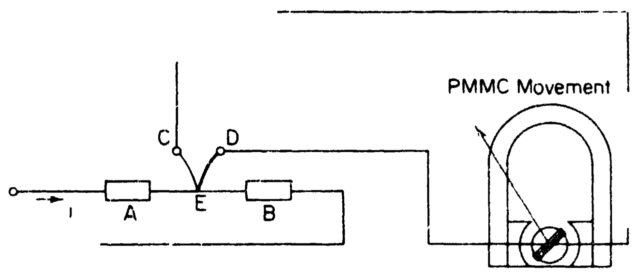


Figure 5-11
Schematic representation of a basic thermocouple instrument using thermocouple CDE and a PMMC movement.

Heating element AB, which is in mechanical contact with the junction of the two metals at point E, forms part of the circuit in which the current is to be measured. (AEB is called the *hot junction.*) Heat energy generated by the current in the heating element raises the temperature of the cold junction and causes an increase in the voltage generated across terminals C and D. This potential difference causes a dc current through the PMMC-indicating instrument. The heat generated by the current is directly proportional to the current squared (I^2R), and the temperature rise (and hence the generated dc voltage) is proportional to the square of the rms current. The deflection of the indicating instrument will therefore follow a square-law relationship, causing crowding at the lower end of the scale and spreading at the high end. The arrangement of Fig. 5-11 does not provide compensation for ambient temperature changes.

The *compensated thermoelement,* shown schematically in Fig. 5-12, produces

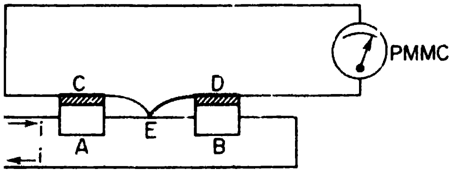


Figure 5-12
Shown here is a compensated thermocouple to measure the thermovoltage produced by the current I alone. The couple terminals C and D are in thermal contact with heater terminals A and B, but are electrically insulated from them.

a thermoelectric voltage in the thermocouple CED, which is directly proportional to the current through the circuit AB. Since the couple-voltage developed is a function of the temperature difference between its hot and cold ends, this temperature difference must be caused only by the current being measured. For accurate measurements, therefore, points C and D must be at the mean temperature of points A and B. This is accomplished by attaching the couple ends C and D to the center of separate copper strips, whose ends are in thermal contact with A and B, but which are electrically insulated from them.

Self-contained thermoelectric instruments of the compensated type are available in the 0.5–20-A range. Higher current ranges are available, but in this case the heating element is external to the indicator. Thermoelements used for current ranges over 60 A are generally provided with air cooling fins.

Current measurements in the lower ranges, from approximately 0.1–0.75 A, use a *bridge-type thermoelement,* shown schematically in Fig. 5-13. This arrangement does not use a separate heater; the current to be measured passes directly through the thermoelements and raises their temperature in proportion to I^2R. The cold junctions (marked *c*) are at the pins which are embedded in the insulating frame and the hot junctions (marked *h*) are at splices midway between the pins. The couples are arranged as shown in Fig. 5-13 and the resultant

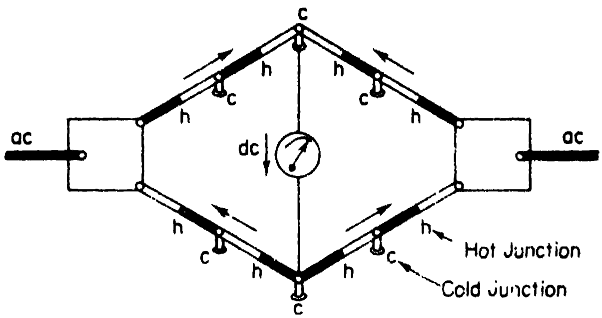


Figure 5-13
The bridge-type thermocouple instrument.

thermal voltage generates a dc potential difference across the indicating instrument. Since the bridge arms have equal resistances, the ac voltage across the meter is 0 V and no ac passes through the meter. The use of several thermocouples in a series arrangement provides a greater output voltage and deflection than is possible with a single element, resulting in an instrument with increased sensitivity.

Thermoinstruments may be converted into voltmeters using low-current thermocouples and suitable series resistors. Thermocouple voltmeters are available in ranges of up to 500 V and sensitivities of approximately 100 to 500 Ω/V.

A major advantage of a thermocouple instrument is that its accuracy can be as high as 1 per cent, up to frequencies of approximately 50 MHz. For this reason, it is classified as an RF *instrument*. Above 50 MHz, the skin effect tends to force the current to the outer surface of the conductor, increasing the effective resistance of the heating wire and reducing instrument accuracy. For small currents, up to 3 A, the heating wire is solid and very thin. Above 3 A, the heating element is made of a tubular design to reduce the errors due to skin effect at higher frequencies.

A thermocouple arrangement, related to the bridge-type heating element, is used in the *thermal watt converter*. This device permits measurement of ac and dc *power* by thermoelectric means. From basic ac theory, we know that power is measured in watts and is expressed as $W = EI \cos \theta$, where E and I represent the phasor quantities of voltage and current, respectively, and θ represents the phase angle between them. Referring to the phasor diagram of Fig. 5-14, where a voltage phasor, E, and a current phasor, I, have been placed at a phase angle θ, we see that the sum, S, of the two phasors can be found by the relationship

$$S^2 = E^2 + I^2 + 2\,EI \cos \theta \tag{5-2}$$

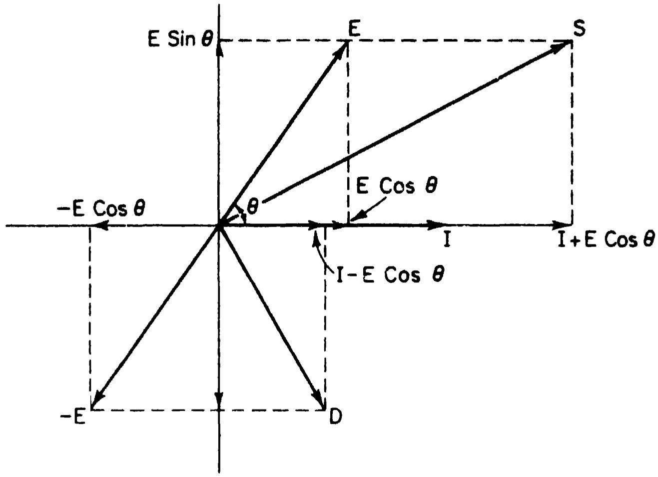


Figure 5-14
Illustrating the geometrical relationship between the sum (S) and the difference (D) of two vectors E and I at a phase angle θ.

where S represents the sum of the E phasor and the I phasor. Similarly, the difference, D, between the two phasors is found by the expression

$$D^2 = E^2 + I^2 - 2\,EI\cos\theta \tag{5-3}$$

Subtracting Eq. (5-3) from Eq. (5-2), we obtain

$$S^2 - D^2 = 4\,EI\cos\theta \tag{5-4}$$

where $EI\cos\theta$ is the power developed by the two phasor quantities in an electric circuit.

A circuit arrangement, therefore, capable of measuring the quantity $S^2 - D^2$ can also measure a quantity proportional to $EI\cos\theta$, representing power. A thermoinstrument, capable of measuring power, is called a ***thermal watt converter.***

Figure 5-15 shows the schematic diagram of the elementary circuit of the thermal watt converter. For a given instant of time, the plain arrow shows the instantaneous direction of current from the current transformer. The flagged arrows show the corresponding instantaneous direction of current in the potential circuit. The heating element of thermocouple A receives the ***sum*** of the currents produced by the current transformer and the potential circuit. The heating element of thermocouple B, however, receives the ***difference*** of these currents. Through proper design, the heat generated in the thermocouples and hence the developed emf, is proportional to the square of the current in the heater. Therefore, thermocouple A develops an emf proportional to S^2 and thermocouple B develops an emf proportional to D^2. The voltage outputs

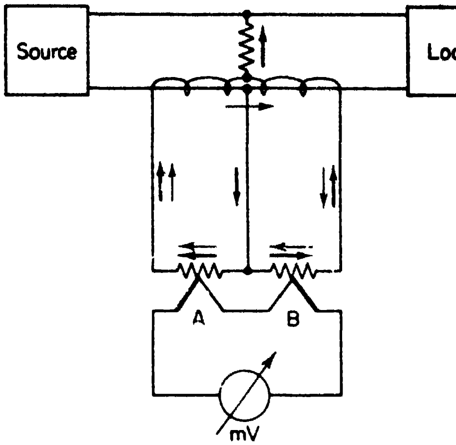


Figure 5-15
Elementary circuit of a thermal converter.

of the thermocouples are connected so that they oppose each other. The total emf measured by the meter is proportional to $S^2 - D^2$, which was shown, by Eq. (5-4), to represent power.

In practice, *chains* of thermocouples are used instead of single couples to obtain greater developed voltages. The couples are *self-heating,* similar to those of the bridge-type element discussed earlier. This results in the practical circuit arrangement shown in Fig. 5-16.

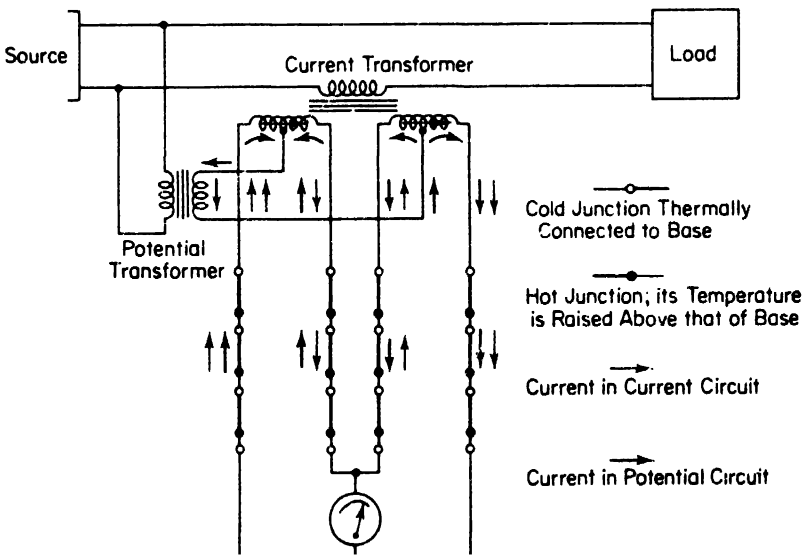


Figure 5-16
Circuit diagram of a thermal watt converter (*Courtesy Weston Instruments Inc.*).

Thermal watt converters are extremely reliable instruments and are widely used to measure power in different circuits, their outputs being summed and applied to a recording potentiometer which will draw a graph of the total power consumed by the circuits. They are also used for dc and ac instrument calibration and for instrumentation process monitoring.

5-6 The Electrostatic Voltmeter

The electrostatic voltmeter is the only instrument that measures voltage *directly* rather than by the effect of the current it produces. This instrument has one distinguishing characteristic: it *consumes no power* (except during the brief transient period of initial connection to the circuit) and it represents therefore an *infinite impedance* to the circuit under measurement. Its action depends on the reaction between two electrically charged bodies (Coulomb's law). The electrostatic mechanism resembles a variable capacitor, where the force existing between the two parallel plates is a function of the potential difference applied to them. Figure 5-17 illustrates the principle of this instrument.

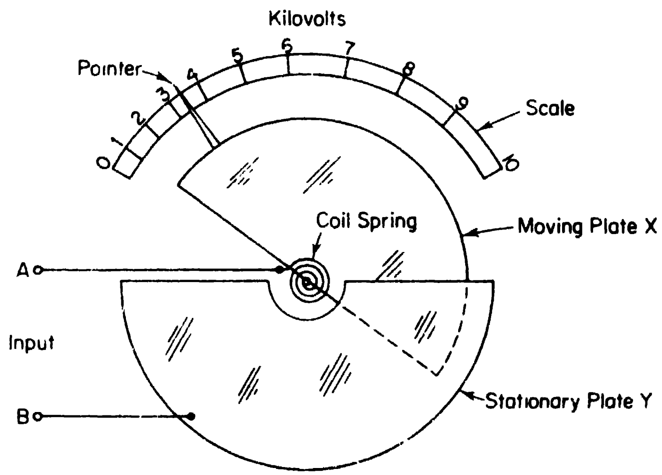


Figure 5-17
Schematic representation of an electrostatic voltmeter.

Plates X and Y constitute a capacitor whose capacitance increases as pointer P moves to the right. The motion of the pointer is opposed by a coil spring which also serves to provide electrical contact between the connecting terminal A and the plate X. When terminals X and Y are connected to points of opposite potential, the plates possess opposite charges, and the force of attraction between the two bodies of equal but opposite charge forces the pointer to move to the right. The pointer will come to rest when the torque caused by the electrical attraction between the plates equals the opposing torque of the coil spring.

Analysis of the energy stored in the electric field between the capacitor plates allows us to determine an expression for the developed torque in terms of the applied voltage. The instantaneous voltage, e, across the capacitor is $e = q/C$, neglecting the leakage resistance of the air capacitor. The instantaneous energy stored in the electric field is

$$W = \frac{1}{2}\frac{q^2}{c} = \frac{1}{2}Ce^2 \tag{5-5}$$

The instantaneous torque may be found by keeping e constant and permitting the movable plates to undergo a small angular displacement, $d\theta$. The developed torque then is

$$T_e = \frac{\partial W}{\partial \theta} = \frac{\partial}{\partial \theta}\left(\frac{1}{2}Ce^2\right) = \frac{1}{2}e^2\frac{\partial c}{\partial \theta} \tag{5-6}$$

Equation (5-6) indicates that the instantaneous torque is proportional to the square of the instantaneous voltage and also depends on the manner in which C changes with θ. The average torque over an entire period T of the alternating voltage is

$$T_{av} = \frac{1}{T}\int_0^T T_e\,dt = \frac{1}{T}\int_0^T \frac{1}{2}e^2\frac{\partial C}{\partial \theta}\,dt = KE_{rms}^2 \tag{5-7}$$

The deflecting torque, expressed in Eq. (5-7), is directly proportional to the square of the applied voltage, whatever its waveform, and the deflection of the electrometer may be calibrated directly in rms volts.

The electrometer can be used on *either* dc or ac and over a fairly large range of frequencies. The instrument may be calibrated with dc and the calibration is valid for ac since the deflection is independent of the waveform of the applied voltage. Since the electrometer is a "square-law" instrument, there will be no waveform error as found in the rectifier-type voltmeter.

When the electrometer is first connected to a source, it draws a momentary charging current which decays exponentially. Once the plates are charged, no more current is drawn from the circuit and the meter represents infinite impedance.

The use of the instrument is limited to certain special applications, particularly in ac circuits of relatively *high voltage,* where the current taken by other instruments would result in erroneous indications. A protective resistor is generally used in series with the instrument to limit the current in case of a short circuit between the plates.

An interesting application of the same principle of electrostatic attraction or repulsion between two parallel plates is found in the *disk electrometer.* This instrument consists of two very large parallel plates, mounted in a shielded case and using quartz support pillars. The force of attraction between the parallel plates caused by the application of a potential difference is measured, and by using the exact dimensions of the plates, and their separation, the voltage between the plates is calculated. The National Bureau of Standards (NBS) uses an instrument of this type for voltage standardization up to 300,000 V. Using this high-voltage electrometer, the ratio of transformation of high-voltage potential transformers can be directly checked by an independent method.

5-7 The Electrodynamometer as a Wattmeter in Single-Phase and Polyphase Power Measurements

The electrodynamometer movement is used extensively in the measurement of power. It may be used to indicate *both* dc and ac power for any waveform of voltage and current and it is not restricted to sinusoidal waveforms. As described in Sec. 5-2, the electrodynamometer used as a voltmeter or an ammeter, has the fixed coils and the movable coil connected in series, thereby reacting to the effect of the current squared. When used as a *single-phase power meter*, the coils are connected in a different arrangement, shown in Fig. 5-18.

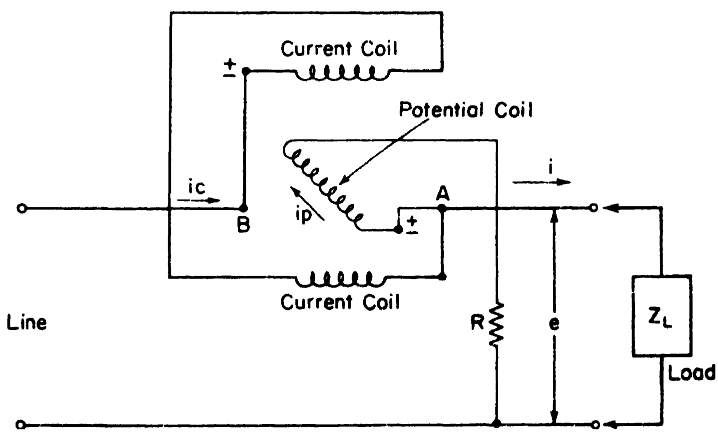


Figure 5-18
Diagram of an electrodynamometer wattmeter, connected to measure the power of a single-phase load.

The fixed coils, or *field coils*, shown here as two separate elements, are connected in series and carry the total line current (i_c). 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 (i_p). The instantaneous value of the current in the movable coil is $i_p = e/R_p$, where e is the instantaneous voltage across the power line and R_p is the total resistance of the movable coil and its series resistor. The deflection of the movable coil is proportional to the product of these two currents, i_c and i_p, and we can write for the average deflection over one period:

$$\theta_{av} = K\frac{1}{T}\int_0^T i_c i_p\, dt \tag{5-8}$$

where θ_{av} = average angular deflection of the coil
K = instrument constant
i_c = instantaneous current in the field coils
i_p = instantaneous current in the potential coil

Assuming for the moment that i_c is equal to the load current, i (actually, $i_c = i_p + i$) and using the value for $i_p = e/R_p$, Eq. (5-8) reduces to

$$\theta_{av} = K\frac{1}{T}\int_0^T i\frac{e}{R_p}\, dt = K_2\frac{1}{T}\int_0^T ei\, dt \tag{5-9}$$

By definition, the average power in a circuit is

$$P_{av} = \frac{1}{T}\int_0^T ei\, dt \tag{5-10}$$

which indicates that the electrodynamometer movement, connected in the configuration of Fig. 5-18, has a deflection proportional to the average power. If e and i are sinusoidally varying quantities of the form $e = E_m \sin \omega t$ and $i = I_m \sin (\omega t \pm \theta)$, Eq. (5-9) reduces to

$$\theta_{av} = K_3 EI \cos\theta \tag{5-11}$$

where E and I represent the rms values of the voltage and the current and θ represents the phase angle between voltage and current. Equations (5-9) and (5-10) show that the electrodynamometer indicates the average power delivered to the load.

Wattmeters have one voltage terminal and one current terminal marked "±." When the marked current terminal is connected to the incoming line and the marked voltage terminal is connected to the line side in which the current coil is connected, the meter will always read up-scale when power is connected to the load. If, for any reason (as in the two-wattmeter method of measuring three-phase power), the meter should read backward, the *current* connections should be reversed and not the voltage connections.

The electrodynamometer wattmeter consumes some power for maintenance of its magnetic field, but this is usually so small, compared to the load power, that it may be neglected. If a correct reading of the load power is required, the current coil should carry exactly the load current and the potential coil should be connected across the load terminals. With the potential coil connected to point A, as in Fig. 5-18, the load voltage is properly metered but the current through the field coils is greater by the amount i_p. The wattmeter therefore reads high by the amount of additional power loss in the potential circuit. If, on the other hand, the potential coil is connected to point B in Fig. 5-18, the field coils meter the correct load current, but the voltage across the potential coil is higher by the amount of the drop across the field coils. The wattmeter will again read high, but now by the amount of the I^2R losses in the field windings. Choice of the correct connection depends on the situation. Generally, connection of the potential coil at point A is preferred for high-current, low-voltage loads; connection at B is preferred for low-current, high-voltage loads.

It is also possible to measure and subtract the instrument loss from the readings, particularly where the power measured at the load is small.*

The difficulty placing the connection of the potential coil is overcome in the *compensated* wattmeter, shown schematically in Fig. 5-19. The current coil

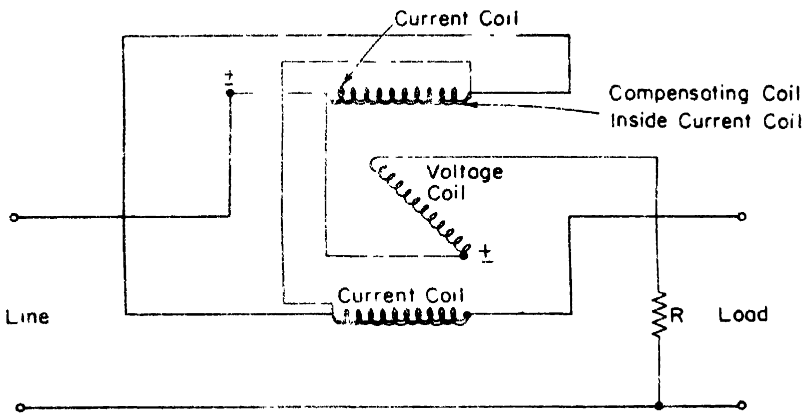


Figure 5-19
Diagram of a compensated wattmeter, where the effect of the current in the potential coil is cancelled by the current in the compensating winding.

consists of two windings, each having the same number of turns. One winding uses heavy wire which carries the load current plus the current for the potential coil. The other winding uses thin wire and carries only the current to the voltage coil. This current, however, is in a direction opposite to the current in the heavy winding, causing a flux which opposes the main flux. The effect of i_p is therefore canceled out and the wattmeter indicates the correct power.

Power measurements in a *polyphase* system require the use of two or more wattmeters. The total real power is then found by algebraically adding the readings of the individual wattmeters. Blondel's theorem states that real power can be measured by one less wattmeter element than the number of wires in any

*Cf. Frank A. Laws, *Electrical Measurements*, 2nd ed. (New York: McGraw-Hill Book Company, Inc., 1961), p. 316.

polyphase system, provided that one wire can be made common to all the potential circuits. Figure 5-20(a) shows the connection of two wattmeters to measure the power consumption of a balanced three-wire delta-connected three-phase load.

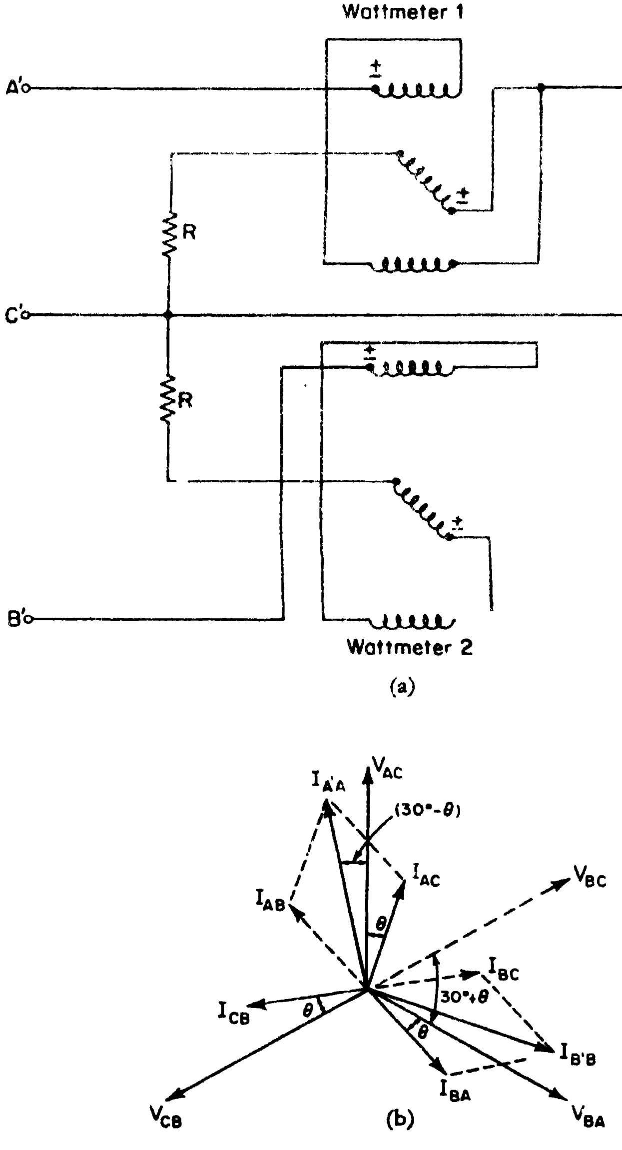


Figure 5-20

(a) Two wattmeters connected to measure the total power in a three-phase three-wire system. (b) Phasor diagram of voltages and currents in the three-phase three-wire system. The angle between phase voltage and phase current is indicated by θ.

The current coil of wattmeter 1 is connected in line A and its voltage coil is connected between line A and line C. The current coil of wattmeter 2 is connected in line B and its voltage coil is connected between line B and line C. The total power, consumed by the balanced three-phase load, equals the algebraic sum of the two wattmeter readings.

The phasor diagram of Fig. 5-20(b) shows the three phase voltages V_{AC}, V_{CB}, and V_{BA} and the three phase currents I_{AC}, I_{CB}, and I_{BA}. The delta-connected load is assumed to be inductive and the phase currents lag the phase voltages by an angle θ. The current coil of wattmeter 1 carries the line current $I_{A'A}$, which is the vector sum of the phase currents I_{AC} and I_{AB}. The potential coil of wattmeter 1 is connected across the line voltage V_{AC}. Similarly, the current coil of wattmeter 2 carries the line current $I_{B'B}$, which is the vector sum of the phase currents I_{BA} and I_{BC}, while the voltage across its potential coil is the line voltage V_{BC}. Since the load is balanced, the phase voltages and phase currents are equal in magnitude and we can write

$$V_{AC} = V_{BC} = V \quad \text{and} \quad I_{AC} = I_{CB} = I_{BA} = I$$

The power, represented by the currents and voltages of each wattmeter, is

$$W_1 = V_{AC} I_{A'A} \cos(30^\circ - \theta) = VI \cos(30^\circ - \theta) \tag{5-12}$$

$$W_2 = V_{BC} I_{B'B} \cos(30^\circ + \theta) = VI \cos(30^\circ + \theta) \tag{5-13}$$

and

$$\begin{aligned} W_1 + W_2 &= VI \cos(30^\circ - \theta) + VI \cos(30^\circ + \theta) \\ &= \cos 30^\circ \cos\theta + \sin 30^\circ \sin\theta \\ &\quad + \cos 30^\circ \cos\theta - \sin 30^\circ \sin\theta \\ &= \sqrt{3}\, VI \cos\theta \end{aligned} \tag{5-14}$$

Equation (5-14) is the expression for the total power in a three-phase circuit and the two wattmeters of Fig. 5-20(a) therefore correctly measure this total power. It may be shown that the algebraic sum of the readings of the two wattmeters will give the correct value for power under any condition of unbalance, power factor or waveform.

If the neutral wire of the three-phase system is also present, as in the case of a four-wire star-connected load—according to Blondel's theorem—three wattmeters would be needed to make the total real power measurement. (In Prob. 12, the reader is asked to prove that three wattmeters measure total power in a four-wire system.)

5-8 Reactive Power Measurement

Reactive power, supplied to an ac circuit, is expressed in a unit called VAR (volt-ampere-reactive), thereby making a distinction between the real power

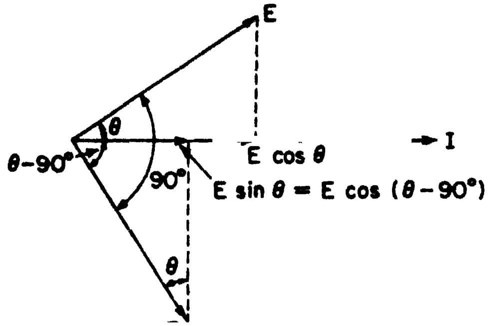


Figure 5-21
Vector diagram of voltage and current phasors, illustrating a shift of −90° of the voltage phasor.

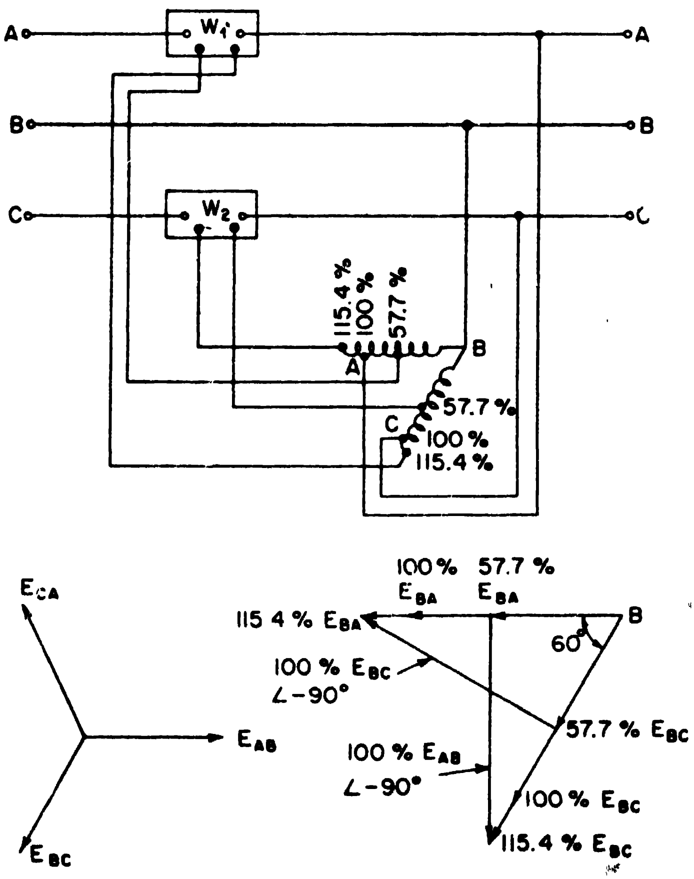


Figure 5-22
Reactive power measurement.

and the quadrature reactive power. Figure 5-21 shows the two phasors E and I, representing the voltage and the current, at a phase-angle θ. Real power is the product of the in-phase components of voltage and current ($EI \cos \theta$), and reactive power, the product of the quadrature components, equals $EI \sin \theta$ or $EI \cos (\theta - 90°)$. If the voltage is shifted through 90° from its true value, the in-phase component of the shifted voltage will be $E \cos (\theta - 90°)$ and the product of the in-phase components will be $EI \cos (\theta - 90°)$, which is the reactive power.

Any ordinary wattmeter, together with a suitable phase-shifting network, may be used to measure reactive power. In a single-phase circuit, a 90° phase shift can be accomplished using R, L, and C components, carefully proportioned. The common application of VAR measurement however is found in three-phase systems, where the required phase-shifting is done with two auto-transformers connected in an "open-delta" configuration (Fig. 5-22). The current coils of the wattmeters are connected in series with the line, as usual. The potential coils are connected to the auto-transformers in the manner indicated. Phase-line B is connected to the common terminal of the two transformers, and the phase A and C lines are connected to the 100 per cent taps on the transformers. Both transformers will produce 115.4 per cent of the line voltage across the total winding. The potential coil of wattmeter 1 is connected from the 57.7 per cent tap of transformer 1 to the 115.4 per cent tap on transformer 2, which produces a voltage equal to the line voltage but shifted by 90°. This is shown in the phasor diagram of Fig. 5-22. The voltage-coil of wattmeter 2 is connected in a similar way. Since both voltage coils now receive an emf equal to the line voltage but displaced by 90°, the wattmeters will read the reactive power consumed by the load. The arithmetic sum of the two wattmeter readings represent the total reactive power supplied to the load. In a single instrument package, the combination of wattmeter and phase-shifting transformer is called the VARmeter.

5-9 The Watthour Meter

Although the watthour meter is not often found in a laboratory situation, it is very widely used for the commercial measurement of electrical energy. In fact it is evident wherever a power company supplies the industrial or domestic consumer with electrical energy. Figure 5-23 shows the elements of a single-phase watthour meter in schematic form.

The current coil is connected in series with the line and the voltage coil is connected across the line. Both coils are wound on a metal frame of special design, providing two magnetic circuits. A light aluminum disk is suspended

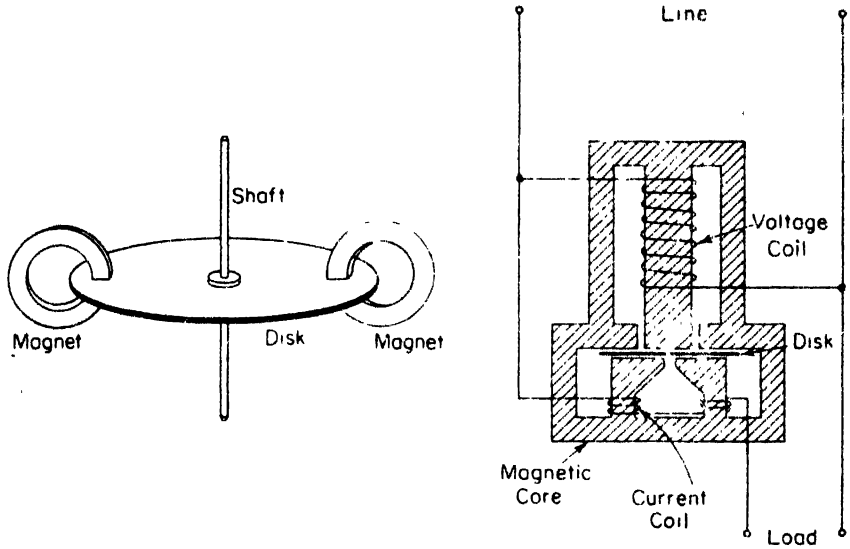


Figure 5-23
Elements of a single-phase watthour meter.

in the air gap of the current-coil field, which causes eddy currents to flow in the disk. The reaction of the eddy currents and the field of the voltage coil creates a torque (motor action) on the disk, causing it to rotate. The developed torque is proportional to the field strength of the voltage coil and the eddy currents in the disk which are in turn a function of the field strength of the current coil. The number of rotations of the disk is therefore proportional to the energy consumed by the load in a certain time interval and is commonly measured in terms of kilowatthours (kWh). The shaft which supports the aluminum disk is connected by a gear arrangement to the clock mechanism on the front of the meter, providing a decimally calibrated read-out of the number of kWh.

Damping of the disk is provided by two small permanent magnets, located opposite each other at the rim of the disk. Whenever the disks rotate, the permanent magnets induce eddy currents in them. These eddy currents react with the magnetic fields of the small permanent magnets, damping the motion of the disk. A typical single-phase watthour meter is shown in the photograph of Fig. 5-24.

Calibration of the watthour meter is performed under conditions of full rated load and 10 per cent of rated load. At full load, the calibration consists of adjustment of the position of the small permanent magnets until the meter reads correctly. At very light loads, the voltage component of the field produces a torque which is not directly proportional to the load. Compensation for the error is provided by inserting a shading coil or plate over a portion of

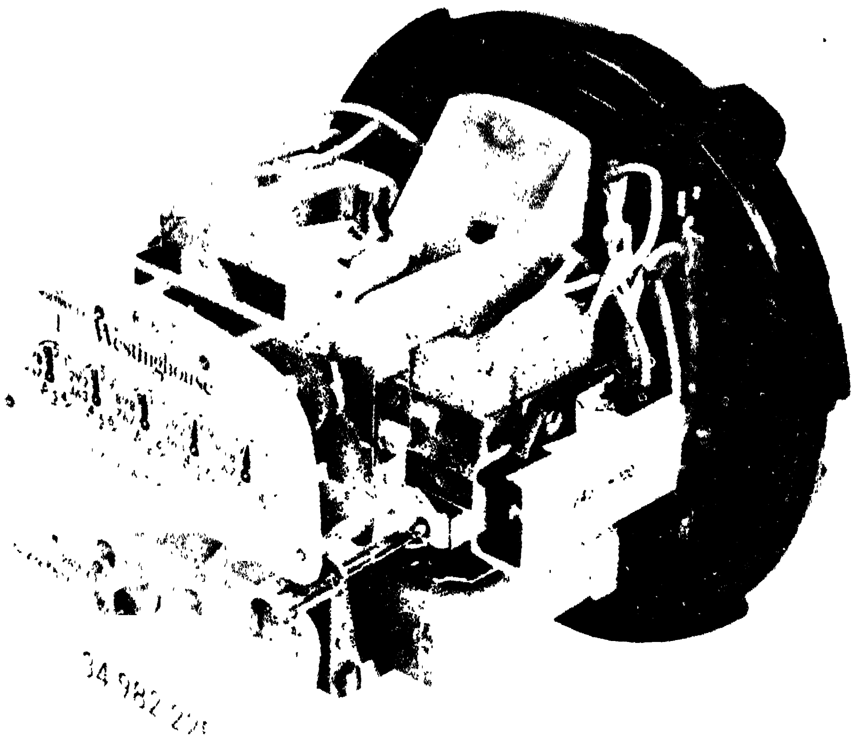


Figure 5-24
A watthour meter for industrial or domestic application. (*Courtesy Westinghouse Electric Corporation.*)

the voltage coil, with the meter operating at 10 per cent of rated load. Calibration of the meter at these two positions usually provides satisfactory readings at all other loads.

A unique design for the suspension of the disk is used in the *floating-shaft* watthour meter. Here the rotating shaft has a small magnet at each end, where the upper magnet of the shaft is attracted to a magnet in the upper bearing and the lower magnet of the shaft is attracted to a magnet in the lower bearing. The movement thus floats without touching either bearing surface, and the only contact with the movement is that of the gear connecting the shaft with the gear train.

Measurements of energy in three-phase systems are performed with polyphase watthour meters. The current coils and voltage coils are connected similar to those of the three-phase wattmeter of Fig. 5-20. Each phase of the watthour meter has its own magnetic circuit and its own disk, but all the disks are mounted on a common shaft. The developed torque on each disk is mechanically summed and the total number of revolutions per minute of the shaft is proportional to the total three-phase energy consumed.

5-10 Power-Factor Meters

Several different types of *power-factor* meters have been developed. Since the power factor, by definition, is the cosine of the phase angle between voltage and current, power-factor measurements usually involve the determination of this phase angle. This is demonstrated in the operation of the *crossed-coil power-factor meter*. The instrument is basically an electrodynamometer movement, where the moving element consist of two *coils,* mounted on the same shaft but at *right* angles to each other. The moving coils rotate in the magnetic field provided by the field coil which carries the line current.

The connections for this meter in a single-phase circuit are shown in the circuit diagram of Fig. 5-25. The field coil is connected as usual in series with the

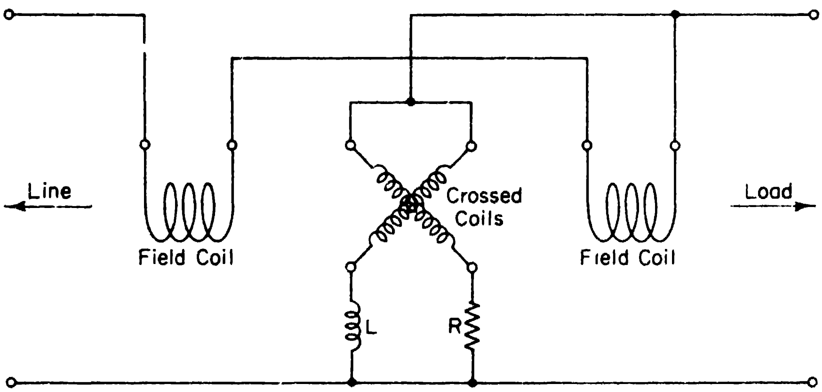


Figure 5-25
Connections for a single-phase crossed-coil power-factor meter.

line and carries the line current. One coil of the movable element is connected in series with a *resistor* across the lines and receives its current from the applied potential difference. The second coil of the movable element is connected in series with an *inductor* across the lines. Since no control springs are used, the balance position of the movable element depends on the resulting torque developed by the two crossed coils. When the movable element is in a balanced position, the contribution to the total torque by each element must be equal but of opposite sign. The developed torque in each coil is a function of the current through the coil and depends therefore on the impedance of that coil circuit.

The torque is also proportional to the mutual inductance between each part of the crossed coil and the stationary field coil. This mutual inductance depends on the angular position of the crossed-coil elements with respect to the position of the stationary field coil. When at balance the two torques are equated, it can be shown that the angular displacement of the movable element is a function of the phase angle between line current (field coil) and line voltage (crossed coils). The indication of the pointer, which is connected to the movable element, is calibrated directly in terms of the phase angle or power factor.

The *polarized-vane power-factor meter* is shown in the construction sketch of Fig. 5-26. This instrument is used primarily in three-phase power systems,

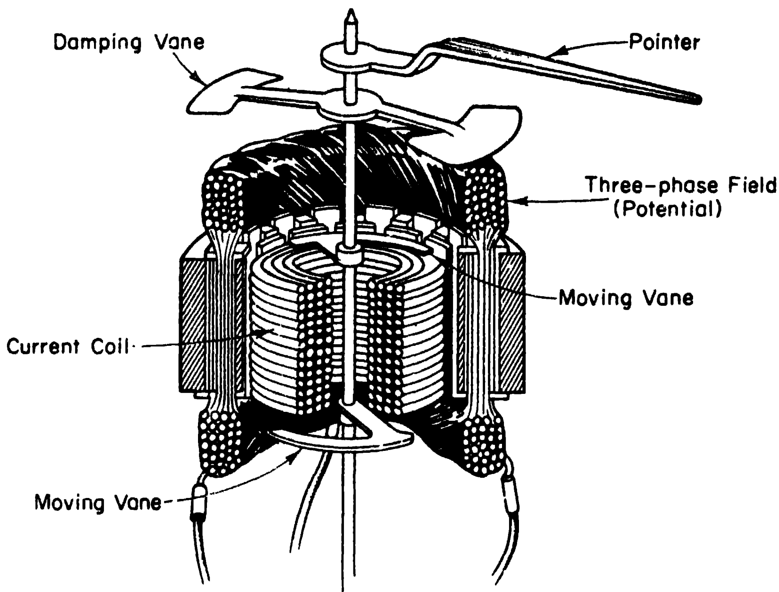


Figure 5-26
Diagram showing the structure of a polarized-vane power-factor meter. (*Courtesy General Electric Company Limited.*)

because its operating principle depends on the application of three-phase voltage. The outside coil in the photograph is the potential coil, which is connected to the three phase lines of the system. The application of three-phase voltage to the potential coil causes it to act like the stator of a three-phase induction motor in setting up a *rotating magnetic flux*. The central coil or current coil, is connected in series with one of the phase lines, and this *polarizes* the iron vanes. The polarized vanes move in a rotating magnetic field and take up the position that the rotating field has at the instant that the polarizing flux is maximum. This position is an indication of the phase angle and therefore the power-factor. The instrument may be used on single-phase systems, provided that a phase-splitting network (similar to that used in single-phase motors) is used to set up the required rotating magnetic field.

Both types of power-factor meters are limited to measurement at comparatively low frequencies and are typically used at the powerline frequency (60 Hz). Phase measurements at higher frequencies often are more accurately and elegantly performed using special electronic instruments or techniques. Methods and instruments for phase measurements at higher frequencies are discussed in Chapters 9, 11, and 13.

5-11 Frequency Meters

Frequency can be determined in a variety of ways, but at the moment we are concerned with indicating instruments; in this category, frequency meters use the effect of frequency upon such factors as mutual inductance, resonance of a tuned circuit, and mechanical resonance.

An example of the use of tuned circuits is found in the *electrodynamometer-type frequency meter*, shown schematically in Fig. 5-27. In this frequency meter,

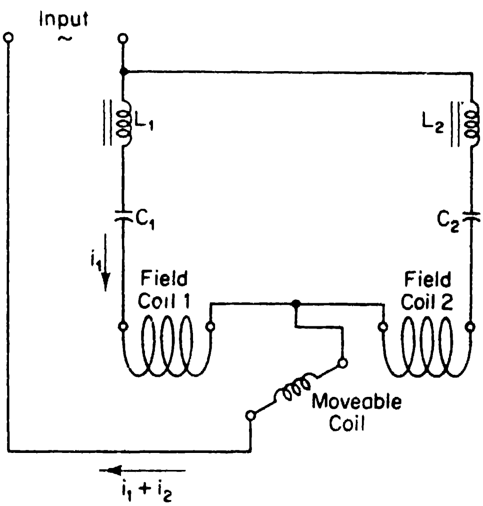


Figure 5-27
Circuit arrangement of the electrodynamometer-type frequency meter.

the field coils form part of two separate resonant circuits. Field coil 1 is in series with an inductor, L_1, and a capacitor, C_1, forming a resonant circuit which is tuned to a frequency slightly below the low end of the instrument scale. Field coil 2 is in series with inductor L_2 and capacitor C_2 forming a resonant circuit which is tuned to a frequency slightly higher than the high end of the instrument scale. In the case of powerline frequencies, the circuits would be tuned to frequencies of 50 Hz and 70 Hz, respectively, with 60 Hz in the middle of the scale. The two field coils are arranged as shown in the diagram and are returned to the powerline through the winding of the movable coil. The torque on the movable element is proportional to the current through the moving coil. This current consists of the sum of the two field-coil currents. For an applied frequency within the limits of the instrument range, the circuit of field coil 1 operates above the resonant frequency with current i_1 lagging the applied voltage. The circuit of field coil 2 operates below its resonant frequency and is therefore capacitive with current i_2 leading the applied voltage. The torques produced by the two currents on the movable coil are therefore in opposition, and the resulting torque is a function of the frequency of the applied voltage. For each given frequency within the range of the instrument, the resulting torque on the movable element causes the pointer to take up a given position and the pointer deflection is calibrated in terms of the given frequency. The restoring torque is provided by a small iron vane mounted on the moving coil. The range of operation of this instrument is usually limited to the powerline frequencies and it finds its major application in this field where it is used for monitoring the frequency of a power system.

The *tuned-reed frequency meter* operates on the principle of *mechanical* resonance. A series of reeds is fastened to a flexible common base which is mounted on the armature of an electromagnet. The coil of the electromagnet is energized from the ac powerline whose frequency is to be determined. The reeds are tuned to an exact natural frequency by careful selection of their length and mass. The reed which has a natural frequency equal to the frequency with which the electromagnet is energized, builds up a vibration. The vibration of the reed is visible at the front of the meter where the vibrating tip of the reed is shown through a window. If the frequency which is to be measured is intermediate between the natural frequencies of two adjacent reeds, both reeds will vibrate and the line frequency will be closest to the reed with the largest vibration. Interpolation between the natural frequencies of the reeds can be made quite easily and accurately, since the reed frequencies are exact. This instrument has the advantage of being very simple of construction and very rugged. It maintains its calibration well, provided that the vibration of the reeds is kept within reasonable limits. Although its operation does not depend on the exact value of the voltage, different voltage ranges are usually provided by the addition of a series resistor.

The *saturable-core frequency meter,* which can comfortably handle and measure a rather wide range of frequencies, is shown schematically in Fig 5-28. The transformer consists of *two cores* and a *yoke*. One core is of nonmagnetic material; the other core is of magnetic material which saturates at very low values of emf and current. The yoke is made of magnetic material but its cross section is so large that it does not reach saturation. The primary winding of the transformer

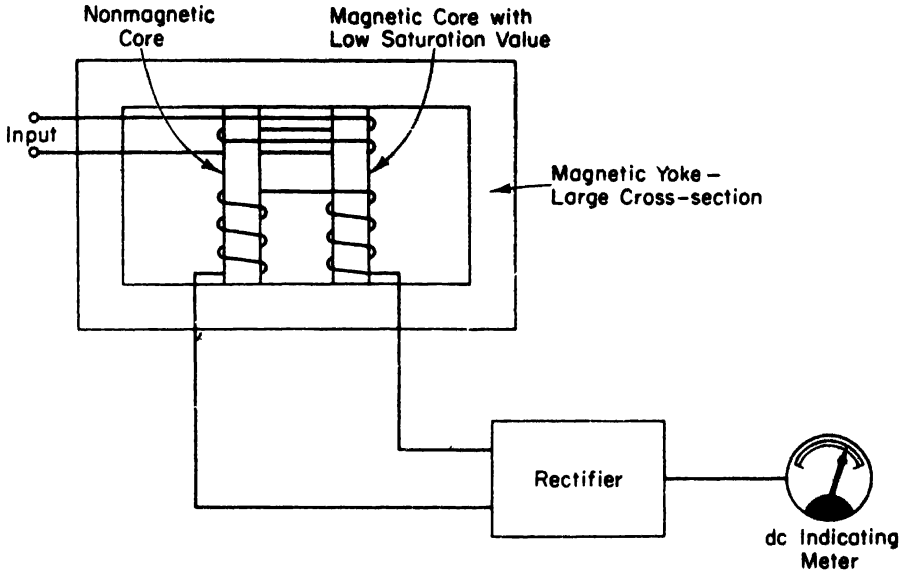


Figure 5-28
Schematic representation of the saturable-core frequency meter.

is wound around both cores simultaneously, as shown in Fig. 5-28. The secondary winding consists of two parts: one half of the winding is wound on the magnetic core and the other half of the winding on the nonmagnetic core. The secondary windings are connected in series in such a way that the voltages induced in the windings oppose each other. When power is supplied to the primary winding, transformer action induces secondary voltages in the secondary windings. Because of the low saturation value of the magnetic core, this core will saturate at very small secondary voltages. As soon as this core is saturated, the rate of increase of induced voltage in that winding will be equal to the rate of increase of the induced voltage in the winding on the nonmagnetic core. Therefore, the rate of increase of induced voltages cancels out, since the emf's in the secondary windings oppose each other. The secondary voltage will then not be a function of the primary applied voltage, but will depend only upon the frequency of the voltage. The secondary output voltage is rectified and applied to a dc meter, whose deflection is proportional to the frequency. The meter scale is calibrated in terms of frequency.

5-12 Instrument Transformers

Instrument transformers are used to measure ac at generating stations, transformer stations, and at transmission lines, in conjunction with ac measuring instruments (voltmeter, ammeters, wattmeters, VARmeters, etc.). Instrument transformers are generally classified according to their use and referred to as *current transformers* (CT) and *potential transformers* (PT).

Instrument transformers perform two important functions: they serve to *extend the range* of the ac measuring instrument, much as the shunt or the multiplier extends the range of a dc meter; they serve to *isolate* the measuring instrument from the high-voltage power line.

The *range* of a dc ammeter may be extended by using a shunt which divides the current under measurement between the meter and the shunt. This method is quite satisfactory for dc circuits but in ac circuits current division depends not only upon the resistances of the meter and the shunt but also upon their reactances. Since ac measurements are made over a wide frequency range, it becomes rather difficult to obtain great accuracy. A CT provides the range extension through its transformation ratio and in addition produces almost the same reading regardless of the meter constants (reactance and resistance), or in fact, of the number of instruments (within limits) connected in the circuit.

Isolation of the measuring instrument from the high-voltage power line is important when we consider that ac power systems frequently operate at voltages of several hundred kV. It would be impractical to bring the high-voltage lines directly to an instrument panel in order to measure voltage or current, not only because of the safety hazards involved but also because of

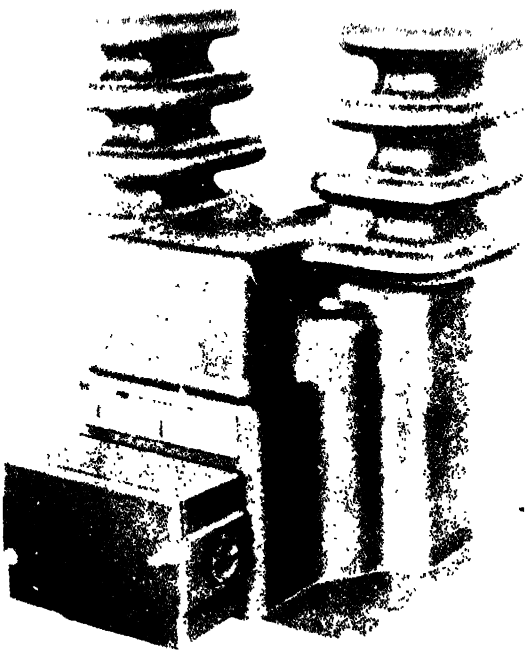


Figure 5-29 (a)
High voltage potential transformer. (*Courtesy Westinghouse Electric Corporation.*)

the insulation problems connected with high-voltage lines running closely together in a confined space. When an instrument transformer is used, only the low-voltage wires from the transformer secondary are brought to the instrument panel and only low voltages exist between these wires and ground, thereby minimizing safety hazards and insulation problems.

Many textbooks develop in detail the theory underlying the operation of transformers. Here these instrument transformers are merely described and their use in measurement situations is shown.*

Figure 5-29(a) presents a photograph of a *potential transformer*; Figure 5-29(b) shows a *current transformer*. The *potential transformer* (PT) is used to trans-

*For fuller treatment of ac machines and circuits consult text books like the following: Michael Liwshitz-Garik and Clyde C. Whipple, *AC Machines*, 2nd ed. (Princeton, N. J.: D. Van Nostrand Company, Inc., 1961), Chaps. 2–5. Russell M. Kerchner and George F. Corcoran, *Alternating Current Circuits*, 4th ed. (New York: John Wiley & Sons, Inc., 1961), pp. 291–317.

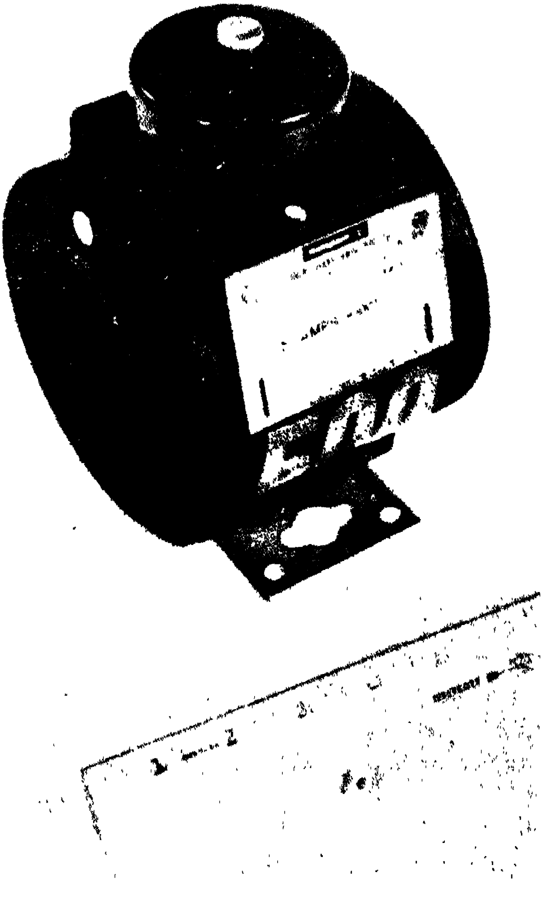


Figure 5-29 (b)
Current transformer. (*Courtesy Westinghouse Electric Corporation.*)

form the high voltage of a powerline to a lower value suitable for direct connection to an ac voltmeter or the potential coil of an ac wattmeter. The usual secondary transformer voltage is 120 V. Primary voltages are standardized to accommodate the usual transmission line voltages which include 2,400 V, 4,160 V, 7,200 V, 13.8 kV, 44 kV, 66 kV, and 220 kV. The PT is rated to deliver a certain power to the secondary load or *burden*. Different load capacities are available to suit individual applications; a general capacity is 200 VA at a frequency of 60 Hz.

The PT must satisfy certain design requirements which include accuracy of the turns-ratio, small leakage reactance, small magnetizing current, and minimal voltage drop. Furthermore, since we may be working with very high primary voltages, the insulation between the primary and the secondary windings must be able to withstand large potential differences, and the dielectric requirements are very high. In the usual case, the high-voltage coil is of a circular pancake construction, shielded to avoid localized dielectric stresses. The low-voltage coil or coils are wound on a paper form and assembled inside the high-voltage coil. The assembly is thoroughly dried and oil impregnated. The core and coil assembly is then mounted inside a steel case, which supports the high-voltage terminals or bushings, made of porcelain. The case is then filled with an insulating oil.

Recent developments in the synthetic rubber industry have introduced the molded rubber potential transformer, replacing the insulating oil and porcelain bushings in some applications. Figure 5-29(a) shows a rubber-molded 25-kV potential transformer, suitable for outdoor use. This unit is less expensive than the conventional oil-filled PT, and since the bushings are made of molded rubber, porcelain breakage is eliminated. A white polarity dot is placed on the proper bushing on the front of the transformer. Two stud-type secondary terminals are enclosed in a removable conduit box. The power rating of a potential transformer is based on considerations other than load capacity, for the reasons previously outlined. A typical load rating is 200 VA at 60 Hz for a transformer having a ratio of 2,400/120 V. For most metering purposes however, the burden will be significantly less than 200 VA.

The *current transformer* (CT) sometimes has a primary and always a secondary winding. If there is a primary winding it has a small number of turns. In most cases, the primary is only one turn or a single conductor connected in series with the load whose current is to be measured. The secondary winding has a larger number of turns and is connected to a current meter or a relay coil. Often the primary winding is a single conductor in the form of a heavy copper or brass bar running through the core of the transformer. Such a CT is called a *bar-type* current transformer. The CT secondary winding is usually designed to deliver a secondary current of 5 A. An 800/5-A bar-type current transformer would have 160 turns on the secondary coil.

Since the primary is connected directly in the load circuit, when the secondary winding of a current transformer is open-circuited, the voltage developed across the open terminals may be very high (because of the step-up ratio) and could easily break down the insulation between the secondary windings. Therefore, the secondary windings of a current transformer are always short-circuited or a meter or relay coil is connected to the secondary. A current transformer should *never* have its secondary open while the primary is carrying current, but should *always* be closed through a current meter, relay coil, wattmeter current coil, or simply a short. Failure to observe this precaution may cause serious damage to either equipment or operating personnel.

The current transformer, shown in Fig. 5-29(b), consists of a core with the secondary winding encased in molded rubber insulation. The window in the core allows for the insertion of one or more turns of the current-carrying high-voltage conductor. A single conductor constitutes a one-turn primary winding. The nominal ratio of the transformer is given on the name plate; this is not the turns ratio (since more than one turn can be used as the primary) but only indicates that a primary current of 500 A will cause a secondary current of 5 A, when the secondary coil is connected to a 5-A ammeter. Within practical limits, the current in the secondary winding is determined by the primary excitation current and not by the secondary circuit impedance. Since the primary cur-

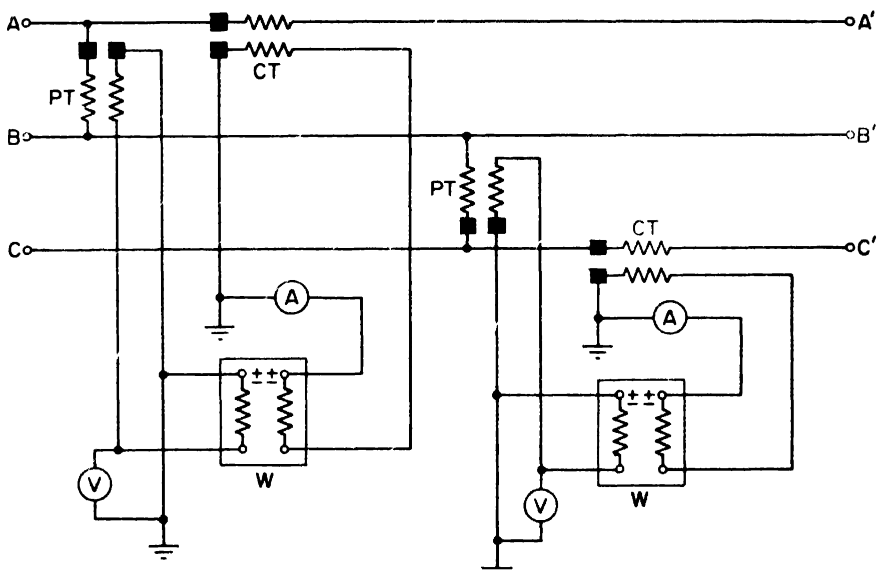


Figure 5-30
Instrument transformers in a three-phase measurement application. Polarity markings of the potential and current transformers are indicated by black squares.

rent is determined by the load in the ac system, the secondary current is related to the primary current by approximately the inverse of the turns ratio. This is true within rather wide limits of the nature of the secondary burden.

Figure 5-30 indicates the use of instrument transformers in a typical measurement application. This diagram illustrates the connection of the instrument transformers in a three-wire, three-phase circuit, including two wattmeters, two voltmeters, and two ammeters. The potential transformers are connected across phase lines A and B, and phase lines C and B; the current transformers are in the phase lines A and D. The secondary windings of the potential transformers are connected to the voltmeter coils and potential coils of the wattmeters; the current transformer secondaries feed the ammeters and the current coils of the wattmeters.

The polarity markings on the transformers, indicated by a dot at the transformer leads, aid in making the correct polarity connections to the measuring instruments. At any given instant of the ac cycle, the dot-marked terminals have the *same* polarity and the marked wattmeter terminals must be connected to these transformer leads as shown.

5. (a) What is the principal advantage of the electrostatic voltmeter? (b) Explain why this instrument has a "square law" scale. (c) Can this instrument be used as a transfer instrument? Why?

6. Outline a calibration procedure for an electrodynamometer-type ac voltmeter. State what laboratory equipment is required for this calibration and indicate the expected accuracy.

7. The circuit diagram of Fig. 5-5 shows a full-wave rectifier ac voltmeter. The meter movement has an internal resistance of 250 Ω and requires 1 mA for full-scale deflection. The diodes each have a forward resistance of 50 Ω and infinite reverse resistance. Calculate (a) the series resistance R_s required for full-scale meter deflection when 25 V rms is applied to the meter terminals, (b) the ohms-per-volt rating of this ac voltmeter.

8. Calculate the indication of the meter of Question 7 when a triangular waveform with a peak value of 20 V is applied to the meter terminals.

9. A resistance of 250 Ω is placed in parallel with the meter movement of the instrument of Question 7. (a) What is the function of this resistor? (b) What effect does this resistance have on the ohms-per-volt rating of the voltmeter? (c) Calculate the new value of resistor R_s to give full-scale deflection for an input voltage of 25 V rms.

10. The commercial voltmeter of Fig. 5-7 uses a 1-mA meter movement with an internal resistance of 100 Ω. The shunting resistance across the movement is 200 Ω. Diodes D_1 and D_2 each have a forward resistance of 200 Ω and infinite reverse resistance. (a) Explain the function of the shunting resistor across the meter movement. (b) Explain the function of diode D_2. (c) Calculate the values of series resistors R_1, R_2, and R_3 if the required meter ranges are 10 V, 50 V, and 100 V, respectively. (d) Determine the ohms-per-volt rating of this ac voltmeter.

11. A thermocouple instrument reads 10 A at full-scale deflection. Calculate the current which causes half-scale deflection.

12. Prove that three wattmeters correctly measure the total power in a three-phase, four-wire system. Assume that the load is star-connected, balanced, and purely resistive. Draw a complete phasor diagram of all line and phase voltages and currents.

13. How many wattmeters are needed to measure the total power in a three-phase three-wire circuit when the load consists of a Y-connected induction motor? Assume that it is necessary to use current and potential transformers and draw a complete circuit diagram of the measurement setup.

14. What is the significance of the marking dots on a current or potential transformer?

SIX

PRINCIPLES AND APPLICATIONS OF POTENTIOMETERS

6-1 Introduction

A *potentiometer* is an instrument designed to measure an unknown voltage by *comparing* it with a known voltage. The known voltage may be supplied by a *standard cell* or any other known voltage-reference source. Measurements using the *comparison method* are capable of a very high degree of accuracy because the result obtained does not depend on an actual pointer deflection, as is the case with a moving-coil instrument, but *only* upon the accuracy of the known voltage standard to which the comparison is made.

Since the potentiometer makes use of a *balance* or *null* condition, no power is consumed from the circuit containing the unknown emf when the instrument is balanced; as a result, the voltage determination is quite *independent* of the source resistance. Since the potentiometer measures voltage, it can also be used to determine current simply by measuring the voltage drop produced by the unknown current through a known *resistance standard.*

The potentiometer is used extensively for the calibration of voltmeters and ammeters and has in fact become the *standard* for the calibration of these instruments. For this reason the potentiometer is very important in the field of electrical measurement and calibration.

6-2 The Basic Potentiometer Circuit

The principle of operation of all potentiometers is based on the circuit of Fig. 6-1, which shows the schematic of the basic *slide-wire potentiometer*. We

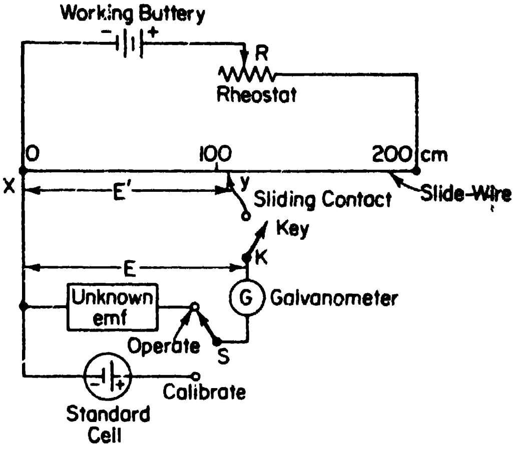


Figure 6-1
Circuit diagram of the basic slide-wire potentiometer.

shall study the operation of this basic circuit on a qualitative basis and then proceed to more sophisticated potentiometric instruments.

With switch S in the "operate" position and the galvanometer key K open, the battery supplies *working current* through the rheostat and the *slide wire*. The working current through the slide wire may be varied by changing the rheostat setting. The method of measuring the unknown voltage, E, depends on finding a position for the sliding contact such that the galvanometer shows *zero deflection* (*a null*) when the galvanometer key, K, is closed. Zero galvanometer current or a null means that the unknown voltage, E, is equal to the voltage drop E' across portion xy of the slide wire. Determination of the value of the unknown voltage now becomes a matter of evaluating the voltage drop E' along the slide wire.

The slide wire is carefully manufactured and has uniform resistance along its entire length. A calibrated *scale*, usually in centimeters and fractions of centimeters, is placed along the slide wire so that the sliding contact can be placed accurately at any desired position along the slide wire. Since the resistance of the slide wire is known accurately, the voltage drop along the entire slide wire, or along any portion of it, can be controlled by adjusting the working current. The working current is adjusted or "standardized" by reference to known voltage-reference source.

The procedure for *standardizing* the potentiometer is illustrated by the following example:

The slide wire of Fig. 6-1 has a total length of 200 cm and a resistance of 200 Ω. The emf of the voltage reference, indicated by the standard cell in Fig. 6-1, is given as 1.019 V. Switch *S* is thrown to the "calibrate" position and the sliding contact is placed at the 101.9-cm mark on the slide-wire scale. The rheostat is now adjusted to provide a working current such that the galvanometer shows no deflection when the key, *K*, is depressed. In this condition of balance or null, the voltage drop along the 101.9-cm portion of the slide wire is equal to the standard cell voltage of 1.019 V. Since the 101.9-cm portion of the slide wire represents a resistance of 101.9 Ω, the working current has in fact been adjusted to 10 mA. The voltage at any point along the slide wire is proportional to the length of the slide wire and is obtained by converting the calibrated length into the corresponding voltage, simply by placing the deci-

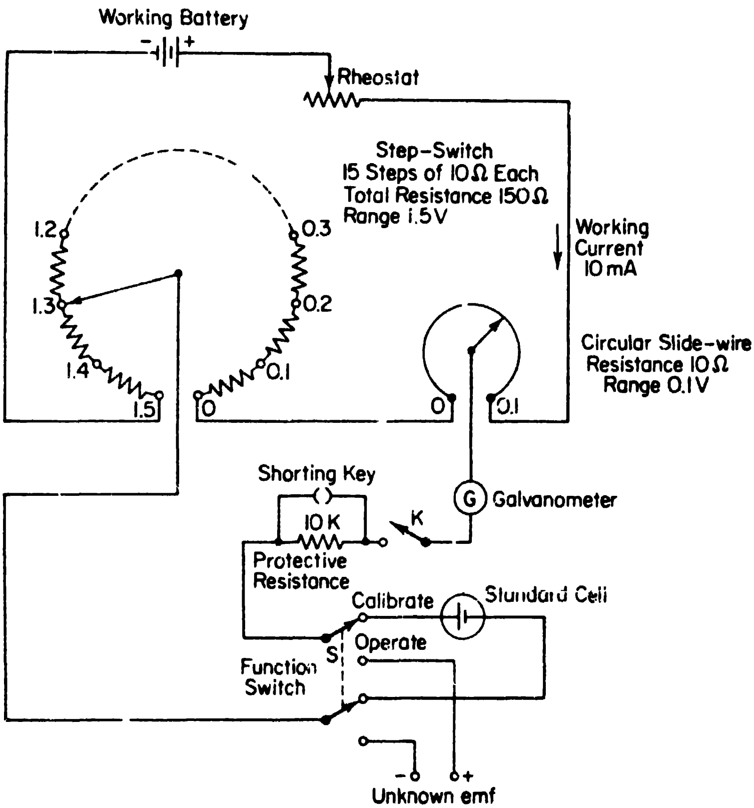


Figure 6-2
Circuit diagram of a simple potentiometer illustrating the use of dial resistors and a circular slide-wire.

mal point in the proper position (e.g., 146.3 cm = 1.463 V). Once calibrated, the working current is never varied.

After the potentiometer has been standardized, any unknown dc voltage may be measured. Switch S is thrown to the "operate" position, and the sliding contact is moved along the wire until the galvanometer shows no deflection when key K is closed. At this null condition, the slide-wire scale reading is converted into its corresponding voltage value. Refer to the voltage standard at regular intervals to make sure that the working current is maintained at its correct value.

The slide-wire potentiometer is a rather impractical form of construction. Modern laboratory-type potentiometers use *calibrated* dial *resistors* and a *small circular slide wire* of one or more turns, thereby reducing the size of the instrument. Figure 6-2 shows the schematic diagram of a simple potentiometer where the long slide wire has been replaced by a combination of fifteen precision resistors and a single-turn circular slide wire. In this case, the resistance of the slide wire is 10 Ω and the dial resistors have a value of 10 Ω each for a total resistance of 150 Ω in the step switch. The slide wire is provided with two hundred scale divisions, and interpolation to one-fifth of a division can be estimated quite comfortably. The working current of this potentiometer is maintained at 10 mA, so that each step of the dial switch corresponds to a voltage step of 0.1 V. Each division on the slide-wire scale corresponds to 0.0005 V and readings can be estimated to approximately 0.0001 V.

The potentiometer is also provided with a double-throw switch allowing the connection of either the standard cell or the unknown dc emf. The galvanometer circuit includes a key and a protective series resistance. To operate the galvanometer at its maximum sensitivity the resistance can be shorted out by inserting a *shorting key* in the contact across the resistance.

Examples 6-1 and 6-2 illustrate some of the characteristics of single-range potentiometers.

Example 6-1: The basic slide-wire potentiometer of Fig. 6-1 has a working battery of 3.0 V with negligible internal resistance. The resistance of the slide wire is 400 Ω and its length is 200 cm. A 200-cm scale, placed along the slide wire, has 1-mm scale divisions and interpolation can be made to one-quarter of a division. The instrument is standardized against a voltage reference source of 1.0180 V with the slider set at the 101.8 cm mark on the scale. Calculate (a) the working current; (b) the resistance of the rheostat; (c) the measurement range; (d) the resolution of the instrument, expressed in mV.

SOLUTION:

(a) When the instrument is standardized, the 101.8 cm mark on the scale corresponds to 1.0180 V (E′ in Fig. 6-1). 101.8 cm of the slide wire represents a resistance of 101.8/200 × 400 Ω = 203.6 Ω. The working current therefore must be 101.8 V/203.6 Ω = 0.5 mA

(b) With a working current of 0.5 mA the voltage drop along the entire slide wire is 0.5 mA × 400 Ω = 2.0 V. The voltage drop across the rheostat then equals 3.0 − 2.0 = 1.0 V and the rheostat setting is 1.0 V/ 0.5 mA = 2,000 Ω

(c) The measurement range is determined by the total voltage across the slide wire which equals 0.5 mA × 400 Ω = 2.0 V.

(d) The resolution of the potentiometer is determined by the voltage represented by one-quarter of a scale division, or 0.25 mm. Since the total length of 200 cm corresponds to 2.0 V, the resolution is 0.25 mm/ 200 cm × 2.0 V = 0.25 mV.

Example 6-2: The single-range potentiometer of Fig. 6-2 is equipped with a 20-step dial switch where each step represents 0.1 V. The dial resistors are 10 Ω each. The 11-turn slide wire has a resistance of 11 Ω, allowing some overlap between settings of the dial switch. The circular slide-wire scale has 100 divisions and interpolation can be made to one-fifth of a division. The working battery has a terminal voltage of 6.0 V and negligible internal resistance. Calculate (a) the measuring range of the potentiometer; (b) the resolution, in μV; (c) the working current; (d) the setting of the rheostat.

SOLUTION:

(a) The total resistance, R_m, of the measuring circuit is

$$R_m = R_{\text{dial}} + R_{\text{slide wire}} = (20 \times 10\,\Omega) + 11\,\Omega = 211\,\Omega$$

Since each 10-Ω step of the dial switch represents a voltage of 0.1 V, the total measuring range is $\frac{211}{10} \times 0.1$ V = 2.11 V.

(b) The 11-Ω slide wire represents a voltage of 0.11 V. Each turn of the slide wire therefore represents 0.11 V/11 = 0.01 V, or 10 mV. Each division on the slide-wire scale represents $\frac{1}{100} \times 10$ mV = 0.1 mV, or 100 μV. The resolution of the instrument therefore is $\frac{1}{5} \times 100\,\mu$V = 20 μV.

(c) To maintain a voltage of 0.1 V across each 10-Ω dial resistor the working current must be 0.1 V/10 Ω = 10 mA.

(d) Since the voltage across the entire measuring resistance is 2.11 V, the drop across the rheostat must be 6.0 V − 2.11 V = 3.89 V. The rheostat setting then is 3.89 V/10 mA = 389 Ω.

The following steps are required in making a potentiometric measurement:

(a) The combination of dial resistors and slide wire is set to the value of the standard cell voltage. (This value is usually printed on the body of the cell.)

(b) The switch is thrown to the "calibrate" position and the galvanometer key is tapped while the rheostat is adjusted for zero deflection on the galvanometer. The protective resistance is left in the circuit to avoid damage to the galvanometer during the initial stages of adjustment.

(c) As zero deflection is approached, the protective resistance is shorted and final adjustments are made with the rheostat control.

(d) After the standardization has been completed, the switch is thrown to the "operate" position, thus connecting the unknown emf into the circuit. The instrument is balanced by the main dial and the slide wire, leaving the protective resistance again in the circuit.

(e) As balance is approached, the protective resistance is shorted, and final adjustments are made to obtain a true valance condition.

(f) The value of the unknown voltage is read directly off the dial settings.

(g) The working current is checked by returning to the "calibrate" position. If the dial settings are exactly the same as in the original calibration procedure, a valid measurement has been made. If the reading does not agree, a second measurement must be made, again returning to a calibration check after completion.

6-3 Multiple-Range Potentiometers

The single-range potentiometer of Sec. 6-2 is usually constructed to cover voltage ranges of up to 1.6 V. The circuit may be modified to include a second measuring range of lower value by the addition of two range resistors and a range switch. Figure 6-3 shows the schematic diagram of a *duo-range potentiometer,* where R_1 and R_2 are the range resistors and switch S is the range switch. The operation of the duo-range potentiometer of Fig. 6-3 may more easily be understood and analyzed by redrawing it in simplified form, omitting some of the details of the galvanometer and calibration circuitry. The simplified schematic is shown in Fig. 6-3.

In Fig. 6-4 the total measuring resistance R_m consists of the slide wire in series with the main dial. The main dial has fifteen steps of 10 Ω each for a total resistance of 150 Ω, and the resistance of the slide wire is 10 Ω. To pro-

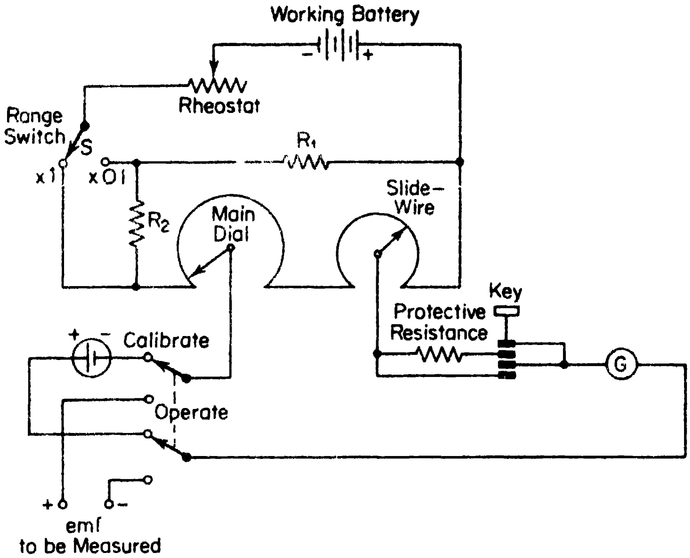


Figure 6-3
Schematic diagram of a duo-range potentiometer.

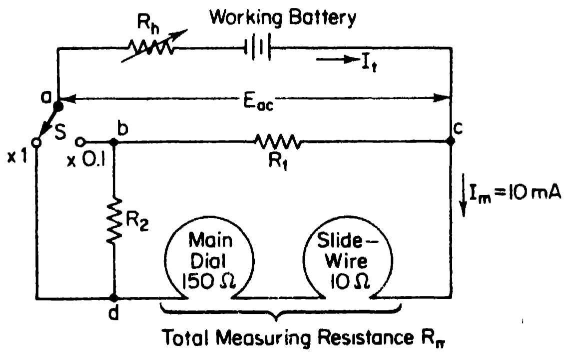


Figure 6-4
Simplified schematic diagram of the duo-range potentiometer.

duce a voltage drop of 1.6 V across the main dial and slide wire, the measuring current I_m must be 10 mA. With switch S in the position shown, the instrument has identical components and current as the single-range potentiometer of Sec. 6-2. When the range switch S is thrown to the $\times 0.1$ position, the measuring current I_m must be reduced to one-tenth of its original value, or 1 mA, to produce a voltage drop of 0.16 V across the measuring resistance R_m.

It is essential in the design of the circuit to be able to change measuring ranges without readjusting the rheostat or changing the value of the working battery voltage. Once the instrument has been calibrated on the $\times 1$ range, following the standardization procedure of Sec. 6-2, calibration of the $\times 0.1$

range should not be necessary. This requires that the voltage E_{ac} in Fig. 6-4 remain the same for both positions of the range switch S. This condition is satisfied only when the total battery current has the same value for each measuring range.

For an analysis of the operation of the duo-range potentiometer of Fig. 6-4, see Fig. 6-5, which shows the elementary circuits of the $\times 1$ range and the

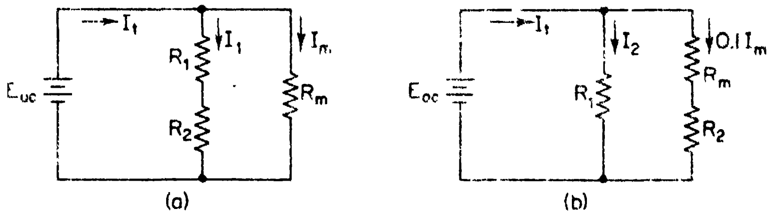


Figure 6-5
Elementary circuits of the duo-range potentiometer, showing the circuit (a) on the $\times 1$ range and (b) on the $\times 0.1$ range.

$\times 0.1$ range. On the $\times 1$ range, Fig. 6-5(a), the range resistors R_1 and R_2 are in parallel with the total measuring resistance R_m. On the $\times 0.1$ range, Fig. 6-5(b), range resistor R_1 is in parallel with the series combination of R_2 and R_m. A constant battery current is possible only when the total circuit resistance R_t on each range is the same. Equating the resistance of Fig. 6-5(a) and Fig. 6-5(b), we obtain

$$\frac{R_m(R_1 + R_2)}{R_1 + R_2 + R_m} = \frac{R_1(R_2 + R_m)}{R_1 + R_2 + R_m} \tag{6-1}$$

and simplifying,

$$R_2R_m = R_1R_2 \quad \text{or} \quad R_1 = R_m \tag{6-2}$$

Equation (6-2) indicates that range resistor R_1 must have the same resistance as the measuring resistance R_m for the battery to supply constant current on either range.

Voltage E_{ac} must be the same for either position of the range switch in order to change ranges without upsetting the initial calibration. E_{ac} can be evaluated by referring to Fig. 6-5. With the range switch in the $\times 1$ position [Fig. 6-5(a)],

$$E_{ac} = I_mR_m \tag{6-3}$$

With the range switch in the $\times 0.1$ position [Fig. 6-5(b)],

$$E_{ac} = I_2R_1 \tag{6-4}$$

Combining Eq. (6-3) and (6-4), we obtain

$$E_{ac} = I_mR_m = I_2R_1 \tag{6-5}$$

Substituting Eq. (6-2) into Eq. (6-5), we obtain

$$I_m = I_2 \tag{6-6}$$

Equation (6-6) indicates that the shunt current, I_2, on the $\times 0.1$ range must be equal to the measuring current, I_m, on the $\times 1$ range.

The battery current I_{t1} in Fig. 6-5(a) is

$$I_{t1} = I_1 + I_m \tag{6-7}$$

The battery current I_{t2} in Fig. 6-5(b) is

$$I_{t2} = I_2 + 0.1 I_m \tag{6-8}$$

Combining Eq. (6-7) and (6-8) and using Eq. (6-6), we obtain

$$I_1 + I_m = I_2 + 0.1 I_m \quad \text{or}$$

$$I_1 = 0.1 I_m \tag{6-9}$$

Finally, to establish the resistance of R_2, the only unknown left in the potentiometer circuit, consider again Fig. 6-5(a). The voltage drop across R_m must be equal to the voltage drop across the series combination of R_1 and R_2; therefore,

$$I_1(R_1 + R_2) = I_m R_m \tag{6-10}$$

Substituting Eq. (6-2) and (6-9) into Eq. (6-10), we obtain

$$0.1 I_m(R_1 + R_2) = I_m R_1 \quad \text{or}$$

$$R_2 = 9R_1 \tag{6-11}$$

For the circuit of Fig. 6-4, where the measuring resistance $R_m = 160\ \Omega$, we find that $R_1 = R_m = 160\ \Omega$ and $R_2 = 9\,R_1 = 9 \times 160 = 1{,}440\ \Omega$. Since we had already assumed a measuring current of 10 mA on the $\times 1$ range, shunt current $I_1 = 0.1 \times 10$ mA $= 10$ mA, and the total battery current $I_t = 11$ mA. On the $\times 0.1$ range, the measuring current is $0.1 I_m = 1$ mA and the shunt current $I_2 = I_m = 10$ mA, again giving a total working current of 11 mA. The condition for constant working current on both ranges has therefore been met.

Calibration of the duo-range potentiometer is accomplished in the usual manner on the $\times 1$ range position. Range resistors R_1 and R_2 are both precision resistors and the initial calibration should be valid for the lower range. The potentiometer of Figs. 6-3 and 6-4 can be used to measure voltages up to 0.16 V on the lower range. The dial readings are simply multiplied by the range factor of 0.1. If the slide wire has one hundred scale divisions which can be interpolated to one-fifth of a division, the resolution of the potentiometer reading is $1/5 \times 1/100 \times 0.01\ \text{V} = 20\ \mu\text{V}$, on the $\times 0.1$ range.

The duo-range potentiometer of Fig. 6-3 was constructed for a voltage ratio of 10/1. A similar arrangement may be used for any other ratio by proper selection of the range resistors R_1 and R_2.

Precision laboratory-type potentiometers often have three voltage ranges, usually of 1.0, 0.1, and 0.01. A typical *multiple-range potentiometer* is shown in Fig. 6-6. Inspection of the circuit indicates that the multiple-range poten-

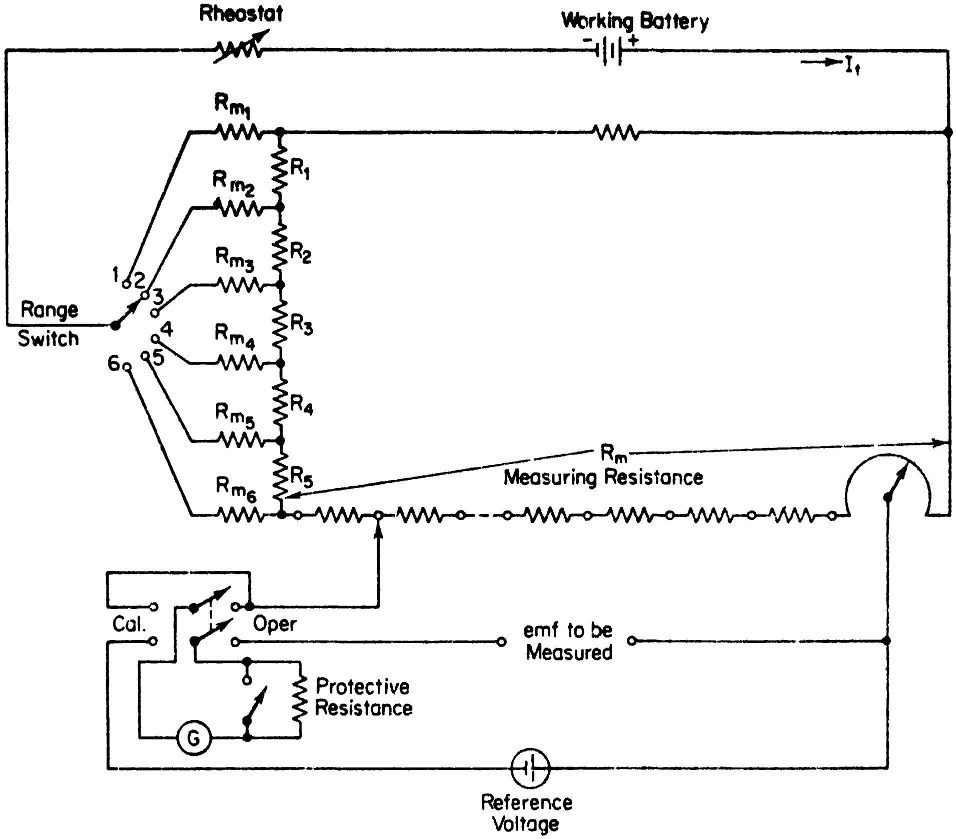


Figure 6-6
A multiple-range potentiometer.

tiometer consists of a basic measuring resistance, R_m, and an associated circuit which will multiply the values read on the basic scale of the instrument by fixed multiplying factors. These multiplying factors, marked as 1 to 6 on the range switch in Fig. 6-6, are the factors by which the voltage along the measuring resistance is multiplied. The relation of these multiplying factors is such that position 1 gives the greatest voltage drop across R_m and each succeeding position gives an increasingly smaller voltage drop across R_m, until at position 6, the smallest drop across the measuring resistance is reached. The ratios of the multiplying factors are proportional to the current through the measuring resistance, when these factors are used. The values of all the resistors in the range-switching circuit can be computed in terms of the measuring resistance and the required ranges of the potentiometer by basic circuit analysis and some rather tedious algebraic manipulation.

In designing a multiple-range potentiometer circuit, the current supplied by the battery should again remain constant on all ranges to allow the instrument to remain calibrated in all range positions. Different arrangements may be used for the range switching circuitry but the circuit of Fig. 6-6 conveys the general idea of the problems involved in designing a multiple-range potentiometer.

Circuits of the type briefly described in this section are used in some laboratory precision potentiometers and also in *self-balancing* and *recording* instruments, where a wide range of emf's is to be measured.

Figure 6-7 shows a *portable* potentiometer designed primarily for calibrating thermocouple-operated instruments and measuring thermocouple voltages. The simplified circuit diagram of this instrument is given in Fig. 6-8. (Try

Figure 6-7
A modern portable instrument designed for calibrating thermocouple-operated instruments and measuring thermocouple voltages over the -1 to $+15$ mV and 0 to 80 mV ranges. (*Courtesy Honeywell Test Instruments Division, Denver, Colo.*)

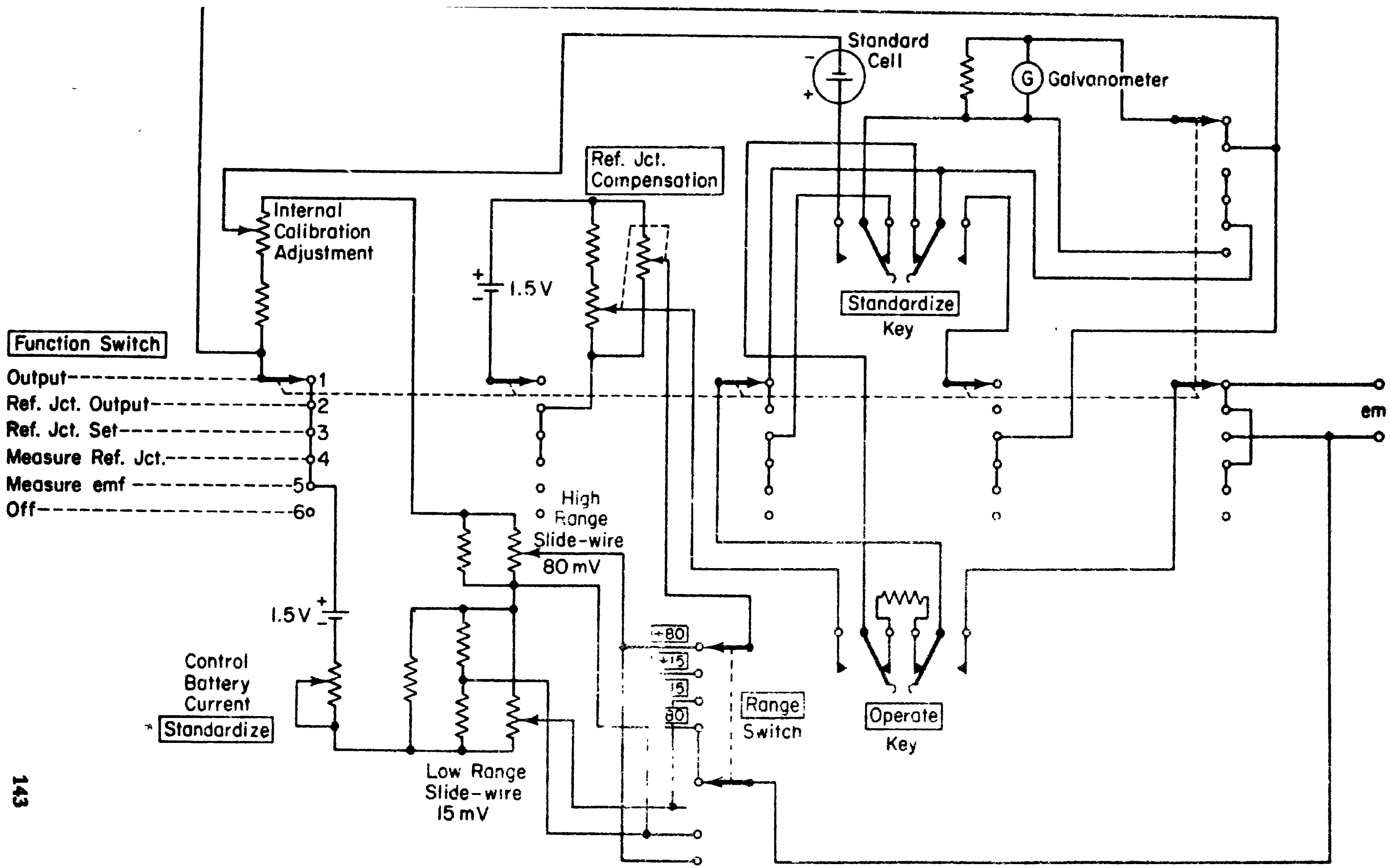


Figure 6-8. Simplified schematic diagram of the portable potentiometer of Fig. 6-7. (*Courtesy Honeywell Test Instruments Division, Denver, Colo.*)

to relate the various controls and switches shown in the photograph to the actual circuit diagram.)

6-4 The Voltbox

General purpose potentiometers generally cover a maximum voltage range of approximately zero to 1.6 V dc. If higher values of voltage are to be measured a precision voltage divider, called a *voltbox,* is used. The voltbox provides multiple ranges in the manner indicated in Fig. 6-9. Selection of the appropriate voltage range is usually made by choosing the corresponding binding posts on the voltbox or, in some instances, by means of a selector switch.

In Sec. 6-1 it was stated that the potentiometer consumes no power from the source when it is balanced. When a voltbox is used, however, this statement is no longer true. Inspection of the circuit of Fig. 6-9 shows that the

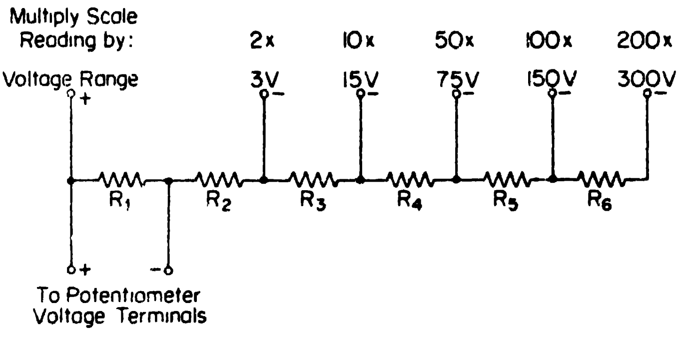


Figure 6-9
Circuit diagram of a voltbox.

"source" of emf connected to the potentiometer for measurement is now represented by resistor R_1 which is a portion of the total divider resistance. Current will be drawn from the *actual* source of emf in order to develop the voltage across R_1. This current drain can be made quite small by using a high-resistance divider. Choice of resistance values for the voltbox involves a compromise: high values are desirable to reduce the current drain on the measured source, yet low values are preferred because these generally give more stability. In addition, low-value resistors give better galvanometer sensitivity and also minimize the effects of high-resistance leakage paths around the binding posts.

Voltboxes are usually constructed with resistances of 100 Ω/V or 200 Ω/V of the nominal voltage range, allowing a maximum current drain from the measured source of 10 mA or 5 mA, respectively.

6-5 The Voltage Reference Source

Every potentiometer requires a voltage reference source for its calibration. In the past, a *standard cell* has been used exclusively for this purpose but in recent years semiconductor devices, such as *Zener-controlled reference sources,* have replaced the standard cell in practically all industrial applications. The standard cell, whose certified terminal voltage is traceable to NBS standards, is discussed in some detail in Chapter 3. The silicon-diode reference standard provides an alternative source of reference voltage with an acceptable degree of accuracy. It is used in many industrial applications where standard cell accuracy is not essential.

Silicon diodes, when properly processed, have voltage-current characteristics as shown in Fig. 6-10. The extremely sharp increase in reverse current occurs at a point on the curve known as the *Zener voltage*. This point indicates a *breakdown* of the diode under reverse voltage application, but the process is reversible if safe current and heating limits are not exceeded. The *Zener voltage* may be controlled over a wide range by the processing techniques used during the manufacture of Zener diodes. The slope of the reverse current at the Zener point is very steep, indicating that the diode has a very low dynamic resistance on this part of the characteristic curve.

The circuit of Fig. 6-11 shows the Zener diode in a *typical* (simplified) *application* as a reference-source. The supply voltage E is usually much higher than the desired reference voltage (the Zener voltage of the diode) assuring breakdown of the diode. Resistor R usually has a high value and is placed in series with the diode and the battery. It serves to limit the current through the diode during breakdown to a safe value.

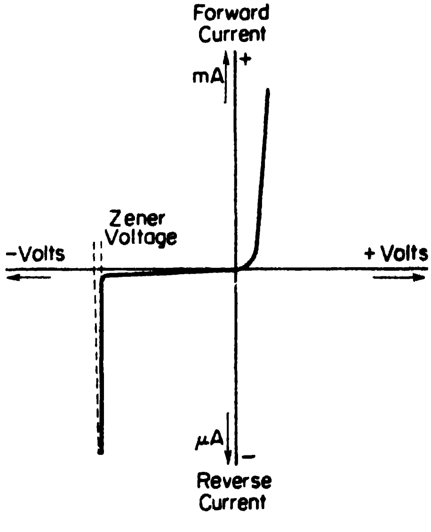


Figure 6-10
Volt-ampere characteristics of a Zener diode.

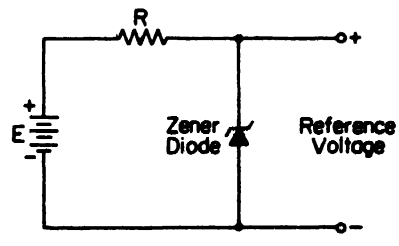


Figure 6-11
Practical circuit showing the Zener diode as a voltage reference.

In addition, when the supply voltage varies for any reason, most of this variation is taken up in voltage drop across R with a very small change in diode current and as a result the reference voltage stays practically the same. Since the dynamic resistance of the Zener diode is very small under reverse current conditions, the regulation of the circuit is excellent. Although the Zener voltage changes with ambient temperature, it is possible to select suitable diodes with temperature coefficients very nearly zero.

It is not expected that Zener-diode reference sources will replace the standard cell in high-precision work, but they may be used satisfactorily to maintain a calibrated voltage applied to a potentiometer circuit, for example, in thermocouple measurements or with recording potentiometers.

Commercial voltage-reference standards using Zener diodes plus additional circuitry to improve regulation and stability are available with voltage outputs in the neighborhood of 1.0190 V $\pm$ 0.01 per cent, which allows them to be substituted for a standard cell. An important advantage of these standards over the standard cell is that, unlike the standard cell, the semiconductor device is immune to the effects of short-circuit and overload conditions. The final choice of voltage reference depends on the application in which the reference is used, the required accuracy, and the cost of the standard.

6-6 Calibration of Voltmeters and Ammeters

The potentiometer method is the usual basis for the calibration of all voltmeters, ammeters, and wattmeters. Since the potentiometer is a dc measurement device, the instruments to be calibrated must be of the dc or electrodynamometer type. The circuit of Fig. 6-12 shows the measurement setup for the calibration of a dc voltmeter. One of the first requirements in this calibration procedure is that a suitable, *stable dc supply* be available, since any variation in the supply voltage causes a corresponding change in the voltmeter calibration voltage.

Figure 6-12 shows that a potential divider network, consisting of two rheo-

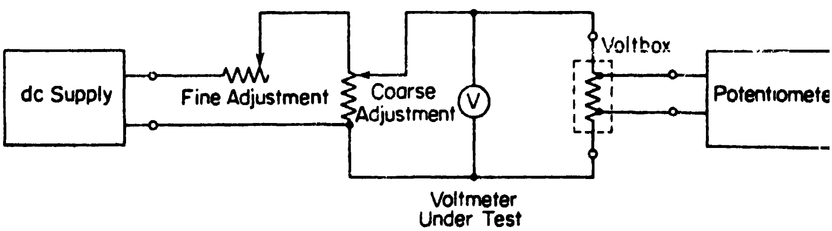


Figure 6-12
Calibration of a dc voltmeter by the potentiometer method.

stats for coarse and fine control of the calibrating voltage, is connected across the supply source. The voltage across the voltmeter is stepped down to a value suitable for application to a potentiometer by means of a voltbox. The voltage applied to the voltmeter is adjusted by the rheostats until the pointer rests on a major scale division. The potentiometer is used to determine the *true* value of this voltage. If the potentiometer reading does not agree with the voltmeter indication, positive or negative voltmeter error is indicated (see Table 6-1). A selected number of major scale divisions is checked in this way.

After the readings have been taken at the selected scale points, first with increasing and then with decreasing applied voltage, a *calibration curve* is plotted. The calibration curve is a plot of the correction values as the ordinate against the scale readings as the abscissa. The *correction value* for each point checked along the scale is the true value as measured by the potentiometer minus the scale reading of the voltmeter under calibration. The observed

TABLE 6-1

RESULTS OF A CALIBRATION OF A DC VOLTMETER BY THE POTENTIOMETER METHOD

dc Voltmeter Instrument Scale Reading (volts)	Potentiometer True Reading (volts)	Correction (volts)
0.0	0.00	0.00
1.0	0.95	−0.05
2.0	2.00	0.00
3.0	3.05	+0.05
4.0	4.10	+0.10
5.0	5.10	+0.10
6.0	6.15	+0.15
7.0	7.10	+0.10
8.0	8.15	+0.15
9.0	9.20	+0.20
10.0	10.25	+0.25

points are then joined by straight lines, since nothing is known of the variation between the observed points. The correction, as defined here, must be *added* to the observed value to obtain the true value. An example of the data needed for the construction of a calibration curve is given in Table 6-1.

The first column of Table 6-1 shows the major scale divisions at which the calibrations were made. The second column lists the true values of the voltages, as determined by the potentiometer. The difference between the two values, the correction value, is shown in the third column. Figure 6-13 shows the calibration curve constructed from the values in Table 6-1.

When the voltmeter receives its *initial* calibration, for instance, in the case where the instrument is to be supplied with a new scale, the procedure is as follows: The voltage impressed on the instrument is adjusted by means of

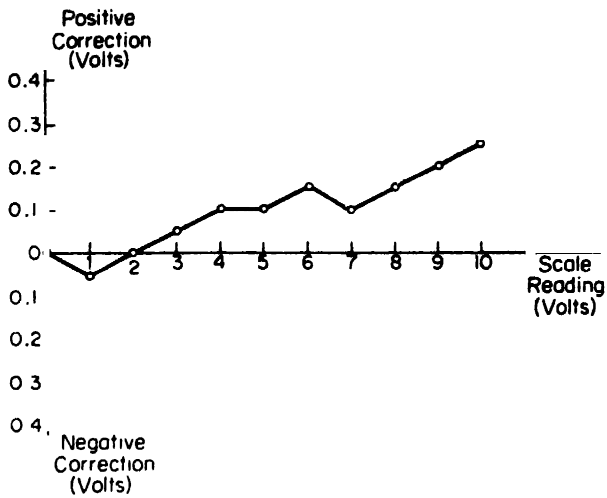


Figure 6-13
Typical calibration curve.

the two rheostats and set at one of the major voltage values (e.g., 1.0 V, 2.0 V, etc.). Adjustment is made until the potentiometer indicates that the desired value has indeed been reached, when the division mark is placed on the scale. A number of selected major voltage points are marked on the scale, and the intermediate values are interpolated.

Since the calibration process is rather time consuming, the potentiometer method is often used to calibrate a laboratory *standard voltmeter.* A standard voltmeter of this type is a very precise instrument, with a large mirror-backed scale to increase the accuracy of reading. The accuracy of such an instrument usually is better than 0.1 per cent of full-scale reading. Ordinary laboratory meters and panel instruments are then checked *by comparison* with this laboratory standard, rather than against a potentiometer.

Figure 6-14 shows the circuit used for calibrating an ammeter. A

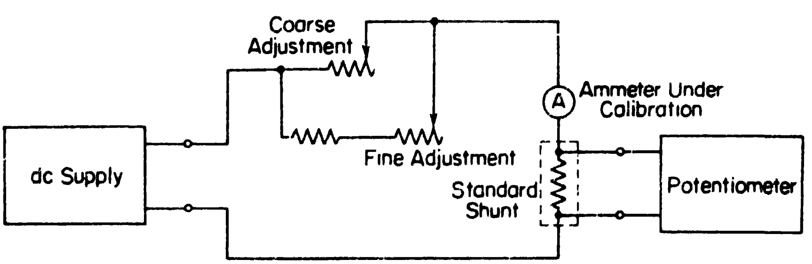


Figure 6-14
Calibration of an ammeter by the potentiometer method.

standard shunt of suitable resistance and sufficient current-carrying capacity is placed in series with the ammeter under calibration. The voltage across the shunt is measured by a potentiometer and the current through the shunt, and hence through the ammeter, is computed. Since the resistance of the standard shunt is accurately known and the voltage across the standard shunt is measured by the potentiometer, this method of calibrating an ammeter is very accurate. The procedure of calibration is very similar to that for the voltmeter. An *error curve* (or calibration curve) can be constructed in the manner discussed.

6-7 The Self-Balancing Potentiometer

The *self-balancing* potentiometer is widely used in industry because it eliminates the constant attention of an operator. Besides the automatic balancing feature, it also draws a curve of the quantity being measured and can be mounted on a switchboard or panel as a monitoring device. In the self-balancing instrument, the unbalance emf, which in a normal potentiometer would produce a current through the galvanometer, is applied to an amplifier via a converter. The output of the amplifier drives a two-phase induction motor which moves the potentiometer slider to balance. The converter, inserted between the potentiometer output and the amplifier input, converts the dc unbalance voltage into an ac unbalance voltage which can easily be amplified to the desired value by an ac amplifier. This scheme avoids use of a dc amplifier with its inherent problems of instability and drift.

The circuit diagram of Fig. 6-15 shows the schematic details of the self-balancing potentiometer which in this case is used for measuring temperature by a thermocouple. The converter consists of a vibrating reed, driven synchronously from the 60-Hz line voltage. The reed operates as a switch which reverses the current through the split winding of the transformer primary for each vibration of the reed. The constant reversal of current for each vibration cycle of the reed converts the unbalance dc voltage of the potentiometer circuit into an alternating voltage at the secondary of the transformer. The ac output of the converter, proportional to the dc input to the converter, is applied to the amplifier. The amplifier output is impressed on the control winding of the two-phase induction motor. The other winding of the motor is supplied by the line voltage. The ac line voltage is shifted 90° in phase with respect to the converter output voltage by the capacitor in the converter driving circuit. Depending on the polarity of the dc unbalance voltage applied to the converter input terminals, the phase of the amplifier output voltage will either lead or lag by 90° the line voltage applied to the induction motor. The direction of rotation of the motor is determined by the phase relationship between the two

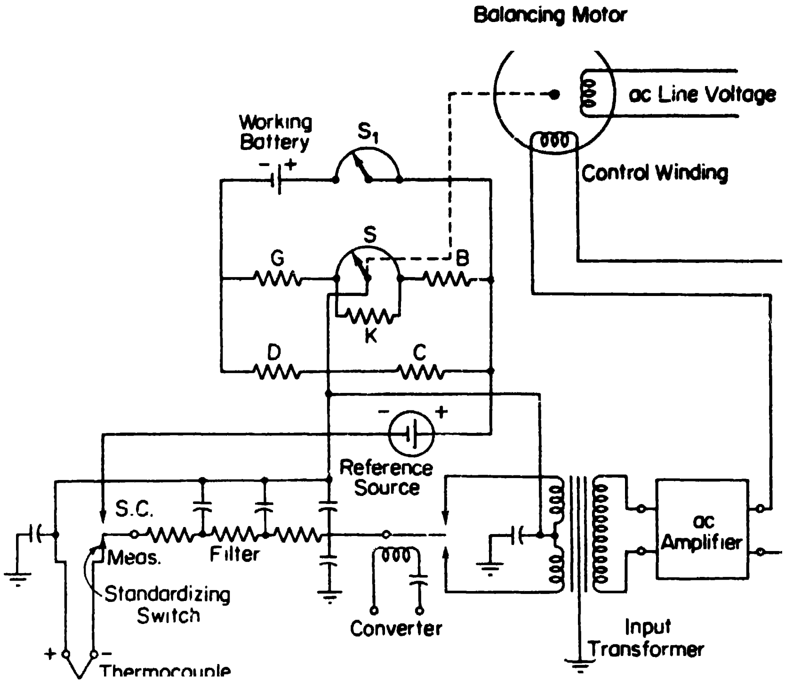


Figure 6-15
Circuit diagram of the Speed-0-Max self-balancing potentiometer.
(*Courtesy Leeds & Northrup Company.*)

voltages at the two windings, and this in turn is determined by the polarity of the voltage supplied to the converter. Thus, if the emf being measured is greater than the balancing voltage produced by the potentiometer, the motor will turn in one direction. If the emf being measured is smaller than the balancing voltage, then the amplifier output will be shifted by 180° and the motor will turn in the opposite direction.

The shaft of the motor is connected mechanically to the slide-wire contact in such a way, that the rotation of the motor decreases the unbalance in the potentiometer circuit. When the emf being measured is equal to the potentiometer voltage, the amplifier output voltage is zero and the motor does not rotate. Therefore, under any condition of unbalance, the amplifier output-voltage will cause the motor to move the potentiometer to balance.

The motor, which moves the slide-wire contact to maintain potentiometer balance, is mechanically coupled to a pen mechanism, and any movement of the slide-wire contact is followed by a simultaneous movement of the pen on a strip chart. The chart is driven by a separate clockmotor with an adjustable gear train to obtain the desired chart speed.

The emf produced by the thermocouple of Fig. 6-15 is a function of the temperature difference between the hot and the cold junction. (The operation of thermocouples is discussed in Chapter 5.) The variation in temperature of the reference junction is compensated by an electrical *compensating circuit*. The voltage drop across resistor D, which is made of a nickel-copper alloy, compensates for the change in temperature of the reference junction. Resistor G balances the voltage drop across D at the desired base temperature. Resistance K and slide wire S form the actual measuring circuit and resistor B produces the correct voltage drop for calibration of the circuit with the reference voltage, in this case a Zener-diode reference.

The signal supplied to the input of the potentiometer circuit is passed through a low-pass filter. The filter capacitors have no effect on the dc voltage supplied to the input, but any rapid variations in the input signal and any stray ac signals that may be impressed upon the input signal, are smoothed out by the action of the capacitors.

The photographs of Figs. 6-16, 6-17 show details of the construction of a *self-balancing recording potentiometer*. The balancing- and chart-drive motors, together with two ink-supply vessels, are shown in Fig. 6-16. The amplifiers of this dual-recording potentiometer are located in the right-hand corner of the instrument case. The power supply plus the Zener reference source is only partly visible at the left of the amplifiers inside the case. Figure 6-17 shows some

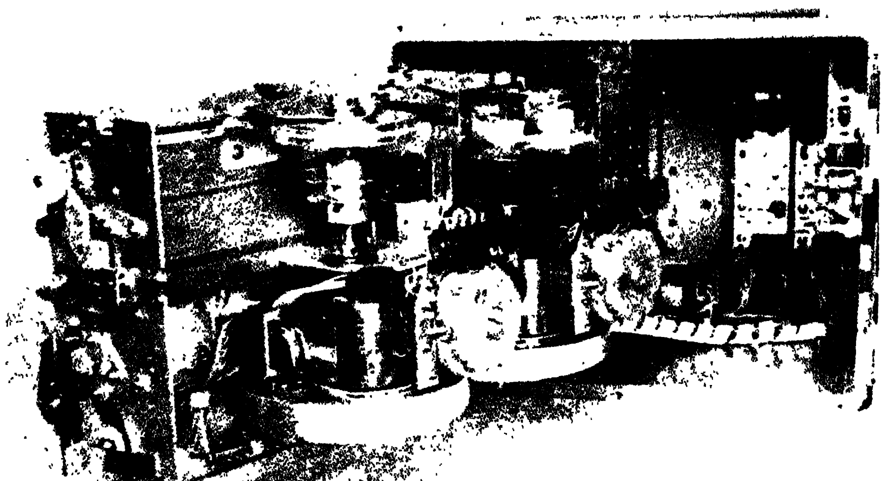


Figure 6-16
Inside view of a self-balancing recording potentiometer: Speedomax W/L recorder. (*Courtesy Leeds & Northrup Company.*)

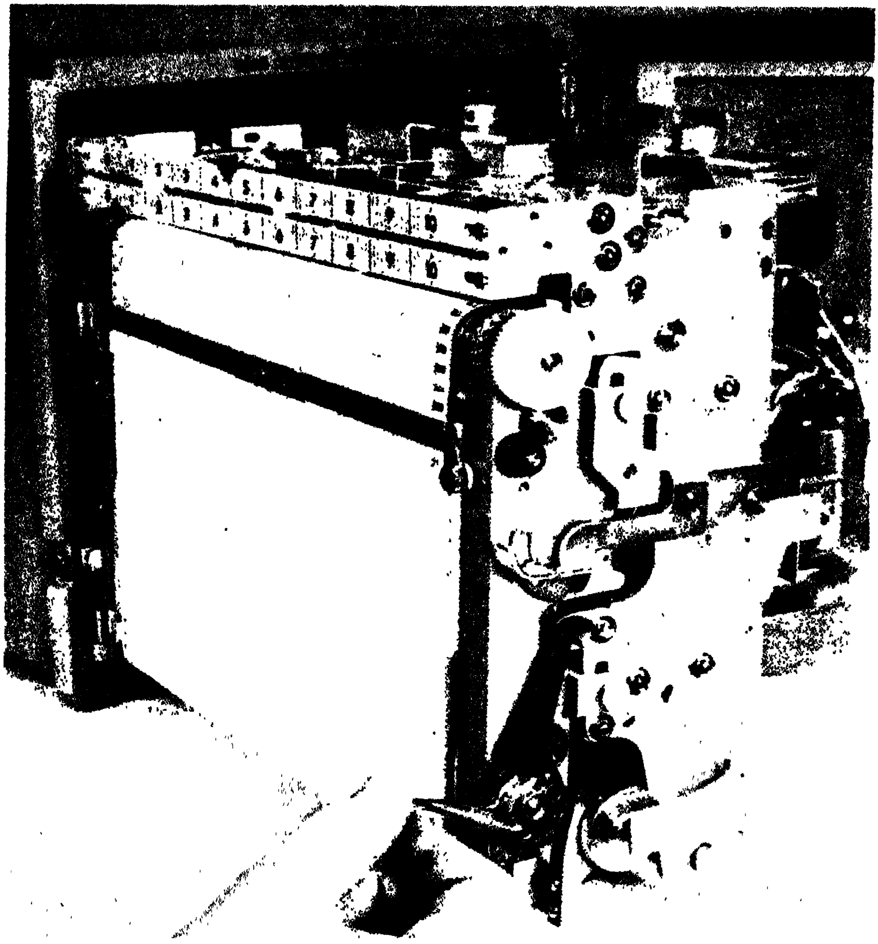


Figure 6-17
Photograph of the chart drive of the Speedomax recorder. (*Courtesy Leeds & Northrup Company.*)

of the construction details of the chart-drive mechanism. As can be clearly seen on this photograph, the instrument has two scales and two recording pens.

2. Frank, Ernest, *Electrical Measurement Analysis*, Chap. 9. New York: McGraw-Hill Book Company, Inc., 1959.

3. Stout, Melville B., *Basic Electrical Measurements*, 2nd ed., Chap. 7. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

4. Honeywell Catalog C-15a, *Electronik 15 Potentiometers*. Ft. Washington, Pa: Honeywell, Inc., Industrial Division, 1966.

Problems

1. The emf of a standard cell is measured with a potentiometer which gives a reading of 1.01892 V. When a 1-MΩ resistor is connected across the standard cell terminals, the potentiometer reading drops to 1.01874 V Calculate the internal resistance R_i of the standard cell.

2. A standard cell has an emf of 1.0190 V and an internal resistance of 250 Ω. A dc voltmeter with a full-scale range of 3 V and internal resistance of 3,000 Ω is connected across the standard cell. (a) Calculate the reading of the voltmeter. (b) Calculate the current drawn from the standard cell. (c) If the standard cell current in (b) exceeds 10 μA, calculate the value of the internal resistance which the voltmeter should have to limit the current to 10 μA.

3. The potentiometer of Fig. 6-1 has a working battery with a terminal voltage of 4.0 V and negligible internal resistance. The 200-cm slide wire has a resistance of 100 Ω and the internal resistance of the galvanometer is 50 Ω. The standard cell has an emf of 1.0191 V and internal resistance of 200 Ω. The rheostat is adjusted so that the potentiometer is standardized with the slider set at the 101.91-cm mark on the slide wire.
(a) Calculate the working current and the resistance of the rheostat.
(b) If the connections to the standard cell are accidentally reversed, calculate the current through the standard cell.
(c) A protective resistance is connected in series with the galvanometer to limit the current through the galvanometer to 10 μA for the conditions of (b). Calculate the resistance of this protective resistor.

4. The voltage between two points in a dc circuit is measured with a slide-wire potentiometer which gives a reading of 1.0 V. A 20-kΩ/V dc voltmeter registers only 0.5 V on its 2.5-V scale when connected to the same two points in the circuit. Calculate the circuit resistance between the two measured points.

5. For the circuit of Fig. 6-3, the main dial consists of fifteen steps of 20 Ω each, and the slide-wire resistance is 30 Ω. The standard cell voltage is 1.0190 V. The potentiometer is designed to have a measuring range of 1.65 V dc on the ×1 range. Calculate (a) the value of the measuring current

I_m on each range, (b) the current supplied by the battery I_t for each range, (c) the resistances of range resistors R_1 and R_2, (d) the required resistance of the rheostat if the working battery has a voltage of 6.0 V.

6. A potentiometer has fifteen steps of 5 Ω each and a slide wire of 5.5 Ω, in series with the working battery of 2.40 V and a rheostat. The maximum range of the instrument is 1.61 V. The galvanometer has a sensitivity of 0.05 μA/mm and an internal resistance of 50 Ω. (a) Calculate the resistance setting of the rheostat. (b) Determine the resolution of the instrument if the slide wire has eleven turns, one hundred divisions per turn, and can be interpolated to one-fifth of a division. (c) A 1.10-V source with negligible internal resistance is measured with this potentiometer. Calculate the error (in V) from the true balance, necessary to deflect the galvanometer spot 1 mm.

7. The potentiometer of Prob. 6 is first standardized and then balanced correctly against a dc voltage source of 1.50 V and 20 Ω internal resistance Calculate the deflection of the galvanometer when the slide wire is moved three divisions.

8. The potentiometer of Fig. 6-2 is designed with a dial of fifteen steps of 10 Ω, 0.1 V each, and a 10-Ω slide wire. The resistance of the 0–0.1-V dial step, however, is only 9.9 Ω instead of the required 10 Ω. The potentiometer is standardized against a reference voltage of 1.0185 V and is then used to measure an unknown voltage. The instrument reading at balance is 0.6525 V. Calculate (a) the true value of the unknown voltage, (b) the percentage error.

9. Deisgn a voltbox with a resistance of 20 Ω/V and ranges of 3 V, 10 V, 30 V, and 100 V. The voltbox is to be used with a potentiometer having a measuring range of 1.6 V.

10. Design a shunt with ranges of 1 A, 5 A, 10 A, and 20 A. The shunt is to be used with a potentiometer having a measuring range of 1.6 V.

SEVEN

DC BRIDGES AND THEIR APPLICATION

7-1 Introduction

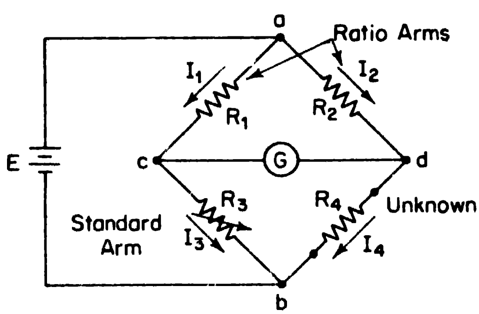
Bridge circuits are widely used for the measurement of resistance, capacitance, inductance, impedance, frequency, and other circuit parameters. Measurement accuracy of bridge circuits can be very high because their output is measured by the *null-indication* method, which is independent of the characteristics and the calibration of the null indicator. Bridges are not used for calibrating meters and other instruments (as the potentiometer is) but only for determining the value of an unknown component. A bridge circuit also allows *comparison* of an unknown component in terms of an accurately known *standard* component, and we speak of *comparison measurements*.

The common dc bridges are discussed in this chapter. Some of the better known ac bridges are analyzed in Chapter 8. Measurements involving either type of bridge depend on the balanced voltage output, which is characteristically found in the well-known Wheatstone bridge configuration.

7-2 The Wheatstone Bridge

Figure 7-1(a) is a photograph of a laboratory-type *Wheatstone bridge,* whose operation is based on the fundamental diagram of Fig. 7-1(b). It has four resistive arms, together with a source of emf (a battery) and a null detector, usually a galvanometer or other sensitive current meter. The current through the galvanometer depends on the potential difference between points

(a)



(b)

Figure 7-1
A laboratory-type Wheatstone bridge used for the precision measurement of resistances ranging from fractions of an ohm to several megohms. The *ratio* control switches the ratio arms in fixed increments. The remaining four step switches determine the value of the variable resistor in the Wheatstone bridge circuit. (a) Photograph of the instrument. (b) Simplified circuit diagram of the basic Wheatstone bridge. (*Courtesy Beckman Instruments, Inc. Cedar Grove Operations.*)

c and d. The bridge is said to be *balanced* when there is no current through the galvanometer or when the potential difference across the galvanometer is 0 V. This occurs when the voltage from point c to point a equals the voltage from point d to point a; or, by referring to the other battery terminal, when the voltage from point c to point b equals the voltage from point d to point b. For bridge balance, we can write therefore,

$$I_1R_1 = I_2R_2 \tag{7-1}$$

For the galvanometer current to be zero, the following conditions also exist:

$$I_1 = I_3 = \frac{E}{R_1 + R_3} \tag{7-2}$$

and

$$I_2 = I_4 = \frac{E}{R_2 + R_4} \tag{7-3}$$

Combining Eqs. (7-1), (7-2), and (7-3) and simplifying, we obtain

$$\frac{R_1}{R_1 + R_3} = \frac{R_2}{R_2 + R_4} \tag{7-4}$$

from which,

$$R_1R_4 = R_2R_3 \tag{7-5}$$

Equation (7-5) is the well-known expression for balance of the Wheatstone bridge. If three of the resistances have known values, the fourth may be determined from Eq. (7-5), and we obtain

$$R_4 = R_3\frac{R_2}{R_1} \tag{7-6}$$

where R_3 is called the *standard arm* of the bridge and R_2 and R_1 are called the *ratio arms.*

The measurement of the unknown resistance, R_4, is independent of the characteristics or the calibration of the null-deflecting galvanometer, provided that the null detector has sufficient sensitivity to indicate the balance position of the bridge with the required degree of precision.

The Wheatstone bridge is widely used for precision measurement of resistance from approximately 1 Ω to the low megohm range. The main source of *measurement error* is found in the limiting errors of the three known resistors (see Chapter 1, Prob. 11). Other errors may include the following:

(a) Insufficient sensitivity of the null detector. This problem is discussed more fully in Sec. 7-3.

(b) Changes in resistance of the bridge arms due to the heating effect of the current through the resistors. Heating effect (I^2R) of the bridge arm currents may change the resistance of the resistor in question. The rise in temperature affects the resistance during the actual measurement, and excessive currents may cause a permanent change in resistance values. This may not be discovered in time and subsequent measurements could well be erroneous. The power dissipation, therefore, must be computed *in advance,* particularly when low-resistance values are to be measured and the current must be limited to a safe value.

(c) Thermal emf's in the bridge circuit or the galvanometer circuit may again cause problems when measuring low-value resistors. Some of the more sensitive galvanometers have copper coils and copper suspension systems to avoid having dissimilar metals in contact with one another, generating thermal emf's.

(d) Errors due to the resistance of leads and contacts exterior to the actual bridge circuit play a role in the measurement of very low resistance values. These errors may be reduced through use of a Kelvin bridge (see Sec. 7-4).

7-3 Sensitivity of the Wheatstone Bridge

Computation of the current through the galvanometer due to an unbalance condition is necessary to determine whether the galvanometer has the *sensitivity* required to detect this unbalance. Different galvanometers not only may require different currents per unit deflection (current sensitivity) but they also may have a different internal resistance. It is impossible to say, without prior computation, which galvanometer will make the bridge, as a complete circuit, more sensitive to an unbalance condition. The sensitivity to unbalance can be computed by "solving" the bridge circuit for a *small* unbalance. The solution is approached by converting the Wheatstone bridge of Fig. 7-1 to its *Thévenin equivalent* circuit.

Since we are interested in the current through the galvanometer, the Thévenin equivalent circuit is determined by looking into the galvanometer terminals, points c and d in Fig. 7-1. Two steps must be taken to find the Thévenin equivalent: The first involves finding the *equivalent voltage* appearing at terminals c and d when the galvanometer is removed from the circuit. The second step involves finding the *equivalent resistance* looking into terminals c and d with the battery replaced only by its internal resistance. For convenience, the circuit of Fig. 7-1 is redrawn in Fig. 7-2.

The Thévenin, or open-circuit, voltage is found by referring to Fig. 7-2(a) and we can write:

$$E_{cd} = E_{ac} - E_{ad} = I_1 R_1 - I_2 R_2$$

where $$I_1 = \frac{E}{R_1 + R_3} \quad \text{and} \quad I_2 = \frac{E}{R_2 + R_4}$$

Therefore,

$$E_{cd} = E\left[\frac{R_1}{R_1 + R_3} - \frac{R_2}{R_2 + R_4}\right] \tag{7-7}$$

This is the voltage of the equivalent Thévenin generator.

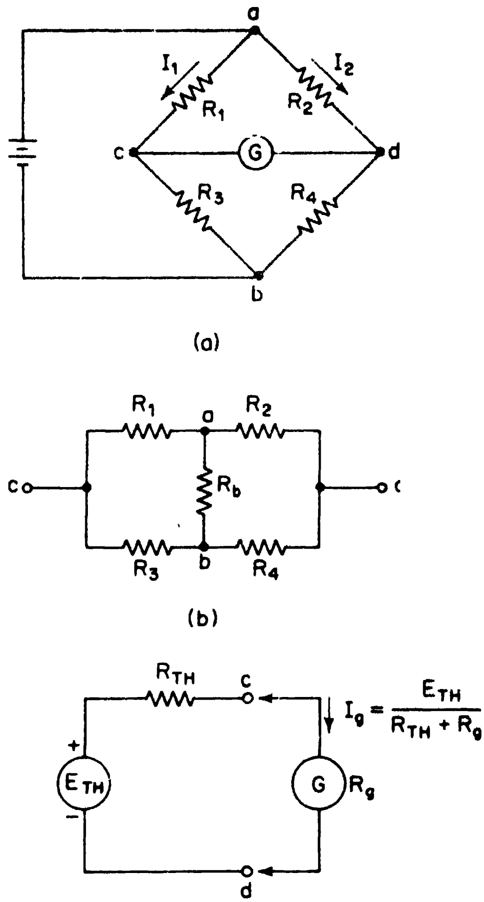


Figure 7-2
Illustrating the application of Thévenin's theorem to the Wheatstone bridge. (a) The Wheatstone bridge configuration. (b) The Thévenin resistance, looking into terminals *c* and *d*. (c) The complete Thévenin circuit, with the galvanometer connected to terminals *c* and *d*.

The resistance of the Thévenin equivalent circuit is found by looking back into terminals c and d, replacing the battery by its internal resistance. The circuit of Fig. 7-2(b) represents the Thévenin resistance. Notice that the internal resistance, R_t, of the battery has been included in Fig. 7-2(b). Converting this circuit into a more convenient form requires use of the *delta-wye transformation* theorem. Readers interested in this approach should consult texts on circuit analysis where this theorem is derived and applied.* In most cases, however, the *extremely low* internal resistance of the battery can be *neglected* and this simplifies the reduction of Fig. 7-2(a) to its Thévenin equivalent considerably.

Referring to Fig. 7-2(b), we see that a short circuit exists between points a and b when the internal resistance of the battery is assumed to be 0 Ω. The Thévenin resistance, looking into terminals c and d, then becomes

$$R_{TH} = \frac{R_1R_3}{R_1 + R_3} + \frac{R_2R_4}{R_2 + R_4} \tag{7-8}$$

The Thévenin equivalent of the Wheatstone bridge circuit therefore reduces to a Thévenin generator with an emf described by Eq. (7-7) and an internal resistance given by Eq. (7-8). This is shown in the circuit of Fig. 7-2(c).

When the null detector is now connected to the output terminals of the Thévenin equivalent circuit, the *galvanometer current* is found to be

$$I_g = \frac{E_{TH}}{R_{TH} + R_g} \tag{7-9}$$

where I_g is the galvanometer current and R_g its resistance.

Example 7-1: Figure 7-3(a) shows the schematic diagram of a Wheatstone bridge with values of the bridge elements as shown. The battery has an emf of 5 V and negligible internal resistance. The galvanometer has a current sensitivity of 10 mm/μA and an internal resistance of 100 Ω. Calculate the deflection of the galvanometer caused by the 5-Ω unbalance in arm *BC*.

SOLUTION: Bridge balance occurs if arm *BC* has a resistance of 2,000 Ω. The diagram shows arm *BC* as a resistance of 2,005 Ω, representing a small unbalance (≪ 2,000 Ω). The first step in the solution consists of converting the bridge circuit into its equivalent Thévenin circuit. Since we are interested in finding the current in the galvanometer, the Thévenin equivalent is determined with respect to galvanometer terminals *B* and *D*. The potential difference from *B* to *D*, with the galvanometer removed from the circuit is the Thévenin voltage using Eq. (7-7):

$$E_{TH} = E_{AD} - E_{AB} = 5\text{ V} \times \left[\frac{100}{100 + 200} - \frac{1,000}{1,000 + 2,005}\right]$$
$$= 2.5\text{ mV}$$

The second step of the solution involves finding the equivalent Thévenin resistance, looking into terminals *B* and *D*, replacing the battery with its internal resistance. Since the battery resistance is 0 Ω, the circuit is represented by the configuration of Fig. 7-3(b) from which we find

$$R_{TH} = \frac{100 \times 200}{300} + \frac{1,000 \times 2,005}{3,005} = 734\ \Omega$$

The Thévenin equivalent circuit is given in Fig. 7-3(c). When the galvanometer is now connected to the output terminals of the equivalent circuit, the current through the galvanometer is

$$I_g = \frac{E_{TH}}{R_{TH} + R_g} = \frac{2.5\text{ mA}}{743\ \Omega + 100\ \Omega} = 3\ \mu\text{A}$$

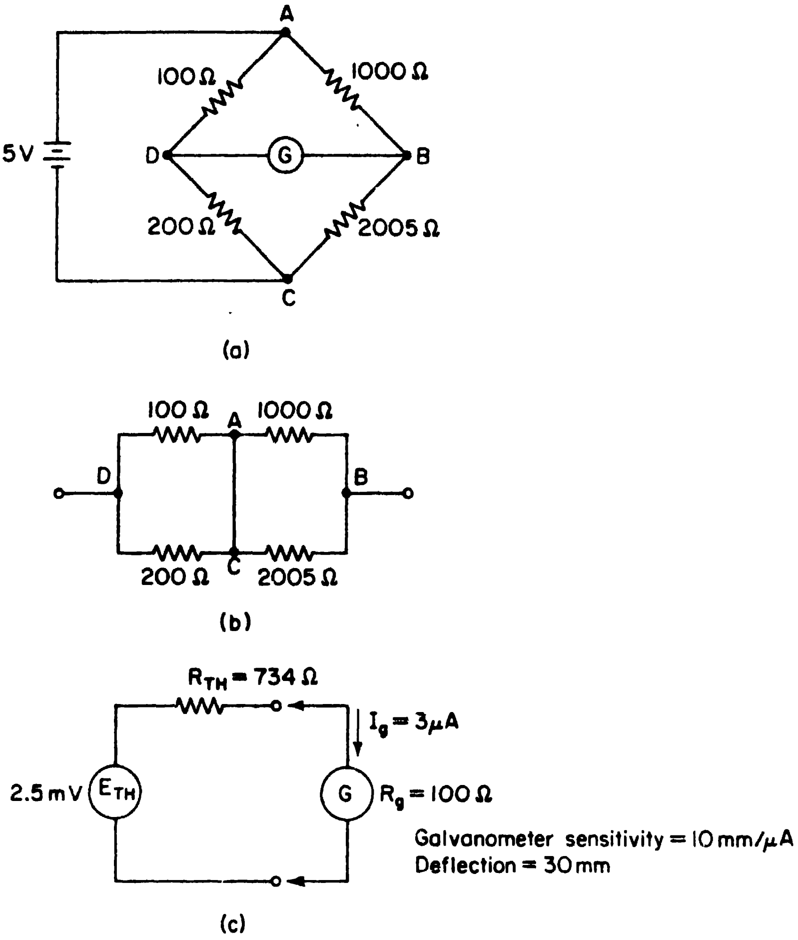


Figure 7-3
Calculation of galvanometer deflection caused by a small unbalance in arm *BC*, using the simplified Thévenin approach.

The galvanometer deflection is

$$d = 3\ \mu\text{A} \times \frac{10\ \text{mm}}{\mu\text{A}} = 30\ \text{mm}$$

At this point the merit of the Thévenin equivalent circuit for the solution of an unbalanced bridge becomes quite evident. If a *different* galvanometer is used (with a change in current sensitivity and internal resistance), the computation of this new galvanometer's deflection is very simple, as is clear from Fig. 7-3(c). On the other hand, if the galvanometer sensitivity is *given*, we can solve for the specific unbalance needed to give a unit deflection (say 1 mm). This value is of interest when we want to determine the sensitivity of the bridge to unbalance, or in response to the question: "Is the galvanometer selected capable of detecting a certain small unbalance?" The Thévenin method is used to find the galvanometer response, which in most cases is of prime interest.

Example 7-2: The galvanometer of Example 7-1 is replaced by one with an internal resistance of 500 Ω and a current sensitivity of 1 mm/μA. Assuming that a deflection of 1 mm can be observed on the galvanometer scale, determine if this new galvanometer is capable of detecting the 5-Ω unbalance in arm *BC* of Fig. 7-3(a).

SOLUTION: Since the bridge constants have not been changed, the equivalent circuit is again represented by a Thévenin generator of 2.5 mV and a Thévenin resistance of 734 Ω. The new galvanometer is now connected to the output terminals, resulting in a galvanometer current

$$I_g = \frac{E_{TH}}{R_{TH} + R_g} = \frac{2.5\text{ mA}}{734\ \Omega + 500\ \Omega} = 2.02\ \mu\text{A}$$

The galvanometer deflection therefore equals $2.02\ \mu\text{A} \times 1\ \text{mm}/\mu\text{A} = 2.02$ mm, indicating that this galvanometer produces a deflection which can be observed adequately.

The Wheatstone bridge is limited to the measurement of resistances ranging from a few ohms to several megohms. The *upper* limit is set by the reduction in sensitivity to unbalance, caused by high resistance values, because in this case the equivalent Thévenin resistance of Fig. 7-3(c) becomes high, thus reducing the galvanometer current. The *lower* limit is set by the resistance of the connecting leads and the contact resistance at the binding posts. The resistance of the leads could be calculated or measured and the final result modified, but contact resistance is very hard to compute or measure. For *low*-resistance measurements, therefore, the *Kelvin* bridge is generally the preferred instrument.

7-4 The Kelvin Bridge

The Kelvin bridge is a modification of the Wheatstone bridge and provides greatly increased accuracy in measurement of *low-value resistances,* generally below 1 Ω. The last paragraph of Sec. 7-3 said that the resistance of the leads connecting the unknown low-resistance resistor to the terminals of the bridge circuit may affect the result of the measurement. Consider the bridge circuit shown in Fig. 7-4, where R_y represents the resistance of the connecting lead from R_3 to R_x. Two galvanometer connections are possible, to point C or to point D. When the galvanometer is connected to point D, the resistance,

R_y, of the connecting lead is added to the *unknown* R_x, resulting in too high an indication for R_x. When connection is made to point C, R_y is added to *bridge arm*, R_3, and the resulting measurement of R_x will be lower than it should be, because now the actual value of R_3 is higher than its nominal value (by the amount of R_y). If the galvanometer is connected to a point P, in between the two points C and D, in such a way that the ratio of the resistances from C to P and from D to P equals the ratio of the resistors R_1 and R_2, we can write

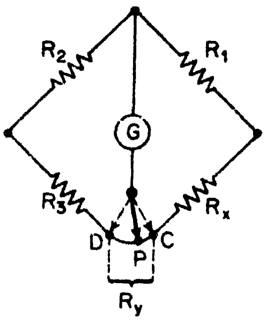


Figure 7-4
A Wheatstone bridge circuit, showing the resistance of the lead from point C to point D.

$$\frac{R_{CP}}{R_{DP}} = \frac{R_1}{R_2} \tag{7-10}$$

and the usual balance equation for the bridge delivers the relationship·

$$R_x + R_{CP} = \frac{R_1}{R_2}(R_3 + R_{DP}) \tag{7-11}$$

Substituting Eq. (7-10) into Eq. (7-11), we obtain

$$R_x + \left(\frac{R_1}{R_1 + R_2}\right)R_y = \frac{R_1}{R_2}\left[R_3 + \left(\frac{R_2}{R_1 + R_2}\right)R_y\right] \tag{7-12}$$

which reduces to

$$R_x = \frac{R_1}{R_2}R_3 \tag{7-13}$$

Equation (7-13) is the usual balance equation developed for the Wheatstone bridge and it indicates that the effect of the resistance of the connecting lead from point C to point D has been eliminated by connecting the galvanometer to the intermediate position, P.

The foregoing development forms the basis for construction of the Kelvin double bridge, commonly known as the *Kelvin bridge*. The name *double bridge* is used because a *second* set of ratio arms is employed. Figure 7-5 shows the schematic diagram of the Kelvin bridge. The second set of ratio arms, a and b,

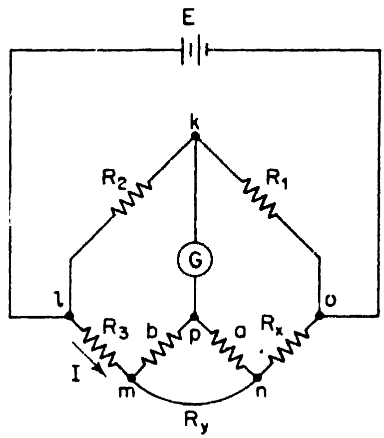


Figure 7-5
The basic Kelvin double-bridge circuit.

is used to connect the galvanometer to a point P at the appropriate potential between C and D, eliminating the effect of the C-to-D connection. 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 will be 0 V when $E_{kl} = E_{lmp}$, where

$$E_{kl} = \frac{R_2}{R_1 + R_2}E = \frac{R_2}{R_1 + R_2}I\left[R_3 + R_x + \frac{(a+b)R_y}{a+b+R_y}\right] \quad \text{and} \qquad (7\text{-}14)$$

$$E_{lmp} = I\left[R_3 + \frac{b}{a+b}\left\{\frac{(a+b)R_y}{a+b+R_y}\right\}\right] \qquad (7\text{-}15)$$

Since, for zero galvanometer deflection, $E_{kl} = E_{lmp}$, we can use Eqs. (7-14) and (7-15) to solve for R_x, obtaining

$$R_x = R_3\frac{R_1}{R_2} + \frac{bR_y}{a+b+R_y}\left\{\frac{R_1}{R_2} - \frac{a}{b}\right\} \qquad (7\text{-}16)$$

Using the initially established condition that $a/b = R_1/R_2$, Eq. (7-16) reduces to the well-known relationship:

$$R_x = R_3\frac{R_1}{R_2} \qquad (7\text{-}17)$$

Equation (7-17) is the usual working equation for the Kelvin bridge and it

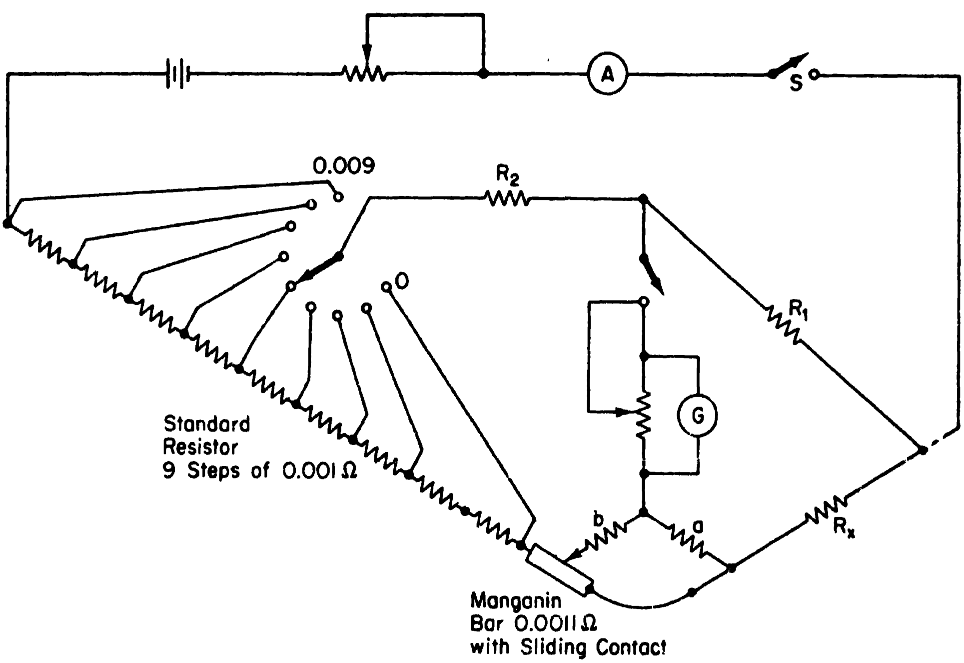


Figure 7-8
The simplified circuit of a Kelvin bridge for the measurement of very low resistances.

indicates that the resistance of the yoke has no effect on the measurement, provided that the two sets of ratio arms have equal ratios.

The Kelvin bridge is used for measuring very low resistances, from approximately 1 Ω to as low as 0.00001 Ω, provided that the correct bridge components are selected. Figure 7-6 shows the circuit diagram of a commercial Kelvin bridge capable of measuring resistances from 10 Ω to 0.00001 Ω. In this bridge, resistance R_3 in Eq. (7-17) is made continuously variable and is represented by the variable *standard* resistor in Fig. 7-6. The ratio arms (R_1 and R_2) are usually provided with a limited number of discreet steps.

Contact potential drops in the measuring circuit may cause large errors and to reduce this effect the standard resistor consists of nine discreet steps of 0.001 Ω each plus a calibrated manganin bar of 0.0011 Ω with a sliding contact. The total resistance of the R_3 arm, therefore, amounts to 0.0101 Ω and is variable in steps of 0.001 Ω plus fractions of 0.0011 Ω by the sliding contact. When both contacts are switched to select the suitable value of standard resistor, the voltage drop between the ratio-arm connection points is changed, but the total resistance around the battery circuit is unchanged. This arrangement, therefore, places any contact resistance in series with the relatively high resistance values of the ratio arms and the contact resistance has negligible effect.

The ratio R_1/R_2 should be so selected that a relatively large part of the standard resistance is used in the measuring circuit. Thus the value of the unknown resistance R_x is determined with the largest possible number of significant figures, and the *accuracy* of measurement is improved.

7-5 The Megohm Bridge and the Measurement of Very High Resistance

The measurement of very high resistances, on the order of 10 megohms to several thousand megohms, presents some special problems. Several instruments and techniques have been developed to make reliable measurements in this range possible. One major difficulty involves the presence of *leakage paths* which occur over and around the component, the binding posts, and inside the instrument itself. Leakage resistance affects the measurement of the unknown by its shunting or paralleling action. It is very difficult to control leakage, particularly since the moisture in the air varies its magnitude from day to day or from location to location.

Generally, some form of *guard circuit* is used to remove the effects of leakage paths. Figure 7-7 illustrates the operation of a guard circuit. In Fig. 7-7(a) a high resistance, mounted on a piece of insulating material, is measured by the voltmeter-ammeter method. The microammeter measures the sum of the current through the resistor (I_R) and the current through the leakage path

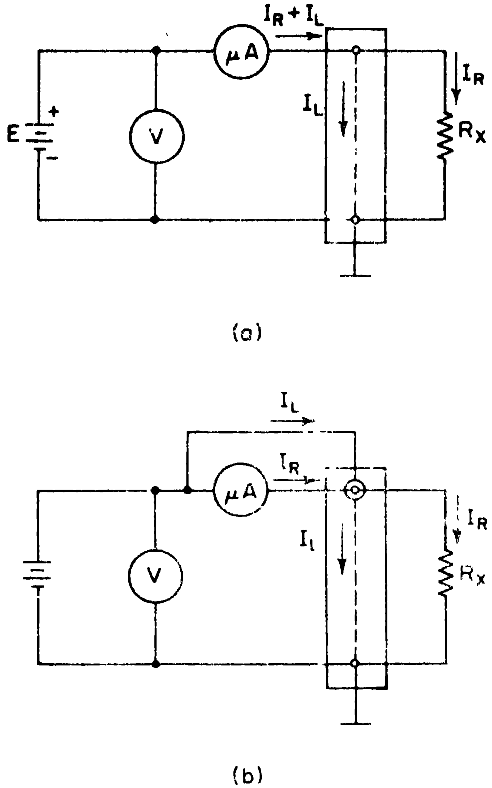


Figure 7-7
The operation of a guard-terminal for the measurement of very high resistances. Distribution of currents: (a) Without a guard circuit. (b) With a guard circuit.

around the resistor (I_L). The resistance, which is computed from the readings indicated on the voltmeter and the microammeter, will not be the true value, but will be in error $[R = E/(I_R + I_L)]$. In Fig. 7-7(b) a guard terminal has been added to the resistance terminal-block. The guard terminal surrounds the resistance terminal entirely and is connected to the battery side of the microammeter. The leakage current (I_L) now bypasses the current meter, which then indicates the current through the resistor (I_R) only and allows accurate determination of the resistance value. The guard terminal and the resistance terminal are at approximately the same potential and practically no current will flow *between* them.

A simple guard circuit for use in a conventional Wheatstone bridge is shown in Fig. 7-8. A high resistance is mounted on two insulating posts which are fastened to a metal mounting-plate. The resistance is connected to the "unknown" terminals of the bridge. The leakage resistances, indicating the leakage paths along the insulating posts, are represented by R_1 and R_2. The metal mounting-plate, which is the common point of R_1 and R_2, is connected to the junction of the ratio arms, R_A and R_B. This connection puts leakage resistance R_1 in parallel with ratio arm R_A, but since R_1 is usually so very much larger than R_A, its shunting effect is negligible. Leakage resistance R_2 is placed in parallel with the galvanometer, but its resistance is so much higher than the galvanometer resistance that its only effect may be a slight reduction in galvanometer sensitivity. If the guard circuit were not used, the leakage resistances would be in parallel with the unknown, reducing its measured value considerably.

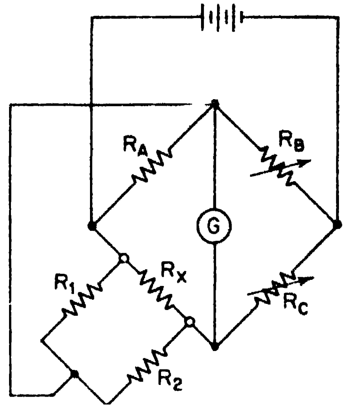


Figure 7-8
A conventional Wheatstone bridge using a guard circuit to improve measurement accuracy.

Commercial megohm bridges usually have the junction of the ratio arms, R_A and R_B, brought out as a separate "guard" terminal on the front panel, allowing the three-terminal resistor to be connected.

An example of a high-voltage megohm bridge is found in Fig. 7-9, where the various controls can easily be identified. This modified Wheatstone bridge uses a vacuum tube voltmeter (VTVM) as a sensitive null indicator. The resistance multiplier, which corresponds to resistor R_c in the diagram, consists of ten parallel resistors and provides the magnification of the variable scale-reading R_B. The dc voltage is adjustable to 1,000 V, providing increased sensitivity to unbalance at high resistance values. The junction of the ratio arms, R_A and R_B, is brought out as a front-panel guard terminal. Provision is made, through binding posts on the instrument panel, for the connection of an external null indicator which may consist of a high-gain amplifier and indicator (an electrometer) or an oscilloscope (CRO).

Measurement of high resistances has many applications in the electrical industry. A few examples are worth mentioning:

(a) Measuring sheets of insulating material used in electric machines.

(b) Measuring insulating material used in wrapping electrical wires and cables.

(c) Measuring the leakage resistance of capacitors.

The high-voltage megohm bridge is not the only instrument used for high-resistance measurements. Other methods may include use of the well-known "*megger*," the *direct-deflection method* for testing insulation samples, the

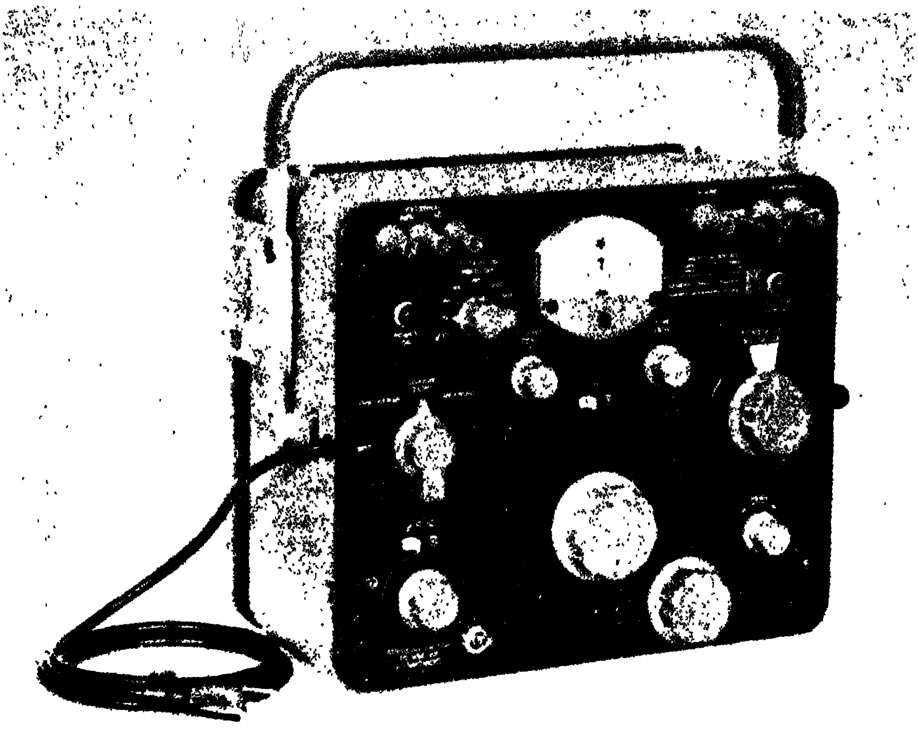


Figure 7-9
A commercial Megohm bridge, used for the measurement of resistances in the terra-ohm range. (*Courtesy General Radio Company.*)

loss-of-charge method for checking the leakage resistance of capacitors. (See several references and manufacturers' publications describing these specialized methods.)*

References

1. Frank, Ernest, *Electrical Measurement Analysis*, Chap. 10. New York: McGraw-Hill Book Company, Inc., 1959.

2. Stout, Melville B., *Basic Electrical Measurements*, 2nd ed., Chap. 4. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

*Cf. Stout, *op. cit.*, pp. 126–133.

Problems

1. The four arms of a Wheatstone bridge have resistances of 100 Ω, 1,000 Ω, 500 Ω, and 50.5 Ω, taken in sequence around the bridge. A galvanometer with internal resistance of 75 Ω is connected from the junction of the 100-Ω and 50.5-Ω resistors to the junction of the 1,000-Ω and 500-Ω resistors. A 4-V battery is connected to the other two corners of the bridge. Use Thévenin's theorem to find (a) the equivalent circuit of this bridge for calculating the galvanometer current; (b) the current through the galvanometer

2. The ratio arms of the Wheatstone bridge of Fig. 7-1 are $R_1 = 1{,}000\ \Omega$ and $R_2 = 100\ \Omega$, the standard resistance $R_3 = 400\ \Omega$, the unknown $R_x = 41\ \Omega$. A 1.5-V battery with negligible internal resistance is connected from the junction of R_1 and R_2 to the junction of R_3 and R_x. The galvanometer, with an internal resistance of 50 Ω and a current sensitivity of 2 mm/μA, is connected to the other corners of the bridge. (a) Find the equivalent circuit of the bridge with respect to the galvanometer terminals. (b) Calculate the deflection of the galvanometer caused by the unbalance in the circuit.

3. Repeat Prob. 2 with galvanometer and battery interchanged and determine which configuration is more sensitive to unbalance.

4. The standard resistance arm R_3 in Prob. 2 can be adjusted from 0 Ω to 1,000 Ω in steps of 0.1 Ω. The galvanometer deflection can be read within 0.5 mm. When the unknown resistance $R_x = 50\ \Omega$, calculate (a) the resolution of the reading, in ohms, (b) the resolution of the reading, in per cent of the unknown R_x.

5. The three known resistance arms of a Wheatstone bridge have limiting errors of $\pm 0.1\ \%$. Determine the limiting error of the unknown resistance when measured with this instrument.

6. For the bridge circuit of Fig. 7-1, $R_1 = 1{,}000\ \Omega$, $R_2 = 4{,}000\ \Omega$, $R_3 = 100\ \Omega$, and $R_4 = 400\ \Omega$, indicating that the bridge is balanced. The galvanometer has an internal resistance of 100 Ω and a current sensitivity of 100 mm/μA. The battery voltage is 3.0 V. Calculate the galvanometer deflection for an unbalance of 1 Ω in resistance arm R_4. (*Hint:* Determine the Thévenin voltage and resistance in terms of the small unbalance x in R_4. Reduce both the expressions for Thévenin voltage and resistance, dropping the increment x from the denominator after reduction.)

7. The Wheatstone bridge of Fig. 7-1 has ratio arms of 1,000 Ω and 100 Ω and is used to measure an unknown resistance of 25 Ω. The battery has negligible internal resistance and an emf of 1 V. Two galvanometers are

available. Galvanometer A has an internal resistance of 25 Ω and a current sensitivity of 20 mm/μA. Galvanometer B has an internal resistance of 200 Ω and a current sensitivity of 100 mm/μA. Calculate (a) the sensitivity of each galvanometer to unbalance in the R_x arm, expressed in mm/Ω; (b) the ratio of the galvanometer sensitivities to unbalance.

8. A Wheatstone bridge is used to measure high resistances (in the megohm range). The bridge has ratio arms of 10 kΩ and 10 Ω. The standard arm is variable and may be adjusted from 0 Ω to 10 kΩ. A battery of 10 V and negligible internal resistance is connected from the junction of the ratio arms to the opposite corner. (a) Calculate the maximum resistance that can be measured by this arrangement. (b) If the galvanometer has a sensitivity of 200 mm/μA and a resistance of 50 Ω, and the maximum resistance of part (a) is connected to the R_x terminals, calculate the unbalance necessary to produce a galvanometer deflection of 1 mm. (c) If the galvanometer is replaced by one with a current sensitivity of 1,000 mm/μA and internal resistance of 1,000 Ω, calculate the unbalance in R_x needed to produce a galvanometer deflection of 1 mm.

9. Each of the ratio arms of a laboratory-type Wheatstone bridge has a guaranteed accuracy of $\pm 0.05\%$, while the standard arm has a guaranteed accuracy of $\pm 0.1\%$. The ratio arms are both set at 1,000 Ω (a 1/1 ratio), and the bridge is balanced with the standard arm adjusted to 3,154 Ω. Determine the upper and lower limits of the unknown resistance, based on the guaranteed accuracies of the known bridge arms.

10. The ratio arms of the Kelvin bridge of Fig. 7-5 are 100 Ω each. The galvanometer has an internal resistance of 500 Ω and a current sensitivity of 200 mm/μA. The unknown resistance $R_x = 0.1002\ \Omega$, and the standard resistance is set at 0.1000 Ω. A dc current of 10 A is passed through the standard and the unknown from a 2.2-V battery in series with a rheostat. The resistance of the yoke may be neglected. Calculate (a) the deflection of the galvanometer, in millimeters, (b) the resistance unbalance required to produce a galvanometer deflection of 1 mm. (*Hint:* In the calculation of the Thévenin voltage and resistance assess the effects of the ratio arms and the rheostat and neglect the appropriate terms.)

11. The ratio arms of a Kelvin bridge are 1,000 Ω each. The galvanometer has an internal resistance of 100 Ω and a current sensitivity of 500 mm/μA. A dc current of 10 A is passed through the standard arm and the unknown from a 2.2-V battery in series with a rheostat. The standard resistance is set at 0.1000 Ω and the galvanometer deflection is 30 mm. Neglecting the resistance of the yoke, determine the value of the unknown.

EIGHT

AC BRIDGES AND THEIR APPLICATION

8-1 The General Equation for Bridge Balance

The ac bridge is a natural outgrowth of the dc bridge and in its basic form consists of four bridge arms, a source of excitation, and a null detector. The power source supplies ac to the bridge at the desired frequency. For measurements at low frequencies, the power line may serve as the source of excitation; at higher frequencies, a transistorized or vacuum tube oscillator may supply power to the bridge. The null detector must respond to ac unbalance currents and in its cheapest (but very effective) form, consists of a pair of headphones. An ac amplifier with a suitable indicator, e.g., an electrometer, an oscilloscope (CRO), an ac current meter, or an electron ray tube (tuning eye) indicator may be used in other applications. Figure 8-1 shows the general form of an ac bridge.

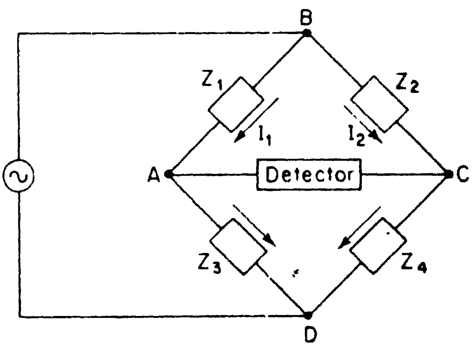


Figure 8-1
General form of the ac bridge.

The four bridge arms $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3$, and $\mathbf{Z}_4$ are indicated as unspecified impedances and the detector is represented by headphones. As was discussed in the case of the basic Wheatstone bridge for dc measurements, the balance condition in an ac bridge is reached when the detector response is zero, or indicates *a null.* Balance adjustment to obtain a null response is made by varying one or more of the bridge arms.

The general equation for bridge balance is obtained by using *complex notation* for the impedances of the bridge circuit. (Boldface type is used to indicate quantities in complex notation.) These quantities may be impedances or admittances as well as voltages or currents. The conditions for bridge balance again require that the potential difference from point A to point C in Fig. 8-1 be zero. This will be the case when the voltage drop from B to A equals the voltage drop from B to C, in *both magnitude and phase.* In complex notation, we can write

$$\mathbf{E}_{BA} = \mathbf{E}_{BC} \quad \text{or} \quad \mathbf{I}_1\mathbf{Z}_1 = \mathbf{I}_2\mathbf{Z}_2 \tag{8-1}$$

For zero detector current, the balance condition, the currents are

$$\mathbf{I}_1 = \frac{\mathbf{E}}{\mathbf{Z}_1 + \mathbf{Z}_3} \tag{8-2}$$

and

$$\mathbf{I}_2 = \frac{\mathbf{E}}{\mathbf{Z}_2 + \mathbf{Z}_4} \tag{8-3}$$

Substitution of Eqs. (8-2) and (8-3) into Eq. (8-1) yields

$$\mathbf{Z}_1\mathbf{Z}_4 = \mathbf{Z}_2\mathbf{Z}_3 \tag{8-4a}$$

or, when using admittances instead of impedances,

$$\mathbf{Y}_1\mathbf{Y}_4 = \mathbf{Y}_2\mathbf{Y}_3 \tag{8-4b}$$

Equation (8-4a) is the most convenient form in most cases and is the *general equation for balance* of the ac bridge. Equation (8-4b) is often used to advantage when dealing with parallel components in bridge arms. A combination of impedances and admittances may be used, since impedance is most convenient for bridge arms consisting of series elements and admittance is most convenient for arms with parallel components.

Equation (8-4a) states that the product of impedances of one pair of opposite arms must equal the product of impedances of the other pair of opposite arms, with the impedances expressed in complex notation. In other words, both the magnitudes and the phase angles of the impedances (or admittances) must be taken into account. If the impedance is written in the form $\mathbf{Z} = Z\angle\theta$, where Z represents the magnitude and θ the phase angle of the complex impedance, Eq. (8-4a) can be rewritten in the form

$$(Z_1\angle\theta_1)(Z_4\angle\theta_4) = (Z_2\angle\theta_2)(Z_3\angle\theta_3) \tag{8-5}$$

Since in multiplication of complex numbers the magnitudes are *multiplied* and the phase angles are *added*, Eq. (8-5) reduces to

$$Z_1 Z_4 \angle(\theta_1 + \theta_4) = Z_2 Z_3 \angle(\theta_2 + \theta_3) \tag{8-6}$$

Equation (8-6) shows that *two conditions* must be met simultaneously when balancing an ac bridge: The first condition is that the magnitude of the impedances satisfy the relationship

$$Z_1 Z_4 = Z_2 Z_3 \tag{8-7}$$

The second requirement demands that the phase angles of the impedances satisfy the relationship

$$\angle\theta_1 + \angle\theta_4 = \angle\theta_2 + \angle\theta_3 \tag{8-8}$$

The phase angles in Eq. (8-8) may be positive, for inductive impedances, or negative, for capacitive impedances. This equation is also useful in determining the balance conditions because it tells us at a glance whether balance is possible at all.

Example 8-1 illustrates how Eqs. (8-7) and (8-8) may be used to determine if bridge balance is possible for a given set of bridge constants.

Example 8-1: The four impedances of the ac bridge of Fig. 8-1, given in terms of their magnitudes and phase angles, are:
$Z_1 = 200\ \Omega \angle 60°$ (inductive impedance)
$Z_2 = 400\ \Omega \angle -90°$ (pure capacitance)
$Z_3 = 300\ \Omega \angle 0°$ (pure resistance)
$Z_4 = 600\ \Omega \angle 30°$ (inductive impedance)
Determine if bridge balance is possible.

SOLUTION: The first condition for bridge balance, Eq. (8-7), requires that $Z_1 Z_4 = Z_2 Z_3$. Substituting the given values for the *magnitudes,* we obtain $200 \times 600 = 400 \times 300$, indicating that the first balance condition indeed is met. The second balance condition, expressed in Eq. (8-8), requires that the sums of the opposite phase angles are equal. Substituting the given values for the *phase angles,* we obtain

$$\angle\theta_1 + \angle\theta_4 = +60° + 30° = +90°$$
$$\angle\theta_2 + \angle\theta_3 = -90° + 0° = -90°$$

indicating that the second balance condition is not satisfied. The bridge, therefore, cannot be balanced.

In summary, then, two conditions must be met in balancing an ac bridge:

(a) The *products* of the *magnitudes* of the opposite arms must be equal.
(b) The *sum* of the *phase angles* of the opposite arms must be equal.

These two conditions will be applied to a study of the more common bridge circuits discussed in the following sections.

8-2 Inductance and Capacitance Comparison Bridges

In its most basic form the ac bridge can be used for the measurement of an unknown inductance or capacitance by comparison with a known inductance or capacitance. Figure 8-2 shows the capacitance comparison bridge. The

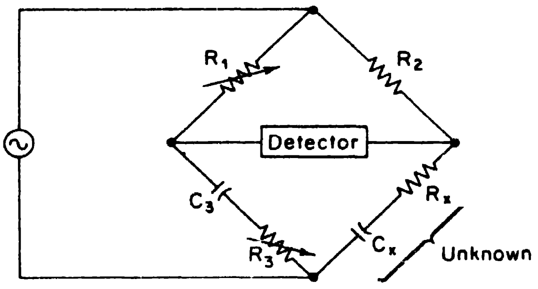


Figure 8-2 The capacitance comparison bridge.

ratio arms are resistive and are represented by R_1 and R_2. The known standard capacitor is represented by C_3 in series with resistance R_3, which represents the resistance inherent in the capacitor. Resistance R_3 usually also includes an *added* variable resistance needed to balance the bridge. The need for this added resistance is shown in the computation which follows.

R_x and C_x represent the unknown capacitance. R_x represents the small leakage resistance inherent in the unknown capacitor. Referring to the general bridge equation, expressed in Eq. (8-4a), we may write, for the impedances of the four arms:

$$\mathbf{Z}_1 = R_1; \quad \mathbf{Z}_2 = R_2; \quad \mathbf{Z}_3 = R_3 - \frac{j}{\omega C_3}; \quad \mathbf{Z}_4 = R_x - \frac{j}{\omega C_x}$$

Substituting these values in the balance equation (Eq. 8-4a), we obtain

$$R_1\left(R_x - \frac{j}{\omega C_x}\right) = R_2\left(R_3 - \frac{j}{\omega C_3}\right) \tag{8-9}$$

which reduces to

$$R_1R_x - R_1\frac{j}{\omega C_x} = R_2R_3 - R_2\frac{j}{\omega C_3} \tag{8-10}$$

Two complex numbers are equal when their real parts on both sides of the equation are equal and also when their imaginary parts on both sides are equal. Equating the real parts of Eq. (8-10), we obtain

$$R_1R_x = R_2R_3 \quad \text{or} \quad R_x = \frac{R_2R_3}{R_1} \tag{8-11}$$

Equating the imaginary terms of Eq. (8-10), we obtain

$$\frac{jR_1}{\omega C_x} = \frac{jR_2}{\omega C_3} \quad \text{or} \quad C_x = C_3\frac{R_1}{R_2} \tag{8-12}$$

Equations (8-11) and (8-12) show that two balance conditions must be met and that the two unknowns C_x and R_x are determined in terms of the known bridge arms.

To satisfy *both* balance conditions, the bridge must contain *two* variable elements in its configuration. Any two of the available four elements (R_1, R_2, R_3, C_3) could be chosen as the variables. In practice, capacitor C_3 is a *standard* capacitor of a fixed precision value and is therefore not available for adjustment.

Inspection of the two balance equations shows that resistance R_3 does not appear in the expression for C_x; to eliminate any interaction between the two balance controls, R_3 is an obvious choice as a variable element. Further, assume that R_1 is the second variable circuit element. Since we are measuring an unknown capacitor, where the resistive effects should be quite small, the first adjustment should be made for balance of the capacitive term; hence R_1 is adjusted for minimum sound in the headphones. The sound will not quite disappear, because the second balance condition has not yet been met. R_3 is now adjusted to decrease the sound in the phones still further; then R_1 is readjusted again. Alternate adjustments of the two resistors is necessary to give zero output in the headset and therefore *true* balance. The need for alternate adjustment becomes clear when we realize that any change in R_1 affects not only the capacitive balance equation (8-12), but also the resistive balance equation (8-11), because R_1 appears in both expressions.

The process of alternate manipulation of R_1 and R_3 is rather typical of the general balancing procedures for ac bridges. The alternate adjustment of the variables causes *convergence* of the balance point. The bridge is said to be *independent of the frequency* of the source of ac exictation since the frequency does not enter either of the balance equations. It is therefore relatively unimportant to know the power-supply frequency.

The *inductance comparison* bridge is very similar to the capacitance comparison bridge. The unknown inductance is determined in terms of a known, *standard* inductance. Figure 8-3 shows the schematic diagram of the inductance comparison bridge. The derivation of the balance equations follows essentially the same steps as for the capacitance bridge and will not be discussed in detail.

The inductive balance equation yields

$$L_x = L_3\frac{R_2}{R_1} \tag{8-13}$$

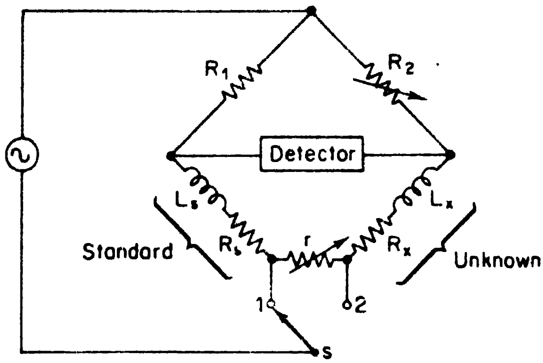


Figure 8-3
The inductance comparison bridge.

and the resistive balance equation gives

$$R_x = R_s\frac{R_2}{R_1} \tag{8-14}$$

In this bridge R_2 is chosen as the inductive balance control and R_s is the resistive balance control.

Some inductance bridges use a variable inductor as the standard-arm element. This variable inductor or *inductometer* consists of fixed and movable coils, which are so mounted that the total inductance changes with the position of the movable coils. One example of the variable inductor is the well-known *Brooks inductometer*.* Figure 8-3 shows a variable resistance r, which can be connected by switch S to either the standard (position 1) or the unknown (position 2). With switch S in position 1, the solution for R_x becomes

$$R_x = (R_s + r)\frac{R_2}{R_1} \tag{8-15}$$

and with the switch in position 2, the solution for R_x gives

$$R_x + r = R_s\frac{R_2}{R_1} \tag{8-16}$$

Since the resistive component of an inductor is usually much larger than that of a capacitor, the resistive adjustment becomes rather important and should be carried out first. Addition of the extra resistor r gives the option of extending the adjustment range for the resistive balance equation.

8-3 The Maxwell Bridge

The *Maxwell bridge* measures an unknown *inductance* in terms of a known *capacitance*. Use of a standard capacitor as the standard-arm element has the

*Cf. Stout, *op. cit.*, p. 357.

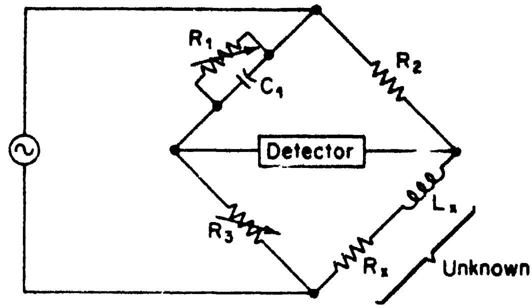


Figure 8-4
Schematic diagram of the Maxwell bridge for inductance measurements.

advantage of compactness and easy shielding of the standard component. (Figure 8-4 shows the schematic diagram of the Maxwell bridge.) Since one of the ratio arms has a resistance and a capacitance *in parallel,* it may now prove somewhat easier to compute the balance equations by using the ***admittance*** of arm 1, rather than its impedance.

Rearranging the general equation for bridge balance, as expressed in Eq. (8-4a), we obtain

$$\mathbf{Z}_x = \mathbf{Z}_2\mathbf{Z}_3\mathbf{Y}_1 \tag{8-17}$$

where $\mathbf{Y}_1$ is the admittance of arm 1. Reference to Fig. 8-4 shows that

$$\mathbf{Z}_2 = R_2; \qquad \mathbf{Z}_3 = R_3; \qquad \text{and} \quad \mathbf{Y}_1 = \frac{1}{R_1} + j\omega C_1$$

Substitution of these values in Eq. (8-17) gives

$$\mathbf{Z}_x = R_x + j\omega L_x = R_2R_3\frac{1}{R_1 + j\omega C_1} \tag{8-18}$$

Separation of the real and imaginary terms yields

$$R_x = \frac{R_2R_3}{R_1} \tag{8-19}$$

$$L_x = R_2R_3C_1 \tag{8-20}$$

where all resistances are expressed in ohms, inductance in henrys, and capacitance in farads.

The Maxwell bridge is limited to the measurement of *low-Q coils.* ($1 < Q < 10$). The measurement of high-Q coils demands that resistance R_1 become extremely large. This can be easily seen from the second balance condition (see Sec. 8-1) which states that the sum of the phase angles of one pair of opposite arms must be equal to the sum of the phase angles of the other pair. Since the phase angles of the resistive elements in arm 2 and arm 3 add up to 0°, the sum of the angles of arm 1 and arm 4 must also add up to 0°.

The phase angle of a high-Q coil will be very nearly 90° (positive), which requires that the phase angle of the capacitive arm must also be very nearly

90° (negative). This in turn means that resistance R_1 must have a very high value indeed, which is rather impractical. (The measurement of high-Q coils is usually performed on the Hay bridge, which is a modification of the Maxwell bridge (see Sec. 8-4)). The Maxwell bridge is also unsuited for the measurement of coils with a *very low value* of Q($Q < 1$). Q values of this magnitude occur in inductive resistors, for example, or in an RF coil if measured at low frequency. The bridge is unsuitable for low-Q measurements because of balance-convergence problems. As can be seen from the equations solving for R_x and L_x, resistors R_1 and R_3 both affect the expression for R_x, whereas R_3 appears only in the expression for L_x. Any adjustment for inductive balance upsets the resistive balance, therefore, and thus gives the effect known as *sliding balance.* (Interaction does not occur, when R_1 and C_1 are used for the balance adjustments, but a variable capacitor is not always suitable.) Sliding balance describes the interaction between controls, so that when we balance with R_1 and then with R_3, then go back to R_1, we find a new balance point. The balance point appears to move or *slide* toward its final point after many adjustments.

The usual procedure for balancing the Maxwell bridge is by first adjusting R_3 for inductive balancing and then varying R_1 to give resistive balance. Returning to the R_3 adjustment, we find that the resistive balance is being disturbed and moves to a new value. This process is repeated and gives *slow* convergence to final balance. For medium-Q coils, the resistance effect is not pronounced, and balance is reached after a few adjustments.

8-4 The Hay Bridge

The *Hay bridge* (Fig. 8-5) differs from the Maxwell bridge by having a resistance R_1 in series with the standard capacitor C_1 instead of in parallel. It is immediately apparent that, for large phase angles, resistance R_1 should have a very low value. The Hay circuit is therefore more convenient for measuring high-Q coils.

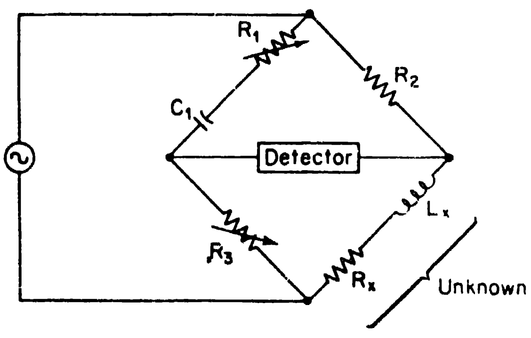


Figure 8-5
The Hay bridge for inductance measurements.

The balance equations are again derived by substituting the values of the impedances of the bridge arms into the general equation for bridge balance, Eq. (8-4a). For the circuit shown in Fig. 8-5, we find

$$\mathbf{Z}_1 = R_1 - \frac{j}{\omega C_1}; \qquad \mathbf{Z}_2 = R_2; \qquad \mathbf{Z}_3 = R_3; \qquad \mathbf{Z}_x = R_x + j\omega L_x$$

Substituting these values in Eq. (8-4a), we get

$$\left(R_1 - \frac{j}{\omega C_1}\right)(R_x + j\omega L_x) = R_2 R_3 \tag{8-21}$$

which expands to

$$R_1 R_x + \frac{L_x}{C_1} - \frac{jR_x}{\omega C_1} + j\omega L_x R_1 = R_2 R_3 \tag{8-21}$$

Separating the real and imaginary terms, we obtain

$$R_1 R_x + \frac{L_x}{C_1} = R_2 R_3 \tag{8-22}$$

$$\frac{R_x}{\omega C_1} = \omega L_x R_1 \tag{8-23}$$

Both Eq. (8-22) and Eq. (8-23) contain L_x and R_x and we must solve these equations simultaneously. This yields

$$R_x = \frac{\omega^2 C_1^2 R_1 R_2 R_3}{1 + \omega^2 C_1^2 R_1^2} \tag{8-24}$$

$$L_x = \frac{R_2 R_3 C_1}{1 + \omega^2 C_1^2 R_1^2} \tag{8-25}$$

These expressions for the unknown inductance and resistance both contain the angular velocity, and it therefore appears that the frequency of the driving source must be known accurately. That this is not true when a high-Q coil is being measured follows from the following considerations: Remembering that the sum of the opposite sets of phase angles must be equal, we find that the inductive phase angle must be equal to the capacitive phase angle, since the resistive angles are zero. Figure 8-6(left) shows that the tangent of the inductive phase angle equals

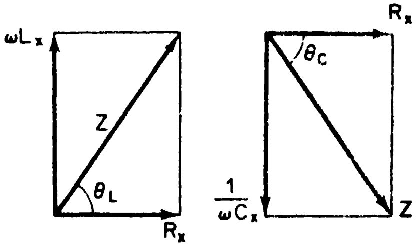


Figure 8-6 Illustrating the inductive and capacitive phase angles by using impedance triangles.

$$\tan\theta_L = \frac{X_L}{R} = \frac{\omega L_x}{R_x} = Q \tag{8-26}$$

The tangent of the capacitve phase angle is shown to be, referring to Fig. 8-6(b),

$$\tan \theta_C = \frac{X_C}{R} = \frac{1}{\omega C_1 R_1} \tag{8-27}$$

When the two phase angles are equal, their tangents are also equal and we can write

$$\tan \theta_L = \tan \theta_C \quad \text{or} \quad Q = \frac{1}{\omega C_1 R_1} \tag{8-28}$$

Returning now to the term $(1 + \omega^2 C_1^2 R_1^2)$ which appears in Eq. (8-24) and Eq. (8-25), we find that, after substituting Eq. (8-28) in the expression for L_x, Eq. (8-25), this reduces to

$$L_x = \frac{R_2 R_3 C_1}{1 + (1/Q)^2} \tag{8-29}$$

For a value of Q greater than ten the term $(1/Q)^2$ will be smaller than $\frac{1}{100}$ and can be neglected. Equation (8-25) therefore reduces to the expression derived for the Maxwell bridge,

$$L_x = R_2 R_3 C_1$$

The Hay bridge is suited for the measurement of high-Q inductors, especially for those inductors having a Q greater than ten. For Q values smaller than ten, the term $(1/Q)^2$ becomes rather important and cannot be neglected. In this case the Maxwell bridge is more suitable.

8-5 The Schering Bridge

The *Schering bridge,* one of the most important ac bridges, is used extensively for the measurement of *capacitors.* It offers some decided advantages over the capacitance comparison bridge, discussed earlier in this chapter (Sec. 8-2). The Schering bridge is used for capacitance measurements in a general sense, but more particularly in measurement of insulating properties, i.e., for phase angles very nearly 90°.

The basic circuit arrangement is shown in Fig. 8-7, and inspection of the circuit shows a strong resemblance to the comparison bridge. Notice that arm 1 now contains a parallel combination of a resistor and a capacitor and the standard arm contains only a capacitor. The standard capaictor is usually a high-quality mica capacitor for general measurement work or an air capacitor for insulation measurements. A good-quality mica capacitor has very low losses (no resistance) and therefore a phase angle of approximately 90°. An air capacitor, when designed carefully, has a very stable value and a very small electric

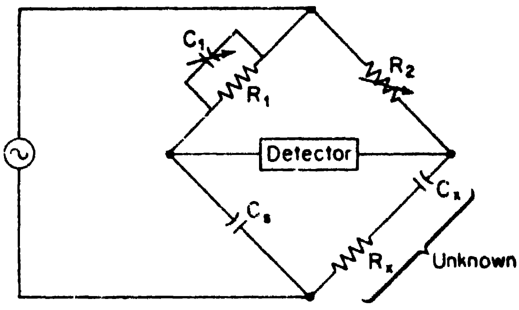


Figure 8-7
The Schering bridge for the measurement of capacitance.

field; the insulating material to be tested can easily be kept out of any strong fields.

The balance conditions require that the sum of the phase angles of arms 1 and 4 equal the sum of the phase angles of arms 2 and 3. Since the standard capacitor is in arm 3, the sum of the phase angles of arm 2 and arm 3 will be $0° + 90° = 90°$. In order to obtain the 90° phase angle needed for balance, the sum of the angles of arm 1 and arm 4 must equal 90°. Since in general measurement work the unknown will have a phase angle smaller than 90°, it is necessary to give arm 1 a small capacitive angle by connecting capacitor C_1 in parallel with resistor R_1. A small capacitive angle is very easy to obtain, requiring a small capacitor across resistor R_1.

The balance equations are derived in the usual manner, and by substituting the corresponding impedance and admittance values in the general equation, we obtain

$$\mathbf{Z}_x = \mathbf{Z}_2\mathbf{Z}_3\mathbf{Y}_1 \qquad \text{or} \qquad R_x - \frac{j}{\omega C_x} = R_2\left(\frac{-j}{\omega C_s}\right)\left(\frac{1}{R_1} + j\omega C_1\right)$$

$$R_x - \frac{j}{\omega C_x} = \frac{R_2 C_1}{C_s} - \frac{jR_2}{\omega C_s R_1} \tag{8-30}$$

Equating the real terms and the imaginary terms, we obtain the results:

$$R_x = R_2\frac{C_1}{C_s} \tag{8-31}$$

$$C_x = C_s\frac{R_1}{R_2} \tag{8-32}$$

As can be seen from the circuit diagram of Fig. 8-7, the two variables chosen for the balance adjustment are capacitor C_1 and resistor R_2. There seems to be nothing unusual about the balance equations or the choice of variable components, but consider for a moment how the quality of a capacitor is defined.

The *power factor* (PF) of a series RC combination is defined as the cosine of the phase angle of the circuit. Therefore, the PF of the unknown equals

PF $= R_x/Z_x$. For phase angles very close to 90°, the reactance is almost equal to the impedance and we can approximate the power factor to

$$\text{PF} \simeq \frac{R_x}{X_x} = \omega C_x R_x. \tag{8-33}$$

The *dissipation factor* of a series RC circuit is defined as the cotangent of the phase angle and therefore, by definition, the dissipation factor

$$D = \frac{R_x}{X_x} = \omega C_x R_x \tag{8-34}$$

Incidentally, since the quality of a coil is defined by $Q = X_L/R_L$, we find that the dissipation factor, D, is the reciprocal of the quality factor, Q, and therefore $D = 1/Q$. The dissipation factor tells us something about the quality of a capacitor; i.e., how close the phase angle of the capacitor is to the ideal value of 90°. By substituting the value of C_x in Eq. (8-32) and of R_x in Eq. (8-31) in the expression for the dissipation factor, we obtain:

$$D = \omega R_1 C_1 \tag{8-35}$$

If the resistor R_1 in the Schering bridge of Fig. 8-7 has a fixed value, the dial of capacitor C_1 may be calibrated directly in dissipation factor D. This is the usual practice in a Schering bridge. Notice that the term ω appears in the expression for the dissipation factor (Eq. 8-35). This means, of course, that the calibration of the C_1 dial holds for only one particular frequency at which the dial is calibrated. A different frequency can be used, provided that a correction is made by multiplying the C_1 dial reading by the ratio of the two frequencies.

8-6 The RC Frequency Bridge or Wien Bridge

The *Wien bridge* is primarily known as a *frequency-determining* device and is presented here not only for its use as an ac bridge to *measure* frequency, but also for its application in various other useful circuits. We find, for example, a Wien bridge in the harmonic distortion analyzer, where it is used as a *notch filter,* discriminating against one specific frequency. The Wien bridge also finds application in audio-and HF oscillators as the *frequency-determining* element. The Wien bridge is therefore discussed here in its basic form, designed to measure frequency, and in other chapters, as an element of different types of instruments.

The Wien bridge has a series RC combination in one arm and a parallel RC combination in the adjoining arm (see Fig. 8-8). The impedance of arm 1 is $\mathbf{Z}_1 = R_1 - j/\omega C_1$. The admittance of arm 3 is $Y_3 = 1/R_3 + j\omega C_3$. Using the basic equation for bridge balance and substituting the appropriate values, we obtain

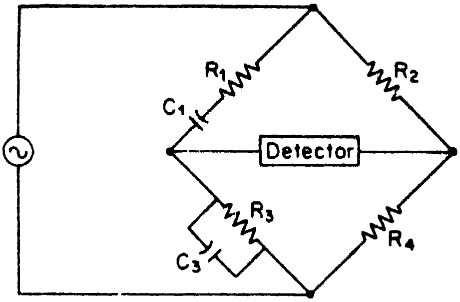


Figure 8-8
Frequency measurement with the RC frequency or Wien bridge.

$$R_2 = \left(R_1 - \frac{j}{\omega C_1}\right)R_4\left(\frac{1}{R_3} + j\omega C_3\right) \tag{8-36}$$

Expanding this expression, we get

$$R_2 = \frac{R_1R_4}{R_3} + (j\omega C_3R_1R_4) - \frac{jR_4}{\omega C_1R_3} + \frac{R_4C_3}{C_1} \tag{8-37}$$

Equating the *real* terms, we obtain

$$R_2 = \frac{R_1R_4}{R_3} + \frac{R_4C_3}{C_1} \tag{8-38}$$

which reduces to

$$\frac{R_2}{R_4} = \frac{R_1}{R_3} + \frac{C_3}{C_1} \tag{8-39}$$

Equating the *imaginary* terms, we obtain

$$C_3R_1R_4 = \frac{R_4}{\omega C_1R_3} \tag{8-40}$$

where $\omega = 2\pi f$

and we obtain an expression for f:

$$f = \frac{1}{2\pi\sqrt{C_1C_3R_1R_3}} \tag{8-41}$$

Notice that the two conditions for bridge balance now result in an expression determining the required resistance ratio, R_2/R_4, and another expression determining the frequency of the applied voltage. In other words, if we satisfy Eq. (8-39) and also excite the bridge with a frequency described by Eq. (8-41), the bridge will be in balance.

In most Wien bridge circuits, the components are chosen such that $R_1 = R_3$ and $C_1 = C_3$. This reduces Eq. (8-39) to $R_2/R_4 = 2$ and Eq. (8-41) to

$$f = \frac{1}{2\pi RC} \tag{8-42}$$

which is the general expression for the frequency of the Wien bridge. In a practical bridge, capacitors C_1 and C_3 are fixed capacitors and resistors R_1 and R_3

are variable resistors controlled by a common shaft. Provided now that $R_3 = 2R_4$, the bridge may be used as a frequency-determining device, balanced by a single control. This control may be calibrated directly in terms of frequency.

Because of its frequency sensitivity, the Wien bridge may be difficult to balance (unless the waveform of the applied voltage is purely sinusoidal). The bridge is *not* balanced for any harmonics present in the applied voltage, so that these harmonics will sometimes produce an output voltage masking the true balance point.

8-7 The Wagner Ground Connection

The discussion so far has assumed that the four bridge arms consist of simple lumped impedances which do not interact in any way. In practice, however, *stray capacitances* exist between the various bridge elements and ground and also between the bridge arms themselves. These stray capacitances shunt the bridge arms and cause errors in the measurement, particularly at the higher frequencies, or when small capacitors or large inductors are measured. One way to control stray capacitances is by shielding the arms and connecting the shields to ground. This does not eliminate the capacitances, but at least makes them constant in value and they can therefore be compensated.

One of the most widely used methods for eliminating some of the effects of stray capacitance in a bridge circuit is the *Wagner ground connection*. This circuit eliminates the troublesome capacitance which exists between the detector terminals and ground. Figure 8-9 shows the circuit of a capacitance bridge,

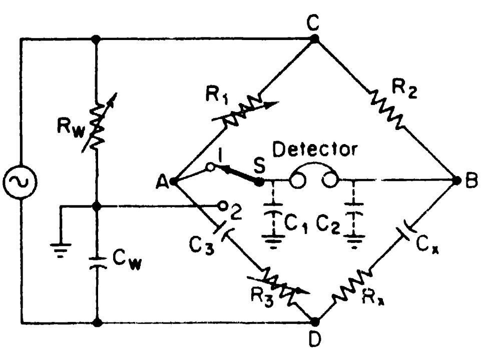


Figure 8-9
The Wagner ground connection, eliminating the effects of stray capacitances across the detector.

where capacitors C_1 and C_2 represent these stray capacitances. The oscillator is removed from its usual ground connection and bridged by a series combination of resistor R_w and capacitor C_w. The junction of R_w and C_w is grounded and called the *Wagner earth connection*. The procedure for initial adjustment of the bridge is as follows: The detector is connected to point 1 and R_1 is adjusted for *null* or minimum sound in the headphones. The switch is then thrown to position 2, which connects the detector to the Wagner ground point. Resistor R_w is now adjusted for minimum sound. When the switch is thrown to position 1 again, some unbalance probably will be shown. Resistors R_1 and R_3 are then adjusted for minimum detector response, and the switch is again thrown to position 2. A few adjustments of R_w and R_1 (and R_3) may be necessary before a null is reached on both switch positions. When null is finally obtained, points 1 and 2 are at the same potential, and this is ground potential. The stray capacitances C_1 and C_2 are then effectively shorted out and have no effect on normal bridge balance. There are also capacitances from points C and D to ground, but the addition of the Wagner ground point eliminates them from the detector circuit, since current through these capacitances will enter through the Wagner ground connection.

The capacitances across the bridge arms are not eliminated by the Wagner ground connection and they will still affect the accuracy of the measurement. The idea of the Wagner ground can also be applied to other bridges, as long as care is taken that the grounding arms duplicate the impedance of one pair of bridge arms across which they are connected. Since the addition of the Wagner ground connection does not affect the balance conditions, the procedure for measurement remains unaltered.

8-8 Shielding of Bridge Elements

Capacitances between the bridge elements and between the elements and ground affect the accuracy of ac bridge measurements. An effective way of controlling these capacitances consists of enclosing the element in a *shield*, thereby placing the capacitance at a position in the circuit where it can do no harm. Figure 8-10 shows how control of stray capacitance in resistors is accomplished by shielding. The stray capacitance across the resistor and from each end of the resistor to ground is shown by the dashed portions of Fig. 8-10(a). A grounded shield, which surrounds the resistor and is connected to one terminal only—see Fig. 8-10(b)— effectively places the stray capacitance between that terminal and ground. There is still capacitance between resistor and shield, but this capacitance is of definite value and does not change. The equivalent circuit of the resistor with its surrounding shield is shown in Fig. 8-10(c) and shows C_{AB} between the resistor terminals and C_{BG} from terminal B to ground.

Two resistors in series are often found in a bridge configuration. In this case, each resistor is shielded and the shields are connected to the common point between the resistors, as indicated in Fig. 8-10(e). This point is then grounded and results in the equivalent circuit of Fig. 8-10(f). C_{AB} and C_{AC}

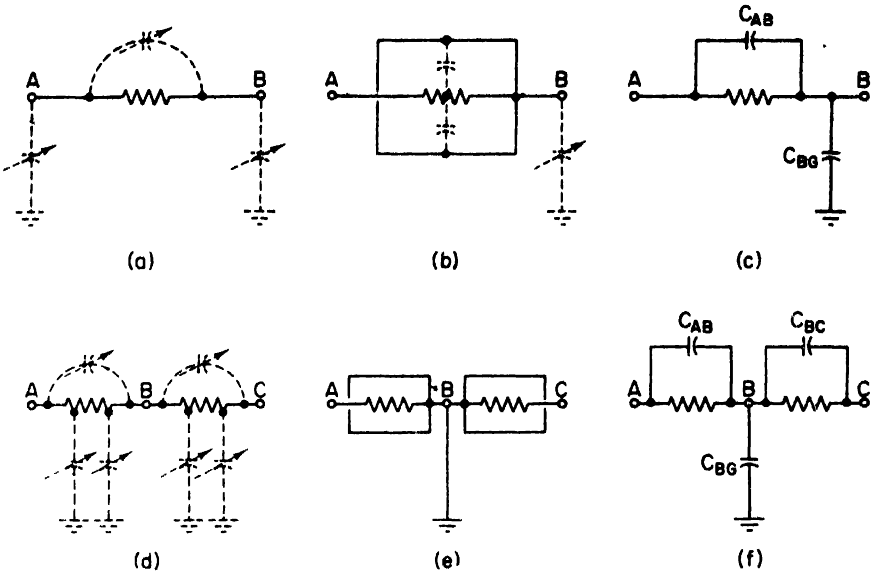


Figure 8-10
The shielding of resistors to reduce the effect of stray capacitances.

represent the capacitances within the shield, which are fixed in value, and C_{BG} is the capacitance from the junction of the two resistors to ground. Notice that the stray capacitances are now concentrated at the terminals of the resistors. This means that for a resistive bridge circuit, the four corners of the bridge exhibit stray capacitance to ground, but there is no longer any stray capacitance between the bridge elements themselves. Of course, the resistors are no longer purely resistive, having the shunting capacitance within the shield across them, but these capacitances are fixed in value and the bridge can therefore be balanced without any problem.

If both terminals of a capacitor are insulated from ground, we have the condition shown in Fig. 8-11(a), where C_{AG} and C_{BG} represent the stray capacitances to ground. The equivalent circuit is shown in Fig. 8-11(b), and we see that we have a delta connection of three capacitors, sometimes called a *three-terminal capacitor*. When the capacitance is surrounded by a shield which is connected to ground, we have the situation as shown in Fig. 8-11(c). The equivalent circuit of this arrangement is shown in Fig. 8-11(d) and we have again a delta connection of three capacitors. In this case, however, C_{AS} and C_{BS} are definite in amount and do not change owing to external effects. When the shield is connected to terminal B of the capacitor, the capacitance from B to shield is short-circuited and the capacitance from A to shield is added to the main capacitance C_{AB}, as shown in Fig. 8-11(e). If point B may be grounded, the

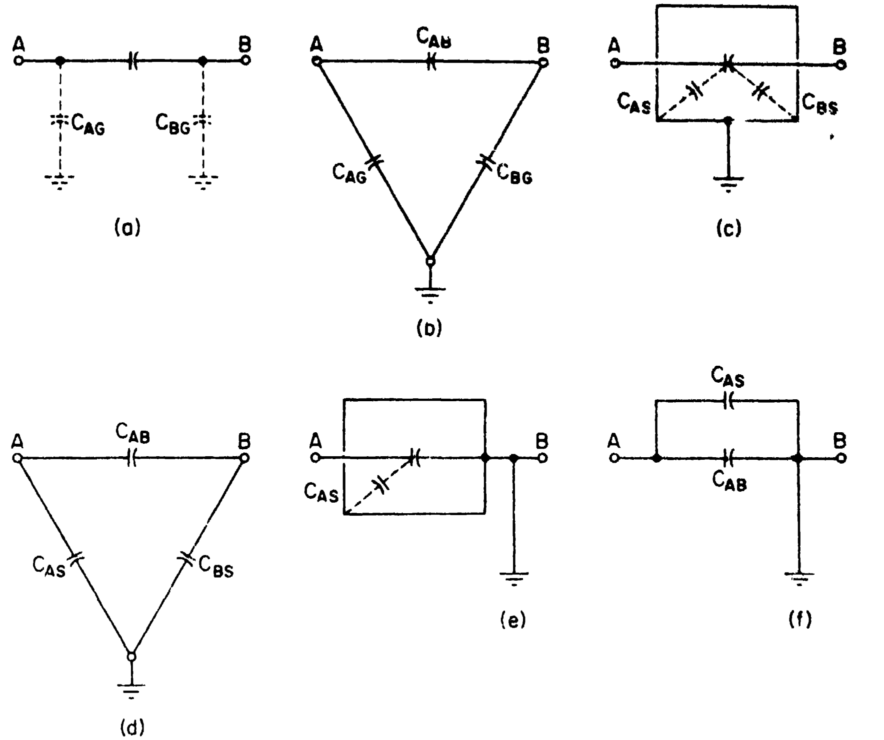


Figure 8-11

Stray capacitances in capacitors, illustrating the effect of shielding and the resultant equivalent circuits.

capacitance from shield to ground is short-circuited and we have a two-terminal capacitor, as shown in the equivalent circuit of Fig. 8-11(f).

Figure 8-12(a) shows an interesting application of shielding and is used in the testing of samples of insulating material in sheet form. A metal foil (A) is connected to one side of the sample of insulating material. The opposite side of the sample has a circular piece of metal foil (B) attached to it, which is surrounded by a guard-ring electrode, also consisting of metal foil. The guard-ring insures uniform capacitance between it and the center foil. The equivalent circuit of this arrangement is a *three-terminal capacitance,* shown in Fig. 8-12(b). It is desired to eliminate the effects of the guard-ring capacitances in the bridge circuit and to measure only the capacitance C_{AB} of the insulating material. The bridge should therefore be so arranged that the guard-ring has the same potential as the upper foil (B).

Before discussing measurement of a three-terminal capacitance, consider the effects of stray capacitances of a bridge circuit in general. Figure 8-13 shows stray capacitances existing in a bridge circuit. Capacitance C_1 is connected from

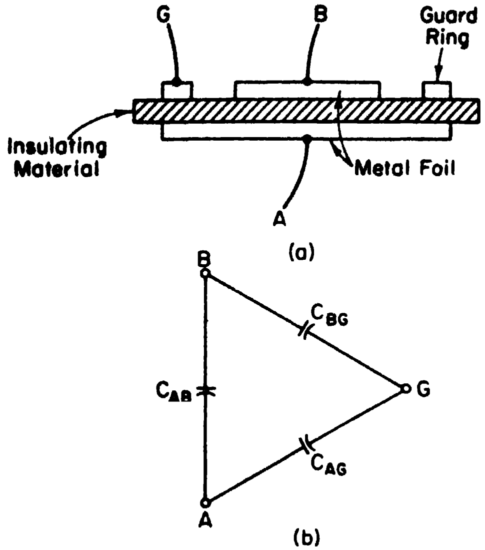


Figure 8-12
Use of a guard circuit for measuring samples of insulating material. (a) Placement of the guard ring. (b) Equivalent circuit showing the fixed capacitances between the contact foil and the guard ring (C_{BG} and C_{AG}).

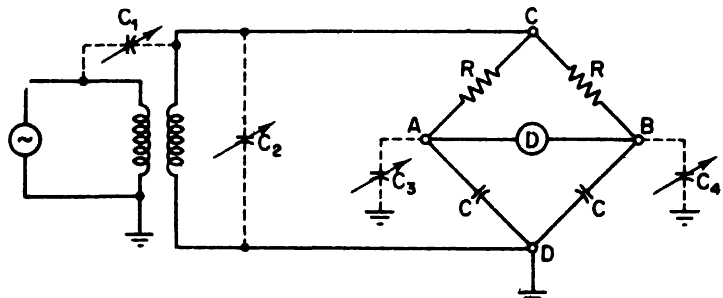


Figure 8-13
Showing the stray capacitances existing in an ac bridge.

point C to ground through the primary winding of the output transformer of the oscillator. Capacitor C_2 represents the capacitance from the top corner of the bridge to ground. These two capacitors shunt the bridge input but have no effect on the balance of the bridge. Capacitor C_3 shunts the bridge capacitance, C_{AD}, and stray capacitance C_4 shunts the bridge arm, C_{BD}. These capacitances do affect the bridge balance, because they are not definite in value and change according to external circumstances. Figure 8-14(a) shows capacitor C_{AD} shielded, with the shield connected to ground. If the capacitor C_{AD} is now calibrated within its shield, this calibration will give the total capacitance between points A and D. The unknown capacitance, C_{BD}, is enclosed within a shield which is connected to point E of a shunting circuit R_s and C_s. This circuit is the same as the shunting circuit of the Wagner ground connection of Fig. 8-9

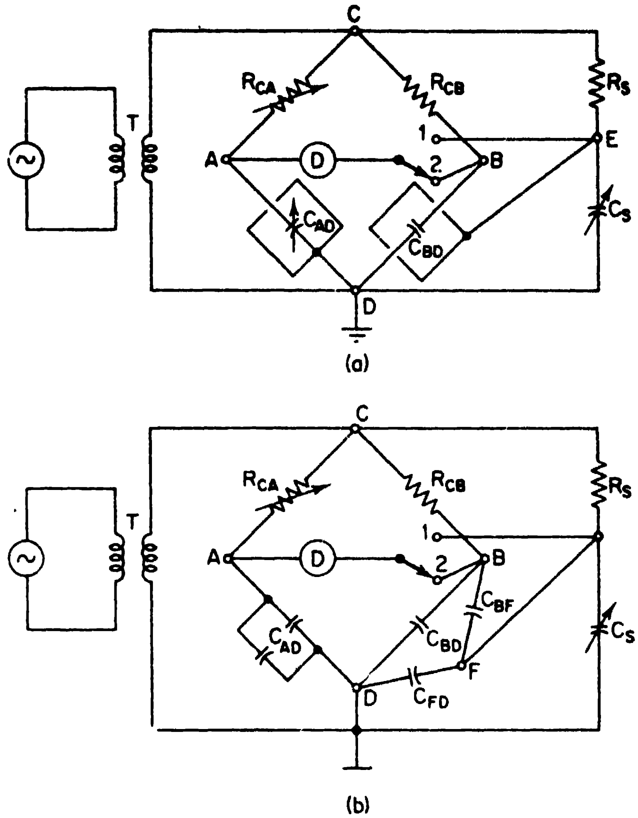


Figure 8-14
(a) Shielding of the capacitors in a capacitance bridge. (b) Equivalent circuit of the bridge.

except that point E is not connected to ground. The bridge is balanced by alternate adjustments of arms AC and AD and the shunting circuit, until there is no deflection of the detector when the switch is in either position 1 or 2. This means that point B is at the same ponteial as point E. If the circuit is redrawn into its equivalent circuit, replacing the shielded capacitors as shown in Fig. 8-14(b), we find that point F of the three-terminal capacitance C_{BD} is at the same potential as point B. This means that capacitor C_{BF} has no effect on the circuit. Capacitor C_{FD} shunts capacitor C_{BD} and, again, has no effect on bridge balance. Therefore, the unknown capacitance, C_{BD}, can be measured without the effects of stray capacitance.

Shielding is effective in preserving the *true magnitude* of the components of a bridge. It is also important in keeping *noise* out of the measuring system. Both magnetic and electrostatic shielding are necessary to keep stray electromagnetic and electrostatic fields from generating unwanted emf's in the bridge circuits. Be careful to shield the leads from an oscillator to a bridge and from the bridge to the detector. Be careful to provide only one ground connection to an ac bridge; otherwise groundloop currents may produce serious noise in the system, resulting in measurement inaccuracy.

8-9 The Universal Impedance Bridge

One of the most useful and versatile laboratory bridges is the *universal impedance bridge*. This bridge combines several of the bridge configurations discussed so far into a single instrument capable of measuring both dc and ac resistance, the inductance and storage factor, Q, of an inductor, and the capacitance and dissipation factor, D, of a capacitor. A representative example of a universal impedance bridge is given in the photograph of Fig. 8-15 which clearly shows the arrangement of the various front panel controls.

The universal bridge consists of four basic bridge circuits, together with

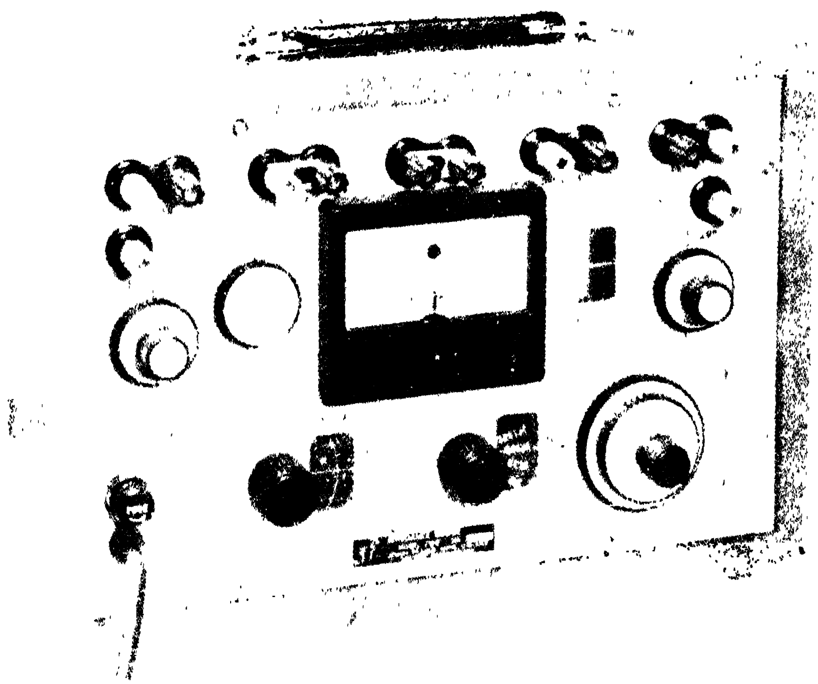


Figure 8-15
A modern universal impedance bridge, the John Fluke Model 710B.
(*Courtesy John Fluke Manufacturing Company.*)

suitable switching, ac and dc detectors, ac and dc generators, and impedance standards. The Wheatstone bridge circuit is used for both *ac and dc resistance measurements*. *Capacitance* is measured in terms of a standard capacitor and precision resistors in a four-arm network, with means for determining the *losses* in the unknown capacitor. The Maxwell configuration is used for low-Q *inductor measurements* and the Hay bridge for measurements of inductors with a Q above ten. For dc resistance measurements a suspension galvanometer with a

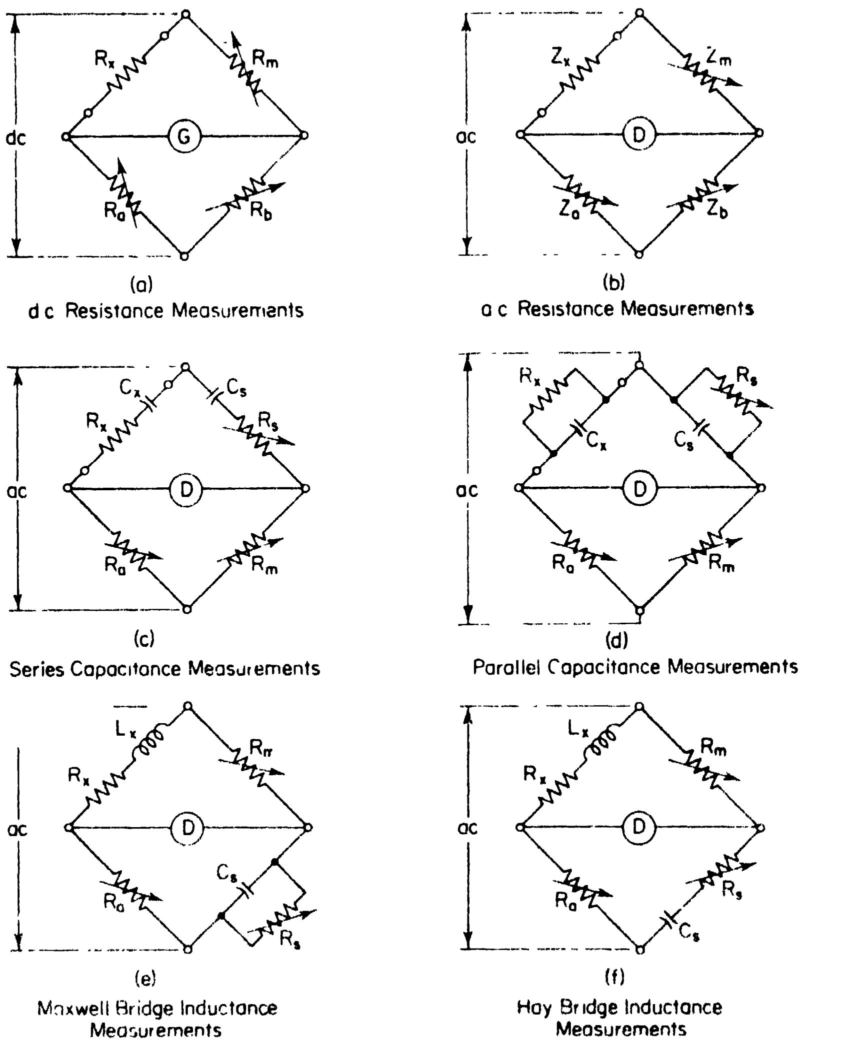


Figure 8-16

The various bridge configurations of the universal impedance bridge of Fig. 8-15.

current sensitivity of 0.5 μA per division is used. A selective amplifier operating an electron ray tube is usually used as a null indicator for all ac measurements. Terminals are provided for connection of external ac and dc null detectors. High-impedance headphones may also be connected and used as an ac detector. The dc generator is a simple dc power supply. The ac generator consists of an oscillator using plug-in RC networks for frequency selection with a frequency of 10 kHz as the standard frequency.

Figure 8-16 shows the different bridge configurations used in this impedance bridge. It will be found that most general-purpose bridges incorporate the same ideas as the instrument described here, but always refer to the appropriate manufacturer's manual accompanying each bridge.

References

1. Stout, Melville B., *Basic Electrical Measurements,* 2nd ed., Chap. 9, 10. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

2. Frank, Ernest, *Electrical Measurement Analysis*, Chap. 9, 13. New York: McGraw-Hill Book Company, Inc., 1959.

Problems

1. A balanced ac bridge has the following constants: arm AB, $R = 2{,}000\ \Omega$ in parallel with $C = 0.047\ \mu F$; BC, $R = 1{,}000\ \Omega$ in series with $C = 0.47\ \mu F$; CD, unknown; DA, $C = 0.5\ \mu F$. The frequency of the oscillator is 1,000 Hz. Find the constants of arm CD.

2. A bridge is balanced at 1,000 Hz and the following constants: AB, 0.2 μF pure capacitance; BC, 500 Ω pure resistance; CD, the unknown; DA, $R = 300\ \Omega$ in parallel with $C = 0.1\ \mu F$. Find the R and C or L constants of arm CD, considered as a series circuit.

3. A 1,000-Hz bridge has the following constants: arm AB, $R = 1{,}000\ \Omega$ in parallel with $C = 0.5\ \mu F$; BC, $R = 1{,}000\ \Omega$ in series with $C = 0.5\ \mu F$; CD, $L = 30$ mH in series with $R = 200\ \Omega$. 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 also as a pure R in parallel with a pure C or L.

4. An ac bridge has in arm AB a pure capacitance of 0.2 μF; in arm BC, a pure resistance of 500 Ω; in arm CD, a series combination of $R = 50\ \Omega$ and $L = 0.1$ H. Arm DA consists of a capacitor $C = 0.4\ \mu F$ in series with a variable resistor R_s. $\omega = 5{,}000$ rad/s (a) Find the value of R_s to

obtain bridge balance. (b) Can complete balance be attained by adjustment of R_s? If not, specify the position and the value of an adjustable resistance to complete the balance.

5. A balanced ac bridge has the following constants: AB, R = 500 Ω; BC, R = 1,000 Ω; CD, the unknown; DA, $C = 0.2\ \mu$F. A voltage of 10 V at 1,000 Hz is impressed on the bridge at points A and C. (a) Find the constants of the unknown. (b) The 1,000-Ω resistor is changed to 1,002 Ω. Find the voltage across the high-impedance detector. (c) Repeat (b), with the detector and the generator interchanged.

6. An unbalanced ac bridge has the following constants: arm AB, R = 2,000 Ω in parallel with $C = 0.2\ \mu$F; BC, $R = 1{,}500$ Ω; CD, $L = 0.8$ H in series with R = 500 Ω; DA, R = 2,000 Ω. The oscillator has an output of 20 V and is connected to A and C. The frequency is 1,000 Hz. (a) What would the values of the constants in arm CD have to be for bridge balance? (b) For the original bridge constants given in this problem, find the voltage across the high-impedance detector.

7. A bridge is balanced at 1,000 Hz and has pure resistance ratio arms, AB 1,500 Ω, BC 1,000 Ω. The unknown is connected from C to D. Arm DA has a standard capacitor of 0.1 μF and negligible internal resistance, to which is added a series resistance of 10 Ω to give balance. The generator has an output of 15 V and is connected from B to D. The detector is a high-impedance voltmeter. (a) Find the constants of arm CD. (b) Find the detector voltage for an increase of 10 Ω in arm BC.

8. In the ac bridge of Fig. 8-2, $R_1 = 521$ Ω, $R_2 = 1{,}200$ Ω, $C_3 = 0.045$ μF, and $R_3 = 12.1$ Ω. The frequency of the oscillator is 10 kHz. (a) Determine the values of R_x and C_x. (b) It is found that R_1 has 2 μH series inductance and 550 pF shunt capacitance, R_2 has 5 μH series inductance and 1,050 pF shunt capacitance, C_3 has 1.5 MΩ shunt resistance, and R_3 is unchanged. Determine the error in measuring R_x and C_x as in part (a).

9. An ac bridge has the following constants: arm AB, R = 1,000 Ω in parallel with $C = 0.159\ \mu$F; BC, R = 1,000 Ω; CD, $R = 500$ Ω; DA, $C = 0.636\ \mu$F in series with an unknown resistance. Find the frequency for which this bridge is in balance and determine the value of the resistance in arm DA to produce this balance.

10. An ac bridge has the following constants: arm AB, $R = 800$ Ω in parallel with $C = 0.4\ \mu$F; BC, R = 500 Ω in series with $C = 1.0\ \mu$F; CD, R = 1,200 Ω; DA, pure R of unknown value. (a) Find the frequency for which the bridge is in balance. (b) Find the resistance required in arm DA to produce balance.

11. An ac bridge has pure resistances in three of its arms: R_1 in arm AB, R_2 in arm BC, and R_3 in arm DA. Arm CD consists of a coil with series components of R and L, shunted by a variable capacitance, C. Derive the

balance equations for this bridge, measuring the constants of the coil in terms of the other components. Express the Q of the coil in terms of bridge-arm constants at balance.

12. An ac bridge, marked ABCD around the corners, has the following constants: AB, pure capacitance 0.01 μF; BC, pure resistance 2,500 Ω; CD, the unknown; DA, capacitance 0.02 μF in series with a resistance of 7,500 Ω. The bridge is balanced at a frequency such that $\omega = 50{,}000$ rad/s. (a) Find the unknown for the bridge constants stated. (b) There is a stray capacitance of 100 pF across arm DA, in addition to the constants above. Find the true value of the unknown.

NINE

THE OSCILLOSCOPE (CRO)

9-1 Introduction

The *cathode ray oscilloscope* (hereafter CRO) is an extremely useful and versatile laboratory instrument used for measurement and analysis of waveforms and other phenomena in electronic circuits. CROs are basically very fast *X*-*Y* plotters, displaying an input signal versus another signal or versus time. The "stylus" of this "plotter" is a luminous spot which moves over the display area in response to input voltages.

In the usual CRO application, the X axis or horizontal input is an internally generated linear ramp voltage or time base, which moves the luminous spot periodically from left to right over the display area or screen. The voltage under examination is applied to the Y axis or vertical input of the CRO, moving the spot up and down in accordance with the instantaneous value of the input voltage. The spot then traces a pattern on the screen which shows the input voltage variation as a function of time. When the input voltage is repetitive at a fast enough rate, the display appears as a stationary pattern on the screen. The CRO therefore provides a means of observing time-varying voltages.

In addition to voltages, the CRO can present visual representations of many *dynamic phenomena* by means of transducers which convert current, pressure, strain, temperature, acceleration, and many other physical quantities into voltages.

CROs are also used to investigate waveforms, transient happenings, and other time-varying quantities from the very low-frequency range to the very high frequencies. Recordings of these happenings can be made by specially developed CRO cameras or strip-chart recorders for quantitative interpretation.

The principles upon which the CRO operates are discussed in detail in the following sections.

9-2 The Cathode Ray Tube (CRT)

The *cathode ray tube* (CRT) is the heart of the CRO, with the rest of the instrument consisting of circuitry necessary to operate the CRT. Figure 9-1 shows a schematic view of a CRT. The main components of the CRT shown in Fig. 9-1 are

(a) Electron gun assembly.
(b) Deflection plate assembly.
(c) Fluorescent screen.
(d) Glass envelope or bottle.
(e) Base through which electrical connections are made to the various components of the CRT.

Before presenting a detailed discussion of the construction and operation of the CRT, a short summary of its operation is appropriate:

The *electron gun* assembly produces a sharply focused beam of electrons, accelerated to a very high velocity. This focused beam of electrons strikes a small area of the *fluorescent screen* with enough energy to cause the screen to light up at one spot. On leaving the electron gun, the beam passes between two pairs of *electrostatic deflection plates*. Voltages applied to these plates deflect the beam: voltages on one pair of plates move the beam vertically up and down; voltages applied to the other pair of plates move the beam horizontally from side to side. These movements are *independent* of each other so that the beam may be positioned anywhere on the screen by appropriate horizontal and vertical voltage inputs.

In order to understand the function and operation of the many different controls found on the CRO front panel, it is necessary to study the CRT more closely. The following sections describe the construction and the fundamentals of operation of the various parts of the CRT.

9-3 The Electron Gun

The source of the focused and accelerated electron beam is the *electron gun*. The name is derived from the analogy between the motion of the electron emit-

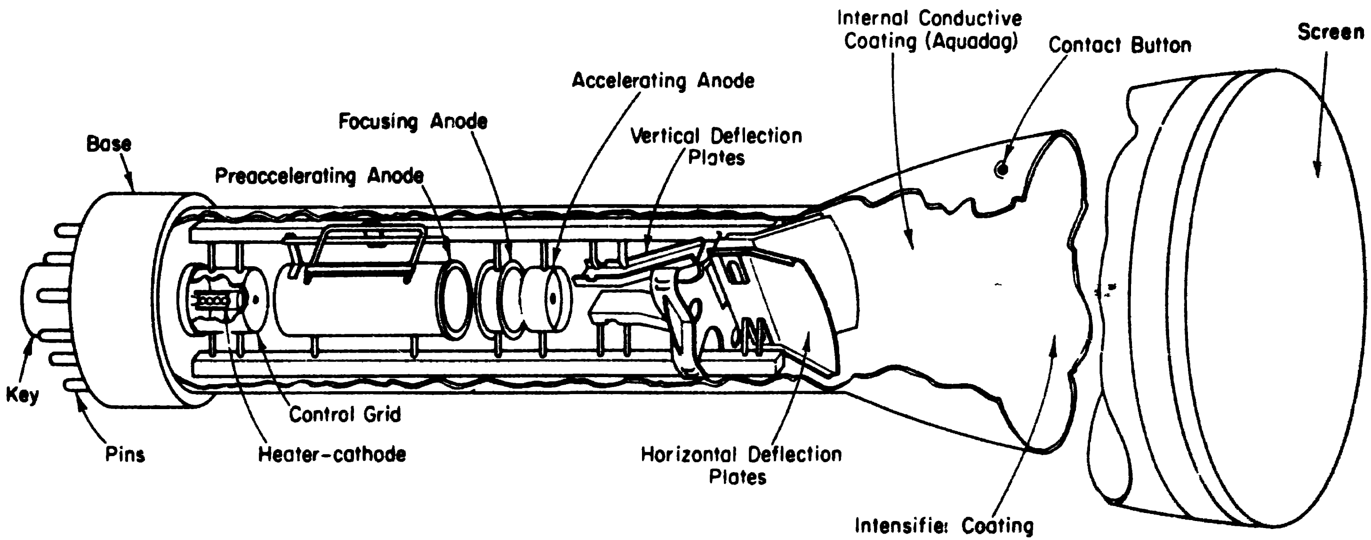


Figure 9-1. Internal structure of a cathode ray tube.

ted from the gun structure and the travel path of a bullet fired from a gun. In fact, the study of the motion of charged particles (for instance, electrons) in an electric field is called *electron ballistics*. A photograph of a conventional general-purpose electron gun is shown in Fig. 9-2.

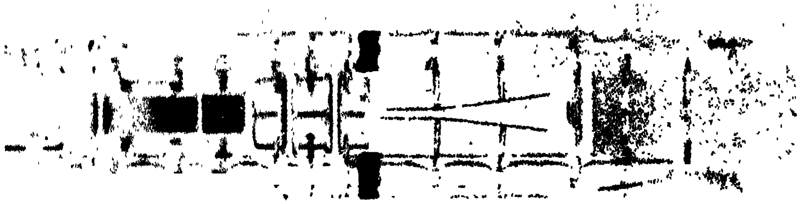


Figure 9-2
The electron gun and deflection plate assembly of a general-purpose CRT. (*Courtesy Tektronix, Inc.*)

Electrons are emitted from the indirectly heated *thermionic cathode*, a, and pass through a small hole in the *control grid*, b. This control grid is usually a nickel cyclinder with a centrally located hole, coaxial with the tube axis. The number of electrons emitted from the cathode is called the *beam intensity* and is controlled by the negative bias on the control grid. This action is identical to the control grid operation of a conventional vacuum tube. The electrons, emitted from the cathode and passing through the hole in the control grid, are accelerated by the high positive potential which is applied to the *accelerating anodes*, c and e, with beam focusing provided by the *focusing anode*, d. The accelerating and focusing anodes are also cylindrical in form, with small openings located in the center of each cylinder, coaxial with the tube axis. These holes in the electrodes permit the passage of the electrons past the vertical and horizontal deflection plates toward the fluorescent screen.

Because of the mutual repulsion of the negatively charged particles (the electrons) which together make up the beam, a method for focusing the electrons into a *narrow sharply defined* beam must be provided. The CRO uses the *electrostatic* method of focusing (as compared to a TV picture tube which employs *magnetic* focusing).

9-4 Electrostatic Focusing

Electrostatic focusing is used in all CROs. In order to examine the operation of the electrostatic focusing method it is useful to consider first the behavior of an individual particle in an electric field. Consider the diagram of Fig. 9-3, which shows a hypothetical electron situated at rest in an electric field. The force (f) which acts on the electron is given by the definition of electric field intensity, ϵ. This definition states that the force on a unit positive charge at any point in an electric field is the electric field intensity at that point. By definition, therefore,

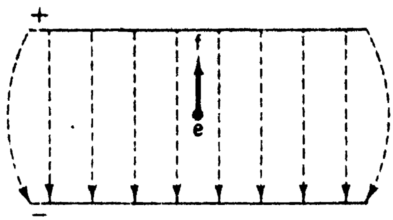


Figure 9-3
Force f on an electron situated in a uniform electric field.

$$\epsilon = \frac{f}{q} \quad (V/m) \tag{9-1}$$

where ϵ = electric field intensity, in V/m
f = force on the charge, in N
q = charge, in C.

An electron is a *negatively* charged particle and its charge is

$$e = 1.602 \times 10^{-19}\text{C} \tag{9-2}$$

The force on the electron in an electric field is, from Eq. (9-1),

$$f_e = -e\epsilon \quad \text{N} \tag{9-3}$$

where the minus sign indicates that the force acts in a direction *opposite* to the direction of the electric field.

This discussion is valid only when the electric field in which the charged particle is situated is of *uniform* intensity. That this is not always so can be seen from Fig. 9-4, where the electric field between two *parallel plates* of finite dimensions is shown. In Fig. 9-4, the field intensity is directed from the positive to the negative plate. The lateral repulsion of the electric field lines causes a spreading of the space between the lines, resulting in a curvature of the field at the

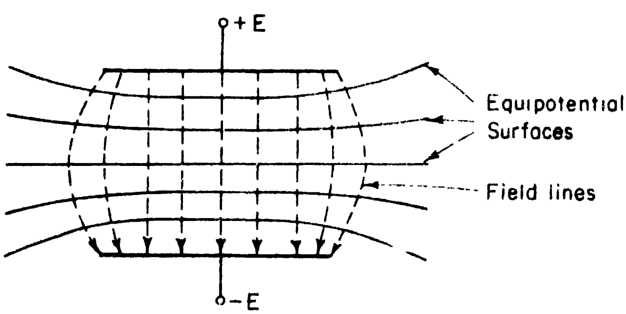


Figure 9-4
Electric field and equipotential surfaces for two parallel plates.

ends of the plates. The *density* of the field lines, therefore, will be less at the ends of the plates than in the center region between the plates. When points of equal potential are connected on each of the field lines, we obtain *equipotential surfaces,* shown as the solid lines in Fig. 9-4. Since the force on an electron acts in a direction opposite to the direction of the field, we can also state that the force on an electron is in the direction *normal to the equipotential surfaces.*

When two *cylinders* are placed end to end and a potential difference is applied to them, the resulting electric field between the two cylinders is not of uniform density. Lateral repulsion will again cause spreading of the lines resulting in a field as shown in Fig. 9-5. The equipotential surfaces are again shown

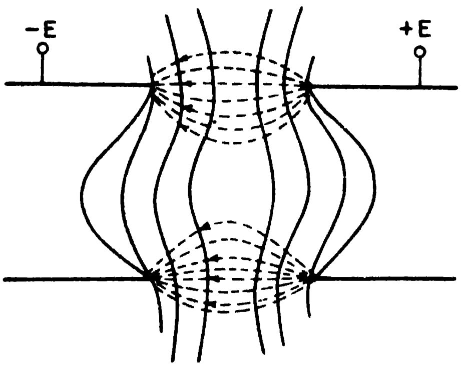


Figure 9-5
Equipotential surfaces for two cylinders placed end to end.

as the solid lines and because of the varying density of the electric field in the area between the cylinders, the equipotential surfaces are *curved.*

Consider now the regions on both sides of an equipotential surface, S, as shown in Fig. 9-6. The potential to the left of the surface S is V^- and to the

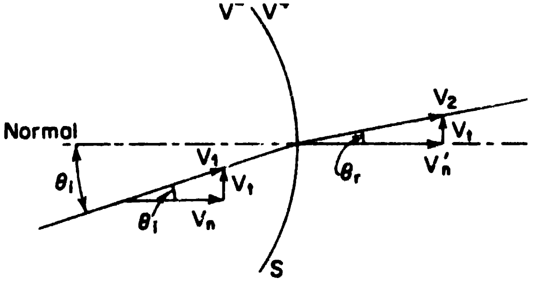


Figure 9-6
Refraction of an electron ray at an equipotential surface.

right of S is V^+. An electron, which is moving in a direction AB at an angle with the normal to the equipotential surface and entering the area to the left of S with a velocity V_1, experiences a force at the surface, S. This force acts in a direction normal to the equipotential surface. Because of this force, the velocity of the electron increases to a new value, V_2, after it has passed S. The tangential component, V_t, of the velocity on both sides of S remains the same. Only the normal component of the velocity, v_n, is increased by the force at the equipotential surface to a new value, v'_n. From Fig. 9-6 then, it follows that

$$v_t = v_1 \sin \theta_i = v_2 \sin \theta_r \tag{9-4}$$

where θ_i is the angle of incidence and θ_r the angle of refraction of the electron ray.

Rearranging Eq. (9-4) we obtain

$$\frac{\sin \theta_i}{\sin \theta_r} = \frac{v_2}{v_1} \tag{9-5}$$

Equation (9-5) is identical to the expression relating the refraction of a light beam in geometrical optics. The refraction of an electron ray or an electron beam at an equipotential surface follows the same laws as the bending of a light beam at a refracting surface, such as a lens. For this reason, an electrostatic focusing system in a CRT is sometimes called an *electron lens*.

Consider now the *electrostatic focusing arrangement* shown in the functional diagram of Fig. 9-7. Electrode c is named the *preaccelerating anode*; it is a metal

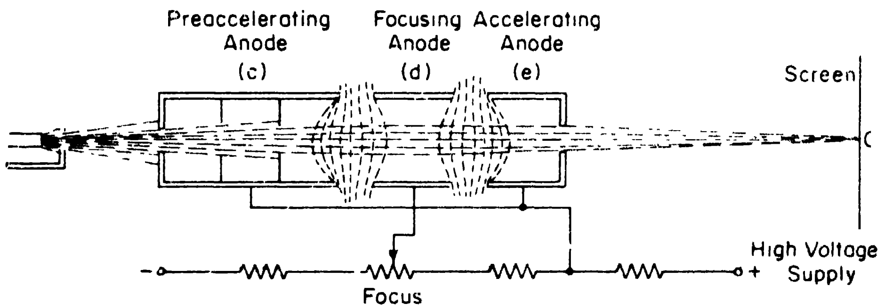


Figure 9-7
Electrostatic focusing arrangement of a CRT.

cylinder containing several baffles to collimate the electron beam which enters through the small opening on the left-hand side. The preaccelerating anode is connected to a high positive potential supplied by the CRT high-voltage power supply. The second electrode is the *focusing anode*, d; the third electrode is the *accelerating anode*, e. These cylindrical anodes are coaxial with the preaccelerating anode. The accelerating anode, e, is electrically connected to the same potential as the preaccelerating anode, c, and the focusing anode, d, is connected to a lower voltage.

Because of the difference in potential between the focusing anode and both accelerating anodes, an electric field exists between them. The electric field lines are nonuniformly spaced, as was shown in Fig. 9-5, causing the equipotential surfaces to form a *double concave* lens system. This is indicated in Fig. 9-7 by the dashed lines in the area between the electrodes. Electrons, entering the field at angles other than normal to the equipotential surfaces, will be refracted towards the normal and the beam of electrons will become focused toward the center of the tube axis. Varying the voltage on the focusing anode with respect to the accelerating anodes changes the *refraction index* of the electron lens and moves the focal point of the beam along the CRT axis.

The complete three-element system of Fig. 9-7 shows the beam focused at point C where the fluorescent screen is located. The potentiometer which allows control of the voltage on the focusing anode is a front panel control, usually marked *FOCUS*.

9-5 Electrostatic Deflection

In discussing the electrostatic deflection method of an electron beam in an oscilloscope we return to the statement made in Sec. 9-4 regarding the force on the electron in a uniform electric field. For convenience, the diagram of Fig. 9-3 is reproduced in Fig. 9-8. By definition of the electric field intensity, ϵ, the force, f, on the electron is, Eq. (9-3), $f_e = -e\epsilon$ N. The action of the force on the electron will accelerate it in the direction of the positive electrode, along the lines of the field flux. Newton's second law of motion allows us to calculate this acceleration since

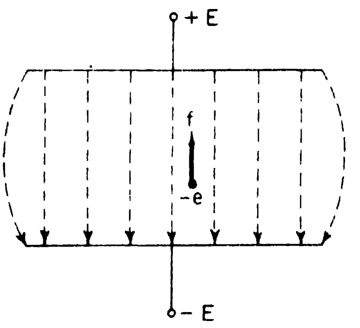


Figure 9-8
Force f on an electron in a uniform electric field.

$$f = ma \tag{9-6}$$

Substituting Eq. (9-2) into Eq. (9-6), we obtain

$$a = \frac{f}{m} = \frac{-e\epsilon}{m} \quad \text{m/s}^2 \tag{9-7}$$

where a = acceleration of the electron, m/s²
f = force on the electron, N
m = mass of the electron, kg.

When the motion of an electron in an electric field is discussed it is usually specified with respect to the customary Cartesian axes, as shown in Fig. 9-9. In discussing the concepts which follow, we shall use subscript notation for the *vector* components of velocity, field intensity, and acceleration. For example,

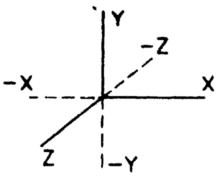


Figure 9-9
Cartesian coordinate system.

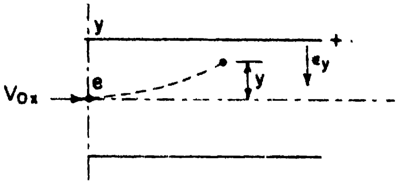


Figure 9-10
Path of a moving electron in a uniform electric field.

the velocity component along the X axis will be written v_x (m/s). The component of the force along the Y axis is written f_y (N), etc. The motion of an electron in a given electric field cannot be determined unless the initial values of velocity and displacement are known. The term *initial* represents the value of velocity or displacement at the time of observation or time $t = 0$. The subscript 0 will be used to indicate these initial values. For example, the initial velocity component along the X axis is written as v_{0x}.

Consider now an electric field of constant intensity with the lines of force pointing in the negative Y direction, shown in Fig. 9-10. An electron, entering this field in the positive X direction with an initial velocity v_{0x}, will experience a force. Since the field acts only along the Y axis, there will be no force along either the X axis or the Z axis, and the acceleration of the electron along these axes must be zero. Zero acceleration means constant velocity, and since the electron enters the field in the positive X direction with an initial velocity v_{0x}, it will continue to travel along the X axis at that velocity. Since the velocity along the Z axis was zero at time $t = 0$, there will be no movement of the electron along the Z axis.

Newton's second law of motion, applied to the force on the electron acting in the Y direction, yields

$$f = ma_y \quad \text{or} \quad a_y = \frac{f}{m} = \frac{-e\epsilon_y}{m} = \text{constant} \tag{9-8}$$

Equation (9-8) indicates that the electron moves with a *constant acceleration* in the Y direction of the uniform electric field. To find the *displacement* of the electron due to this accelerating force, we use the well-known expressions for velocity and displacement:

$$v = v_0 + at \quad \text{(m/s)} \qquad \text{(velocity)} \tag{9-9}$$

$$x = x_0 + v_0 t + \tfrac{1}{2}at^2 \quad \text{(m)} \quad \text{(displacement)} \tag{9-10}$$

Subject to the initial condition of zero velocity in the Y direction, ($v_{0y} = 0$), Eq. (9-9) yields

$$v_y = a_y t \quad \text{(m/s)}$$

which, after substitution of Eq. (9-8), results in

$$v_y = \frac{-e\epsilon_y t}{m} \quad \text{(m/s)} \tag{9-11}$$

The *displacement* of the electron in the *Y direction* follows from Eq. (9-10), which yields, applying the initial conditions of zero displacement ($y_0 = 0$) and zero velocity ($v_{0y} = 0$):

$$y = \tfrac{1}{2}a_y t^2 \quad \text{(m)}$$

which, after substitution of Eq. (9-8) results in

$$y = \frac{-e\epsilon_y t^2}{2m} \quad \text{(m)} \tag{9-12}$$

The *X distance,* traveled by the electron in the time interval t, depends on the initial velocity, (v_{0x}), and we can write, again using Eq. (9-10):

$$x = x_0 + v_{0x}t + \tfrac{1}{2}a_x t^2 \quad \text{(m)}$$

which after applying the initial conditions for the X direction ($x_0 = 0$ and $a_x = 0$), becomes

$$x = v_{0x}t \quad \text{or} \quad t = \frac{x}{v_{0x}} \quad \text{(s)} \tag{9-13}$$

Substituting Eq. (9-13) into Eq. (9-12) we obtain an expression of the vertical deflection as a function of the horizontal distance traveled by the electron:

$$y = \left[\frac{-e\epsilon_y}{2v_{0x}^2 m}\right]x^2 \quad \text{(m)} \tag{9-14}$$

Equation 9-14 shows that the path of an electron, traveling through an electric field of constant intensity and entering the field at right angles to the lines of flux, is *parabolic* in the *X-Y* plane.

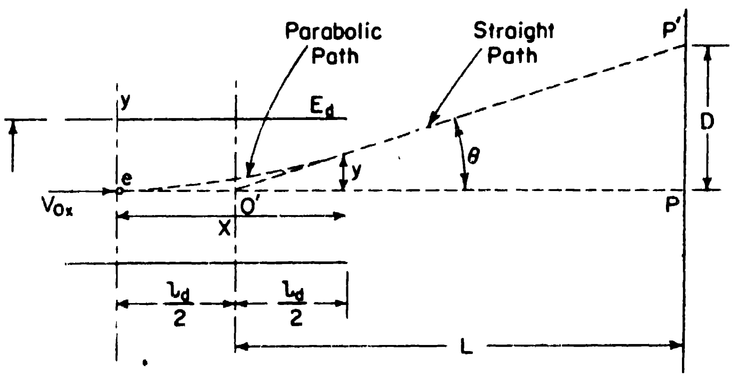


Figure 9-11
Deflection of the cathode ray beam.

In Fig. 9-11 two parallel plates, called *deflection plates,* are placed a distance d apart and are connected to a source of potential difference E_d, so that an electric field ϵ exists between the plates. The intensity of this electric field is given by:

$$\epsilon = \frac{E_d}{d} \quad \text{(V/m)} \tag{9-15}$$

An electron entering the field with an initial velocity v_{0x}, is deflected toward the positive plate following the parabolic path of Eq. (9-14), as indicated in Fig. 9-11. When the electron leaves the region of the deflection plates, the deflecting force no longer exists and the electron travels in a straight line toward point P′, a point on the fluorescent screen. The slope of the parabola at a distance $x = l_d$, where the electron leaves the influence of the electric field, is defined as

$$\tan\theta = \frac{dy}{dx} \tag{9-16}$$

where y is given by Eq. (9-14). Differentiating Eq. (9-16) with respect to x and substituting $x = l_d$ yields

$$\tan\theta = \frac{dy}{dx} = -\frac{e\epsilon_y l_d}{mv_{0x}^2} \tag{9-17}$$

The straight line of travel of the electron is tangent to the parabola at $x = l_d$ and this tangent intersects the X axis at point O′. The location of this *apparent* origin 0′ is given by Eq. (9-14) and Eq. (9-17) since

$$x - 0' = \frac{y}{\tan\theta} = \frac{e\epsilon_y l_d^2/2mv_{0x}^2}{e\epsilon_y l_d/mv_{0x}^2} = \frac{l_d}{2} \quad \text{(m)} \tag{9-18}$$

The apparent origin 0′ is therefore at the center of the deflection plates and a distance L from the fluorescent screen.

The deflection on the screen is given by

$$D = L\tan\theta \quad \text{(m)} \tag{9-19}$$

Substituting Eq. (9-17) for $\tan\theta$ we obtain:

$$D = L\frac{e\epsilon_y l_d^2}{mv_{0x}^2} \quad \text{(m)} \tag{9-20}$$

The potential energy of the electron entering the area between the deflection plates with an initial velocity v_{0x} is

$$\tfrac{1}{2}mv_{0x}^2 = eE_a \tag{9-21}$$

where E_a is the accelerating voltage in the electron gun.

Rearranging Eq. (9-21), we obtain

$$v_{0x}^2 = \frac{2eE_a}{m} \tag{9-22}$$

Substituting Eq. (9-15) for the field intensity, ϵ_y, and Eq. (9-22) for the velocity of the electron in the X direction, v_{0x}, into Eq. (9-20) for the screen deflection, D, we obtain

$$D = L\frac{e\epsilon_y l_d^2}{mv_{0x}^2} = \frac{Ll_dE_d}{2dE_a} \quad \text{(m)} \tag{9-23}$$

where D = deflection on the fluorescent screen (meters)
L = distance from center of deflection plates to screen (meters)
l_d = effective length of the deflection plates, (meters)
d = distance between the deflection plates (meters)
E_d = deflection voltage (volts)
E_a = accelerating voltage (volts).

Equation (9-23) indicates that, for a given accelerating voltage E_a and for the particular dimensions of the CRT, the deflection of the electron beam on the screen is directly proportional to the deflection voltage E_d. This direct proportionality indicates that the CRT may be used as a *linear voltage-indicating device.*

This discussion assumed that E_d was a fixed dc voltage. The deflection voltage usually is a varying quantity and the image on the screen follows the variations of the deflection voltage in a linear manner, according to Eq. (9-23).

The *deflection sensitivity* S of a CRT is definied as the deflection on the screen (in meters) per volt of deflection voltage. By definition therefore

$$S = \frac{D}{E_d} = \frac{Ll_d}{2dE_a} \quad \text{(m/V)} \tag{9-24}$$

where S = deflection sensitivity (m/V)
D = deflection on the screen (m)
E_d = deflection voltage (V).

The *deflection factor* G of CRT, by definition, is the reciprocal of the sensitivity S and is expressed as

$$G = \frac{1}{S} = \frac{2dE_a}{Ll_d} \quad \text{(V/m)} \tag{9-25}$$

with all terms defined as for Eq. (9-23) and (9-24). Both the expression for deflection sensitivity S and for deflection factor G indicate that the sensitivity of a CRT is independent of the deflection voltage but varies linearly with the accelerating potential. High accelerating voltages therefore produce an electron beam which requires a high deflection potential for a given excursion on the screen. A highly accelerated beam possesses more kinetic energy and therefore produces a brighter image on the CRT screen but this beam is also more difficult to deflect and we sometimes speak of a hard beam. Typical values of deflection factors range from 10 V/cm to 100 V/cm, corresponding to sensitivities of 1.0 mm/V to 0.1 mm/V, respectively.

Referring now again to the diagramatic view of the CRT in Fig. 9-1 and the photograph of the electron gun in Fig. 9-2, we see that the electron beam, after leaving the gun, passes between two pairs of deflection plates. One pair of plates is mounted in the vertical direction, establishing an electric field in the horizontal plane. The force on the electrons passing between these plates is then also in the horizontal direction. This pair of plates, mounted vertically but causing horizontal beam deflection, is called the *horizontal deflection plates*. The other pair of plates is mounted horizontally and produces a vertical deflection of the beam. These plates are called the *vertical deflection plates*. The arrangement of the deflection plates in a conventional CRT is shown in the photograph of Fig. 9-2. Notice that the plates are flared. If the deflection voltage, applied to either set of plates, is large, it is possible for the electrons to strike the edge surface of the plates, be deflected, and therefore never reach the screen. The flared plates allow the beam to pass the area between them without striking the plates.

A simultaneous application of deflection voltages to both sets of deflection plates produces a beam deflection in both the X direction and the Y direction, producing varying patterns on the screen. Certain types of patterns are useful in the determination of frequency or phase relationships of the deflection voltages. The techniques for using these patterns are described in Sec. 9-16.

9-6 Screens for CRTs

When the electron beam strikes the coated screen of the CRT a spot of light is produced. The screen material on the inner surface of the CRT which produces this effect is called the *phosphor*. The phosphor absorbs the kinetic energy of the bombarding electrons and reemits energy at a lower frequency in the visual spectrum. This property of some crystalline structures to emit light when stimulated by radiation, is called *fluorescence*. Fluorescent materials, like phosphor or zinc oxide, have a second characteristic, called *phosphorescence*, which refers to the property of the phosphor to continue light emission even after the source of excitation (in this case the electron beam) is cut off. The amount of time during which phosphorescence occurs, is called the *persistence* or *luminescence* of the fluorescent material. Persistence is usually classified as short (microseconds), medium (milliseconds), or long (seconds).

The *intensity* of the light emitted from the CRT screen depends on several factors. First, intensity is a function of the physical characteristics of the screen material. Second, the intensity is controlled by the number of bombarding electrons per second or *beam current*. The beam current in turn is controlled by the bias voltage on the CRT control grid. Third, the light intensity depends on the energy with which the bombarding electrons strike the screen and this is determined by the velocity of the electrons and therefore the accelerating voltage. High-velocity electrons cause a brighter image on the screen, but, on the other hand, a high-velocity beam is more difficult to deflect, as previously noted.

Several screen materials are available and a proper selection of the type of phosphor must be made on the basis of the application of the CRO. Table 9-1 lists some of the characteristics of the more commonly used phosphors in CRTs.

The bombarding electrons, striking the CRT screen, release secondary emission electrons, keeping the screen in a state of electrical equilibrium. These

TABLE 9-1
CATHODE RAY TUBE PHOSPHOR CHARACTERISTIC (STANDARD RETMA PHOSPHORS)

Phosphor Type	Trace Color Under Excitation	Afterglow	Persistence	Application
P1	green	green	medium	General purpose CRO
P2	yellow-green	yellow-green	medium-short	Observation of low- and medium-speed nonrecurrent phenomena
P4	white	yellow	medium	Television picture tube
P7	white	yellow-green	long	Observation of low-speed recurrent or medium-speed nonrecurrent phenomena
P11	blue	blue	medium-short	Photographic applications
P31	green	green	medium-short	Observation of low- or medium-speed nonrecurrent phenomena

secondary emission electrons are of low velocity and are collected by a conductive coating deposited on the inside surface of the glass envelope. This coating is usually an aqueous solution of graphite, known as *aquadag*, which is electrically connected to the second anode. In some tubes, particularly CRTs using magnetic focusing, the accelerating anode is dispensed with entirely and the conductive coating on the glass walls is used as the final accelerating anode. The TV picture tube is an example of this system.

9-7 CRT Connections

The connections to the various elements inside the glass envelope of the CRT are made through the base of the tube. Connections for a typical CRT with electrostatic deflection is shown in Fig. 9-12.

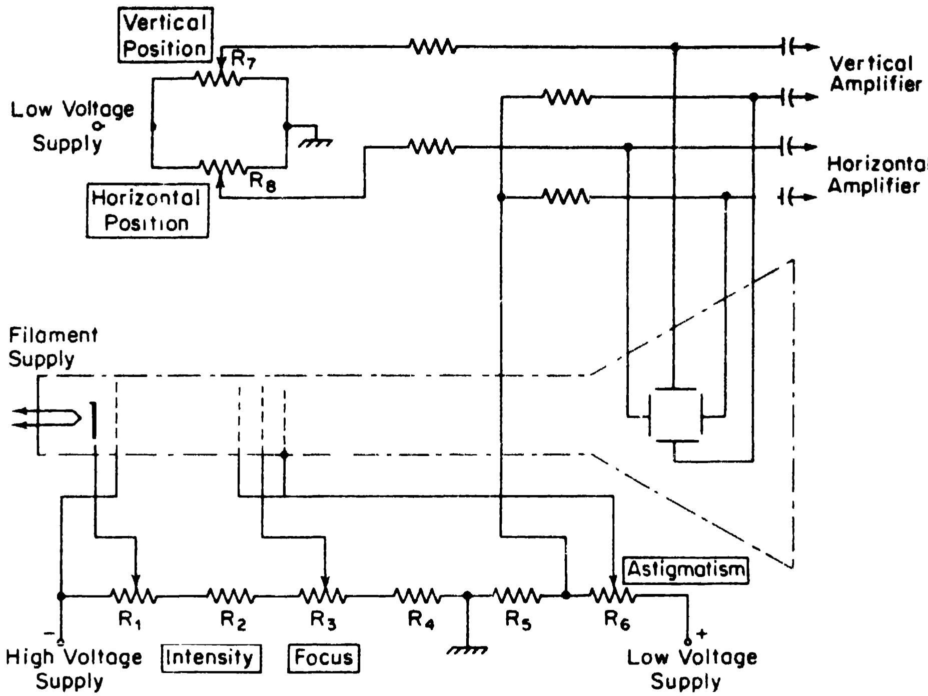


Figure 9-12
Typical CRT connections with the various front panel controls indicated.

The dc supply for the accelerating and focusing anodes is usually obtained from a high-voltage half-wave rectifier circuit. Resistors R_1 to R_6 represent a voltage divider network which provides the correct operating voltages to these anodes. The beam current or intensity of the electron beam may be adjusted by means of the bias control potentiometer R_1 which controls the cathode to control-grid voltage. Potentiometer R_1 is a front-panel control of the oscilloscope, marked INTENSITY. Resistor R_3 adjusts the voltage on the focusing anode with respect to the accelerating anodes and is also a front-panel control, marked FOCUS. The magnitude of the voltage on the focusing anode is usually about one-fourth to one-fifth of the voltage on the accelerating anodes.

It is observed from Fig. 9-12 that the accelerating anodes and the deflection plates are very close to ground potential and that the cathode is below ground potential by the amount of the accelerating voltage. The reason for this is that the waveform to be observed is applied to the plates and to protect the operator from high-voltage shock when making connections to the plates, they should be near ground potential. Also, the deflection voltages are measured with respect to ground and therefore high-voltage blocking capacitors are avoided by this connection. Resistors R_7 and R_8 are shunted across a portion of the low-voltage bleeder and provide varying amounts of dc voltage to each pair of deflection plates. This dc bias voltage allows adjustment of the position of the image on the screen, independent from the deflection voltages. Resistors R_7 and R_8 are front panel controls, marked HORIZONTAL POSITION and VERTICAL POSITION, respectively.

9-8 The CRT Graticule

The graticule is usually rectangular in form and is placed inside the display area to allow accurate measurements of the display. Most CROs have a graticule inscribed on a clear or tinted plastic plate, which is placed over the outside of the CRT face. When the graticule is used for making measurements, errors in reading may be caused by parallax, because the graticule is separated from the phosphor (on which the image is written) by the thickness of the glass face plate. Some CRT manufacturers incorporate an *internal* graticule, which is placed in the same plane as the phosphor. This avoids errors due to parallax.

9-9 Basic CRO Circuitry

The primary subsystems of a CRO are shown in the block diagram of Fig. 9-13. These subsystems are (*see p. 211*)

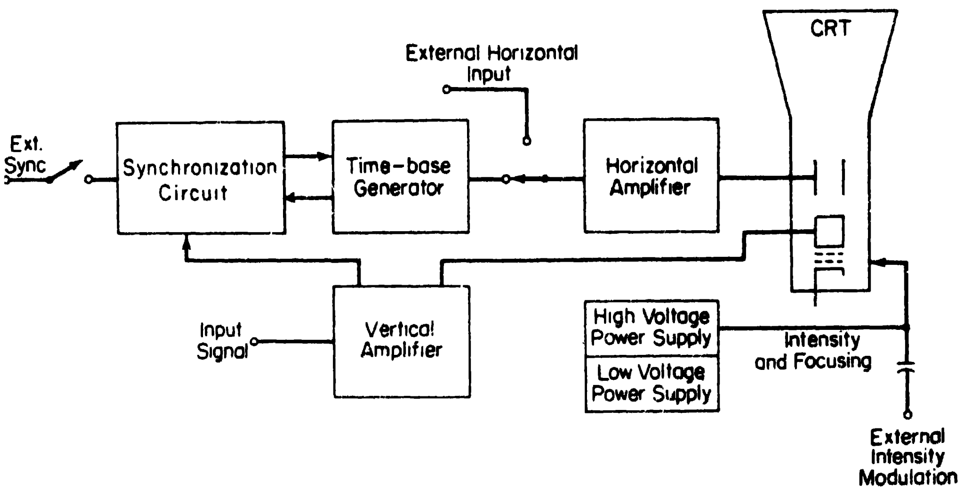


Figure 9-13
The basic subsystems of a general purpose oscilloscope.

(a) The vertical deflection system.

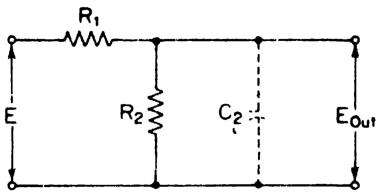
(b) The horizontal deflection system, including the time-base generator and synchronization circuitry.

(c) The CRT.

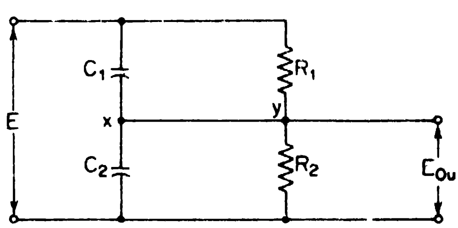
(d) The high-voltage and low-voltage power supplies.

The *vertical deflection system* consists of an input attenuator and a number of amplifier stages, which together must be able to accept a wide range of input voltages and amplify these sufficiently to drive the CRT spot. The amplifier must meet the requirements of *uniform gain* over a *wide band* of frequencies and freedom from nonlinear distortion. Much work has been done by CRO manufacturers to achieve these goals and amplifiers have been designed to meet specific requirements of high gain, wide bandwidth, freedom from noise, and other considerations. (Since there are so many different types of amplifiers, each with its own specific performance characteristics, see the information given in the instrument manuals published by the CRO manufacturers.)

The requirements for wide-band performance of a CRO call for a *compensated input attenuator.* Figure 9-14 shows this type of attenuator. Resistors R_1



(a)



(b)

Figure 9-14
(a) A vertical amplifier input attenuator, showing the input capacitance. (b) The compensated attenuator, redrawn as a bridge circuit.

and R_2 form a simple resistance combination, where the output signal equals $R_2/(R_1 + R_2)$ times the input voltage [9-14(a)]. Consider capacitor C_2 to be the inevitable input capacitance of the amplifier stage. The voltage division will now also be dependent on the *frequency* of the input voltage and it is the function of capacitor C_1 in Fig. 9-14(b) to prevent this dependence on frequency. The circuit suggests that the two resistors and the two capacitors form the four arms of a bridge circuit. The bridge will be in balance when $R_1X_{c2} = R_2X_{c1}$ or $R_1C_1 = R_2C_2$ and no current exists in the connecting branch XY. In order to calculate the output voltage, branch XY may be omitted from the circuit and the output voltage appearing at the junction of R_1 and R_2 will be $R_2/(R_1 + R_2)$ times the input voltage, independent of its frequency. Resistors R_1 and R_2 are usually large, so that the impedance of the attenuator prevents loading of the circuit under measurement. Capacitor C_1 is made adjustable in practice and the final correction for frequency compensation is made experimentally by the method of *square-wave testing*. This procedure is necessary because the compensation is critically dependent on the condition that $R_1C_1 = R_2C_2$.

The effect of an incorrectly compensated input attenuator is illustrated by the following discussion: Suppose that a step voltage is applied to the input terminals of the attenuator of Fig. 9-14(b). Referring to Fig. 9-15 it is seen that

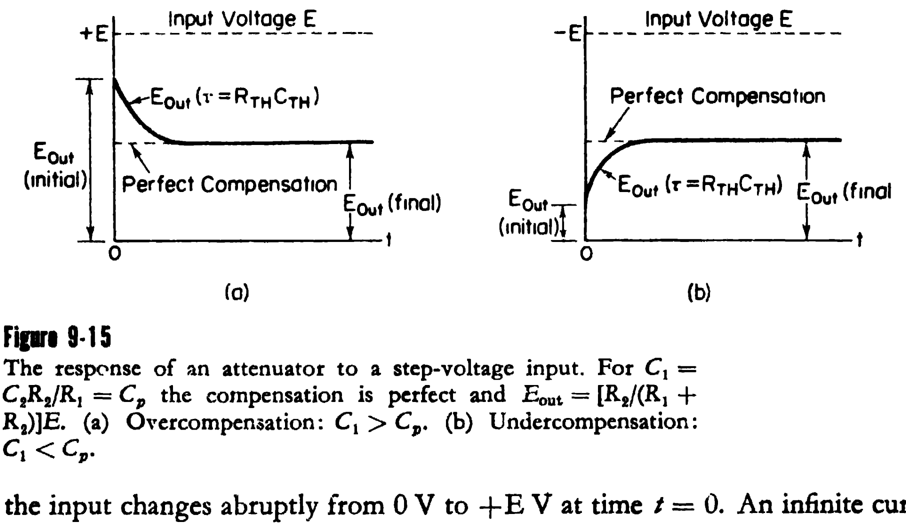


Figure 9-15
The response of an attenuator to a step-voltage input. For $C_1 = C_2R_2/R_1 = C_p$ the compensation is perfect and $E_{out} = [R_2/(R_1 + R_2)]E$. (a) Overcompensation: $C_1 > C_p$. (b) Undercompensation: $C_1 < C_p$.

the input changes abruptly from 0 V to +E V at time $t = 0$. An infinite current exists at $t = 0$ for an infinitesimal time, so that a charge $q = \int_{0-}^{0+} i\,dt$ is delivered to each capacitor. According to Kirchhoff's voltage law, the total voltage E at time $t = 0$ is

$$E = \frac{q}{C_1} + \frac{q}{C_2} = \left[\frac{C_1 + C_2}{C_1C_2}\right]q \quad \text{(V)} \tag{9-26}$$

and the output voltage at $t = 0+$ equals

$$E_{\text{out (initial)}} = \frac{q}{C_2} = \frac{C_1}{C_1 + C_2}E \quad \text{(V)} \tag{9-27}$$

The initial output voltage described in Eq. (9-27) is determined by the capacitors since they behave like *short* circuits for an instantaneous change in voltage. The final output voltage at $t = \infty$ is determined by the resistors because the capacitors act as *open* circuit elements under the steady-state condition of dc voltage.

Therefore,

$$E_{\text{out (final)}} = \frac{R_2}{R_1 + R_2} E \quad (\text{V}) \tag{9-28}$$

The Thévenin equivalent circuit of the attenuator as seen from the output terminals with the input short-circuited shows a series combination of R and C, where $R_{TH} = (R_1R_2)/(R_1 + R_2)$ and $C_{TH} = C_1 + C_2$. The output voltage of the attenuator therefore changes exponentially from its initial value to its final value with a time constant $\tau = R_{TH}C_{TH}$.

Perfect compensation is obtained when $E_{\text{out (final)}} = E_{\text{out (initial)}}$. From Eq. (9-27), (9-28), we get this condition for perfect compensation:

$$\frac{C_1}{C_1 + C_2} = \frac{R_2}{R_1 + R_2} \tag{9-29}$$

This expression is equivalent to the previously stated condition for bridge balance in Fig. 9-14(b). Compensation of the attenuator is perfect when

$$C_1 = C_2\frac{R_2}{R_1} = C_p \tag{9-30}$$

Overcompensation occurs when $C_1 > C_p$ and *undercompensation* when $C_1 < C_p$. These conditions are shown in Fig. 9-15.

The *horizontal deflection system* provides the drive voltages for moving the beam horizontally. Since so many measurements are taken with respect to *time*, the horizontal deflection system of a CRO always includes a *sawtooth oscillator or time-base generator*. (see Sec. 9-10). Also included in the horizontal deflection system are *synchronizing circuits*, so that the sweep can be started at a specific instant with respect to the waveform under observation. Synchronization is discussed in Sec. 9-14 and some actual synchronizing circuits are shown there. Finally, connections are provided for *external horizontal input* signals to deflect the beam horizontally in response to external waveforms other than the internal sweep. This technique is useful in making *X-Y plots*, such as Lissajous figures (see Sec. 9-16). The horizontal amplifier serves to provide sufficient drive voltage for moving the beam across the screen. Its gain and bandwidth requirements are usually much lower than for the vertical amplifier. (Details of the horizontal amplifier may be found in the manufacturer's CRO manuals.)

When a ramp voltage or sawtooth wave form, such as shown in Fig. 9-16, is applied to the horizontal deflection plates of the CRT, the beam will move across the screen at a constant velocity. This constant velocity occurs

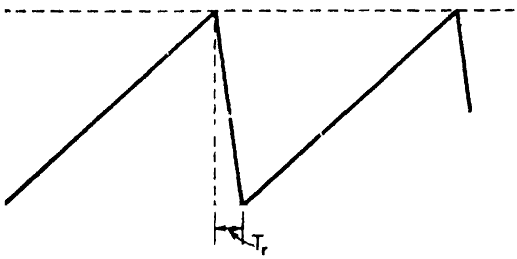


Figure 9-16
Horizontal deflection voltage, linearly increasing with time, is called a sawtooth voltage or ramp voltage.

because the deflection voltage increases linearly with time for the duration of the sweep. Time interval T_s (*sweep time*) in Fig. 9-16 shows how the ramp voltage rises linearly. When the ramp voltage terminates, the beam moves back rapidly to its original position. This takes place in the time interval T_r (*retrace time*) of the sawtooth waveform. The beam will move again only when another ramp voltage is applied to the horizontal deflection plates. The beam therefore traces out a straight line across the CRT screen from left to right.

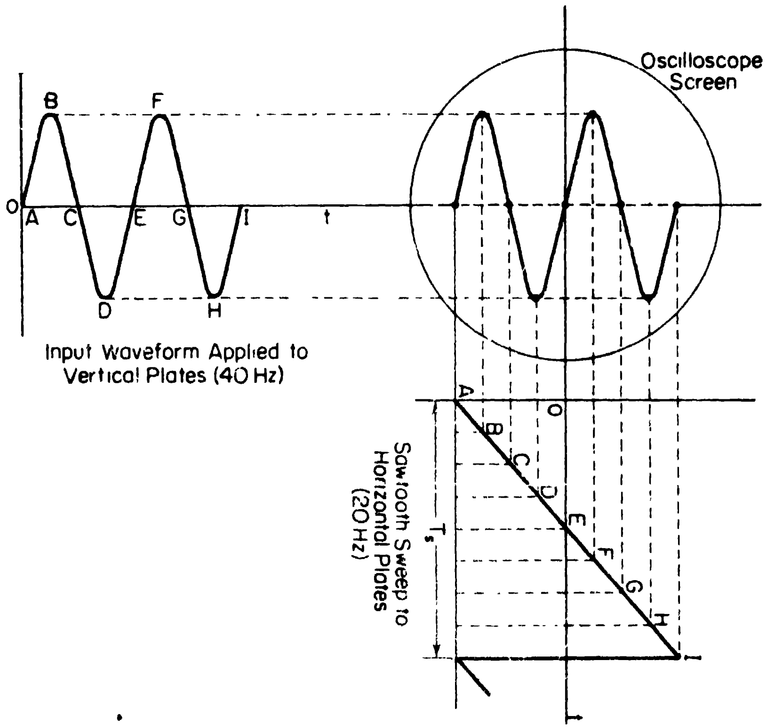


Figure 9-17
Oscilloscope pattern of a 40-Hz signal with a sweep voltage of 20 Hz.

When, simultaneously with the horizontal ramp voltage, an input signal of sinusoidal waveform is applied to the vertical deflection plates, the beam will be under the influence of two forces: one in the horizontal direction, moving the beam at a linear rate across the screen from left to right, and one in the vertical direction, moving the beam up and down according to the magnitude and the polarity of the vertical deflection voltage. Figure 9-17 shows the resultant motion of the beam by these two independent deflection voltages, supplied simultaneously to the X and the Y plates.

If the frequency of the horizontal ramp voltage is known, the frequency of the signal applied to the vertical deflection plates can easily be calculated. Assume, for example, that the time-base generator produces a ramp voltage with a repetition rate (RR) of 20, the beam will sweep across the screen twenty times per second. One sweep of the beam takes place in 1/20 or 0.05 s. If the wave-form portrayed on the screen contains two complete cycles, as in Fig. 9-17, these two cycles must have occurred in the time of 0.05 s. Therefore the frequency of the vertical deflection signal equals 2/0.05 or 40 Hz. The CRO may therefore be used to determine the frequency of an unknown signal, provided that (1) the time-base frequency is calibrated and known accurately; (2) the vertical deflection signal has a frequency which is an even multiple of the time-base frequency. If these two conditions are not met, frequency determinations with the CRO become rather unreliable.

One of the quickest and most accurate methods of determining frequencies uses the Lissajous patterns, which are produced when unknown and known frequencies of sine-wave voltages are applied simultaneously to both pairs of deflection plates (see Sec. 9-16).

The power supplies in the basic subsystem block diagram of Fig. 9-13 supply the electronic circuitry with both low and high voltages. The high-voltage section of the power supply also delivers the high accelerating potentials required for the CRT. These supplies are of conventional design and need no further elaboration.

9-10 Time-Base Generators

Most CRO applications involve measurement of a waveform which is varying with respect to time. This requires that the CRT spot move across the screen from left to right with a constant velocity. To deflect the beam from left to right in a linear manner, the following system is generally used:

(a) The right-hand horizontal deflection plate is supplied with a positive-going ramp voltage.

(b) The left-hand horizontal deflection plate is supplied with a negative-going ramp voltage.

Both waveforms are shown in Fig. 9-18.

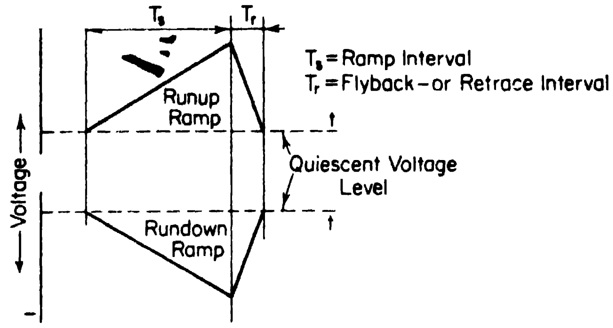


Figure 9-18
Linear deflection voltages: the run-up ramp and the run-down ramp.

The horizontal amplifier accepts either a *run-up* or a *run-down ramp* voltage and supplies both waveforms to its push-pull output circuit for simultaneous delivery to the deflection plates. The circuits which develop these ramp waveforms are called *time-base generators* or *sweep generators*. The output of a sweep generator is called a *sweep voltage*.

The typical form of a time-base voltage is shown in Fig. 9-19. It appears

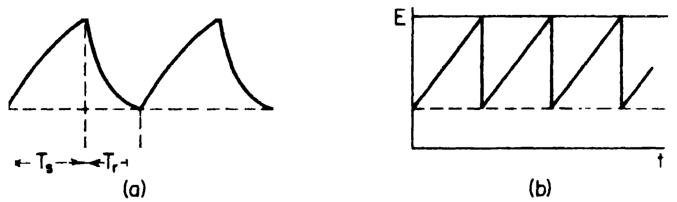


Figure 9-19
Time-base voltage waveforms. (a) A typical waveform. (b) The ideal waveform.

that this voltage, starting from some initial value, increases linearly with time to a maximum value, after which it returns again to its initial value. Time, T_s, is called the *sweep time* and time, T_r, the *retrace* or *flyback time*. Figure 9-19 shows a waveform where the retrace time is very short and a new linear voltage rise is initiated at almost the same instant that the previous one is terminated. These figures therefore suggest the name *sawtooth generator* or *ramp generator*.

Time-base generators do not necessarily provide sweep voltages that are exactly linear, although every attempt is made to obtain reasonable linearity of the voltage rise. Most time-base generators are refinements of the basic RC charging circuit shown in Fig. 9-20. Switch S is initially closed and the voltage

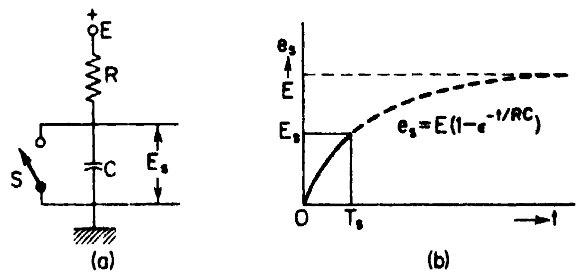


Figure 9-20
A basic RC charging circuit. (a) The actual circuit. (b) The output waveform.

across the capacitor is zero. At $t = 0$ the switch is opened and the voltage across the capacitor is given by the expression

$$e_s = E(1 - \epsilon^{-t/RC}) \tag{9-31}$$

showing an exponential voltage rise across the capacitor toward the final value of the supply voltage E. If the capacitor-charging process is stopped sufficiently early in its exponential rise, the curvature of the voltage rise is only slight and approaches a linear rise in voltage.

A simple RC sweep circuit is useful only in applications where *small* sweep voltages are required. For example, a 20-V sweep can be obtained with a *slope error* (deviation from linearity) of less than 10 per cent by using a supply voltage of at least 200 V. The charging process is ended by closing switch S in Fig. 9-20 and the sweep is terminated at time $t = T_s$. The switch performing this function is usually a vacuum tube, gas-filled tube, SCR, transistor, or other semiconductor device, depending on the application, speed required, and other practical considerations. Consider the use of a thyratron replacing the switch of Fig. 9-20. The circuit of Fig. 9-21(a) is discussed and the waveform of the voltage across the capacitor is shown in Fig. 9-21(b). It will be recalled that a thyratron ionizes, when the anode reaches its breakdown potential corresponding to the critical grid voltage. At this point, the tube loses control over the plate current, regardless of any variations in grid voltage. Once an arc is formed, the voltage drop across the tube remains essentially constant and the current through the tube is determined by the circuit external to the thyratron. The arc will be extinguished only by reducing the tube current below the minimum required to maintain ionization. Once the arc has been extinguished, the grid again regains control and determines the plate voltage that must be applied to cause breakdown. Generation of the sawtooth waveform proceeds as follows: Capacitor C charges exponentially through R, approaching the supply voltage E_{bb} as shown in Fig. 9-21(b). When the plate voltage reaches the value

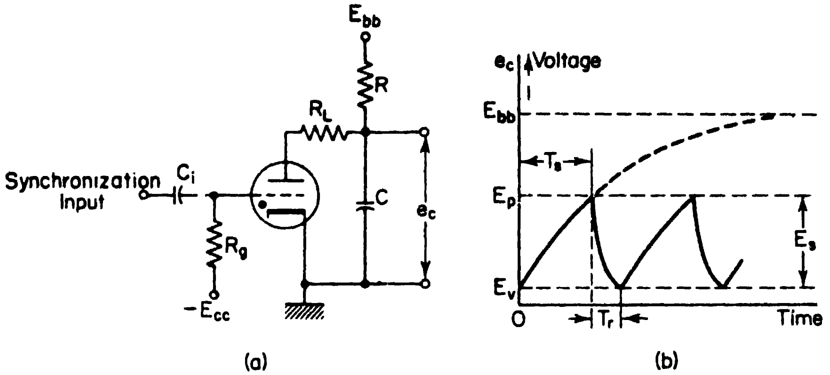


Figure 9-21
The thyratron switch in an *RC* charging circuit. (a) Sweep circuit. (b) Waveform.

E_p, the thyratron ignites and conducts heavily. Capacitor *C* will discharge rapidly through the tube and series resistor R_1, until the capacitor voltage drops to E_v. The tube will extinguish itself when the discharge current is less than the minimum required for ionization and conduction ceases. At this instant, the tube presents infinite resistance and the capacitor starts its charging process again through resistor *R*. The value of resistor R_1 is chosen so that the capacitor discharges rapidly but the current is limited to a value which the tube can safely handle. The process repeats itself, and the resulting sweep voltage is shown in Fig. 9-21(b). Linearity of the sweep is good, provided that E_s is small compared to E_{bb} and a sweep voltage proportional to time is obtained.

9-11 The Free-Running Mode

In the sweep circuit presently under consideration, the sawtooth waveform is repetitive. A new sweep is started immediately after the previous sweep is terminated and the circuit is not initiated by any signal external to the circuit. This mode of operation is called *free-running, stable,* or *recurrent.* The frequency of oscillation is a function of the supply voltage E_{bb}, the critical grid voltage E_{cc}, and the time constant $\tau = RC$. In practice, the frequency of oscillation is varied by changing the values of R and *C*, where R is termed the *timing resistor* and *C* the *timing capacitor*. When only R and *C* are changed, the amplitude of the sweep voltage remains constant, because amplitude depends only upon the supply voltages and the thyratron characteristics. The timing resistor is then used for *continuous* control of the frequency (*vernier control*) and the timing capacitor is changed *in steps* to provide different frequency ranges (*sweep range or time/cm control*).

When a periodic signal of frequency f_v is applied to the vertical input terminals of the CRO while a sweep of frequency f_s is applied to the horizontal axis, a stationary pattern of n cycles will appear on the screen if f_v is an even multiple of $f_s (f_v = nf_s)$. This is shown in Fig. 9-22 where two cycles of the vertical input waveform occur in the same time interval as one cycle of the sweep waveform. A portion of the last cycle of the input signal f_v occurs during the retrace time T_r.

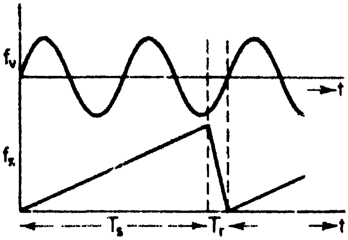


Figure 9-22
Illustrating the relationship between the vertical and horizontal deflection voltages.

From Fig. 9-22 it is seen that the beam retraces very rapidly from right to left on the screen, provided that T_r is very small. In this case, the CRT screen excitation will be so small that the return trace is hardly seen at all. In the thyratron circuit, a small value of capacitor C will reduce the discharge time and hence shorten the retrace time. Usually, however, T_r is not sufficiently small and the return trace will be visible as a faint line on the CRT screen. The trace may be made invisible by turning off the CRT beam during retrace. In the thyratron circuit of Fig. 9-21, this may be achieved by differentiating the retrace voltage and applying the resulting negative pulse directly to the grid of the CRT cutting the electron beam off. This process is called *retrace blanking.*

To maintain a stationary pattern on the CRT screen, the condition that $f_v = nf_s$ must be maintained at all times. In other words, the sweep generator must be *synchronized* to the vertical input signal. If f_v is slightly different from nf_s, the waveform will *drift* across the screen. Synchronization is achieved by applying the vertical deflection signal to the grid of the thyratron in Fig. 9-21, thus controlling the critical grid voltage and hence the tube's ionization potential. This is discussed in detail in Sec. 9-14.

9-12 The Triggered Mode

A waveform to be observed on the CRO may not be periodic but may perhaps occur at irregular intervals (aperiodic). In this case it is desirable that the sweep circuit remain inoperative and that the sweep be initiated by the waveform itself. Or perhaps the waveform is periodic, but it may be that the interesting part of the waveform is of very short time duration compared to the period of the waveform. For example, a waveform may consist of a number of 1-ms pulses, with a time interval of 100 ms between the pulses. If a free-running sweep were to be used, with the waveform synchronized to the sweep frequency, a stationary pattern would appear on the CRT screen if the sweep period were 100 ms. If the time base is spread out over a 10-cm area of the CRT screen, the pulse would occupy only 0.1 cm of the screen and all detail of the pulse would be lost. A sweep period of 1 ms or perhaps a bit larger, would spread the entire pulse out over the 10-cm screen area and the waveform of the pulse could be studied in detail. Therefore, what is required is a sweep period of, say, 1.5 ms initiated by the pulse itself. A circuit which would accomplish this is called a *triggered sweep* circuit.

Figure 9-23(a) shows a modification of the previously discussed thyratron

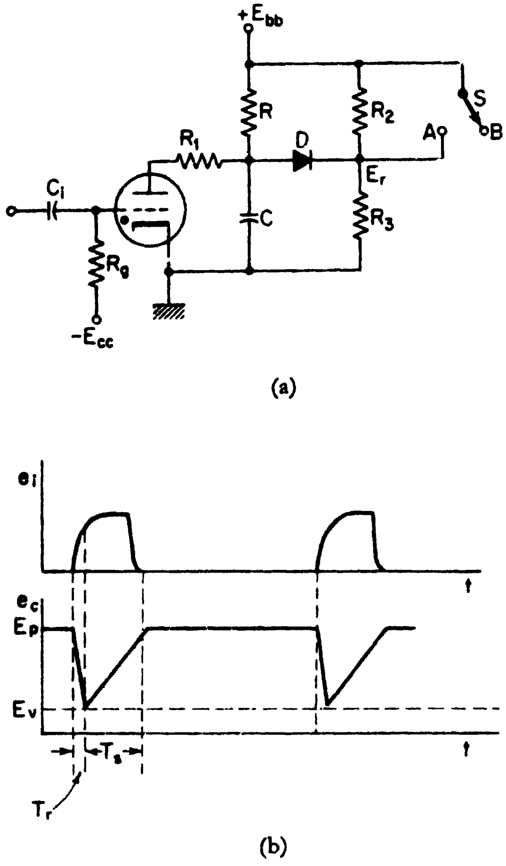


Figure 9-23
(a) A practical circuit for a triggered sweep. (b) The operation of the triggered sweep.

sweep circuit of Fig. 9-21. This modification consists of the addition of diode D and two resistors, R_2 and R_3. The values of R_2 and R_3 are chosen such that the voltage at their junction is less than E_p, the thyratron ionization potential. Therefore, when the thyratron is in the nonconducting state, the timing capacitor C charges up toward the thyratron firing voltage, but a point is reached where the diode conducts and prevents a further rise in capacitor voltage. The capacitor voltage then never reaches the thyratron firing voltage and the capacitor cannot discharge. It is clamped at a voltage E_r, as determined by R_2 and R_3. If now a positive voltage is applied to the grid of the thyratron, its ionization potential is lowered below the value E_r and the tube fires. The capacitor discharges quickly through R_1 (retrace time T_r) and again charges exponentially through R toward E_{bb}, but will be clamped when it reaches E_r.

Figure 9-23(b) shows the waveforms of the triggered sweep for the case just discussed. The pulse train is applied to the vertical deflection amplifier and is also used to trigger the sweep circuit. Note that at the leading edge of the pulse the thyratron is triggered and the capacitor discharges first, after which the charging cycle starts again. Also note that the sweep speed has been adjusted so that the entire pulse falls just within the sweep time T_s. The sweep circuit responds to the trigger signal only after it has been applied, so that a small part of the initial waveform will be lost due to the time-base retrace. Therefore this type of triggered sweep is used only for low-frequency waveforms where a small portion of the wave can be lost without too much harm.

The triggering circuit can be improved by decreasing the time required for retrace. For a thyratron circuit, however, this time is determined by the deionization time of the gas and this is in the order of 10 μs. A circuit using a unijunction transistor instead of a thyratron has a smaller retrace time but it is still in the order of a few microseconds. Most modern high-frequency CROs with fast sweeps use vacuum tube or transistor switches instead of thyratrons or UJTs. The addition of a switch, S, across resistor R_2 in Fig. 9-23(a) converts the triggered sweep circuit into a conventional free-running circuit. When the switch is closed, resistor R_2 is shorted out and diode D is connected directly to E_{bb}. Under these circumstances D never conducts because its cathode is at a higher potential than the ionization voltage of the thyratron.

9-13 A Time-Base Circuit for a General-Purpose CRO

Figure 9-24 is a block diagram of a time-base circuit used in a general-purpose CRO. This circuit uses the *thyratron switch* as indicated in Fig. 9-21. The polarity of the sync signal (positive for the thyratron) is selected by the *sync polarity* switch, which takes the sync signal across either a plate or a cathode resistor. The *sync amplifier* is of conventional design and is provided with a gain control. The tnyratron *sweep circuit* is essentially the same as that discussed so far and the *diode clamp* is the network shown in Fig. 9-23(a). The *blanking circuit* (retrace beam blanking) consists of a high-pass RC differentiatin gnetwork followed by an amplifier. The sweep voltage is applied to a *paraphase amplifier* to provide both a run-up and a run-down ramp as discussed previously.

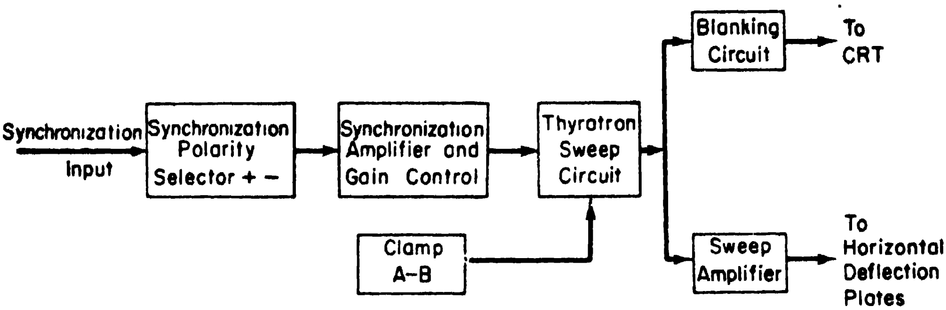


Figure 9-24
Block diagram of a sweep system for a general-purpose oscilloscope.

Laboratory oscilloscopes are designed to make accurate measurements with respect to time and therefore require a sweep with excellent linearity. Several methods may be used to improve sweep linearity. Among the more important ones are

(a) Constant current charging, whereby the timing capacitor is charged linearly from a constant current source.

(b) The Miller sweep circuit, whereby a step input is converted into a linear ramp by using an operational integrator.

(c) The phantastron circuit, which is a variation of the Miller circuit.

(d) The bootstrap circuit, whereby constant charging current is maintained by keeping the voltage across the charging resistor and hence the charging current through it, constant.

(e) Compensating networks, which are used to improve the linearity of the Miller and bootstrap circuits.

Detailed analysis of the circuits just listed falls outside the scope of this text. Students who wish to pursue this matter may consult the many excellent textbooks* in the field of pulse and switching circuits.

9-14 Synchronization of the Sweep Circuit

The operation of *synchronization* may best be understood by referring to the circuit of Fig. 9-21(a), which is redrawn for convenience in Fig. 9-25. In the

*Millman & Taub, *Pulse, Digital, and Switching Waveforms* (New York: McGraw-Hill Book Company, Inc., 1965), pp. 514–596; John M. Doyle, *Pulse Fundamentals* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1963), pp. 314–332; 455–458.

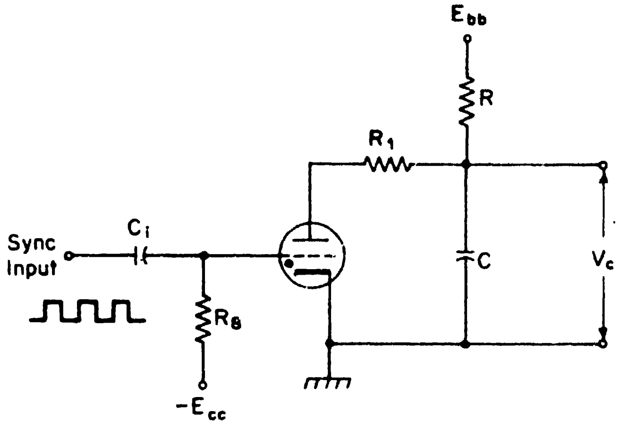


Figure 9-25
A thyratron sweep generator.

absence of an external signal, the capacitor stops charging when the capacitor voltage e_c reaches the breakdown voltage E_p of the thyratron. The capacitor discharges rapidly through the thyratron and in order to simplify the waveform drawings, we assume that the capacitor discharges in zero time. Synchronization to an external signal is possible, by applying this external signal to the *sync terminals* in Fig. 9-25 in such a manner as to change the breakdown voltage E_p. In the case of the thyratron, a *positive* pulse applied to the grid will *lower* E_p.

A pulse train of synchronizing pulses applied to the grid of the thyratron lowers its breakdown voltage for the duration of each pulse. In Fig. 9-26(a), a train of regularly spaced pulses, starting at an arbitrary time $t = 0$, is shown. The first several pulses have no effect on the sweep generator, which continues to run at its natural frequency, unsynchronized. Eventually, the exact moment at which the thyratron fires is determined by the instant at which the pulse occurs ($t = T$), as is also each succeeding beginning of the capacitor discharge time. In order for synchronization to occur, the synchronizing pulses must take place at the time intervals causing the charging process to stop prematurely. The interval T_p between pulses must therefore be less than the natural period (T_o) of the sweep generator. When the pulse interval T_p is larger than the natural frequency, synchronization does *not* occur. Figure 9-26(b) shows that four sweep cycles occur in the time of three pulse periods and this is of no value. Sync pulses with too small an amplitude do not cause synchronization either because the amplitude of the sync pulses must be large enough to bridge the gap between the quiescent breakdown voltage and the instantaneous sweep voltage. Figure 9-26(c) shows the case where the sync pulse amplitude is too small and synchronization does not take place.

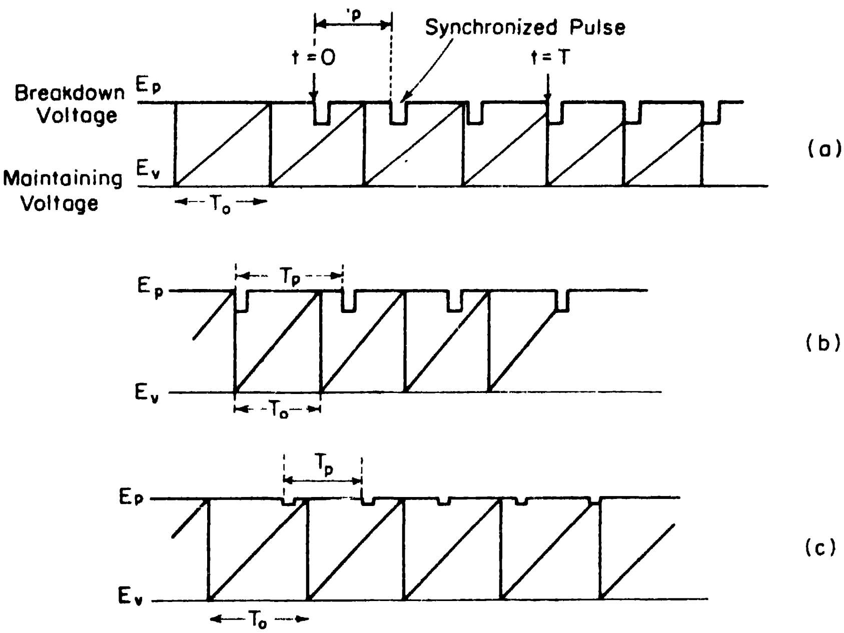


Figure 9-26
The principle of synchronization of the sweep.

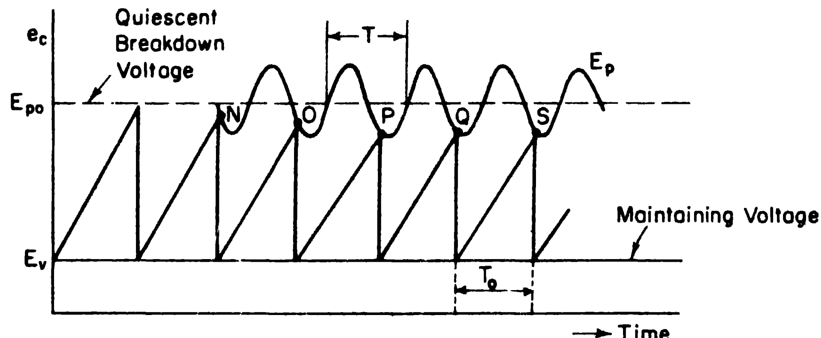


Figure 9-27
The time relationship that must exist between the synchronization signal (E_p) and the sweep voltage in a synchronized sweep.

Consider now the case where the sync signal is sinusoidal, as indicated in Fig. 9-27. The dashed line in Fig. 9-27 is the thyratron breakdown voltage (E_{po}) in the absence of a sync signal and the solid curve is the breakdown voltage in the presence of a sync signal. Synchronization is established when the period (T) of the sync signal is equal to the natural period (T_o). The frequency of the sync signal must be slightly greater than that of the free-running sweep in order to start the process of synchronization. Successive triggering of the sweep takes place at subsequent reduced ionization voltages of the thyratron, indicated by points O, P, Q, and S in Fig. 9-27, and the sweep is synchronized.

Figure 9-28 illustrates the case where the sync amplitude is exceedingly

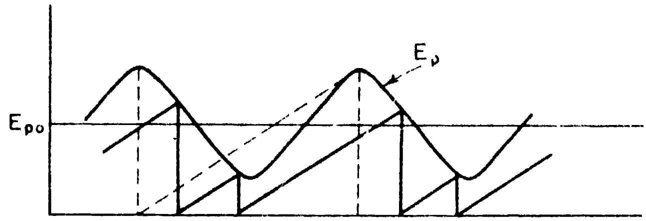


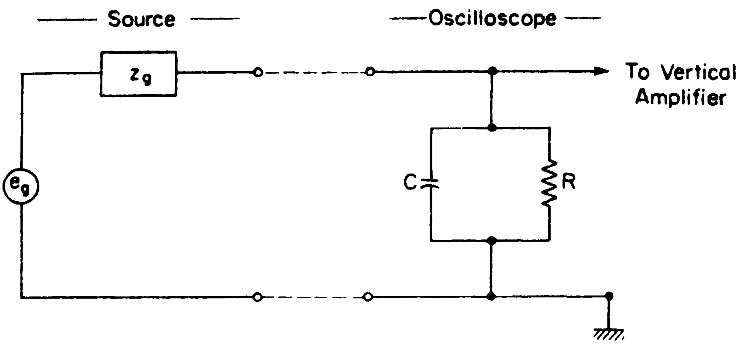
Figure 9-28
A possible result of excessive sync signal amplitude.

large and causes the circuit to initiate two sweeps during its charge-discharge process. The actual sweep waveform however consists of alternate long and short sweeps. Figure 9-28 suggests that when a sweep is used in conjunction with a CRO, it is advisable always to use as small a sync signal as possible. A sweep waveform as shown in Fig. 9-28 will cause a piecewise display of each cycle of the signal being observed on the CRO screen.

9-15 CRO Probes

The CRO *probe* performs the important function of connecting the instrument to the circuit under measurement, without altering, loading, or otherwise disturbing the test circuit. This is a basic requirement for all instrumentation, and the CRO is no exception. The image on the CRT screen must represent a true picture of the waveform, therefore, as it was *before* the CRO was connected into the circuit.

A typical vertical amplifier presents an input impedance of 1 MΩ shunted by a capacitance of 12 to 47 pF. In order to make a good measurement, the amplifier input impedance must be taken into account. To illustrate this point, consider the circuit of Fig. 9-29 which shows the equivalent circuit of a signal source connected to the input terminals of the vertical amplifier. If the source impedance is high, the small current drawn by the CRO input circuit may cause an appreciable voltage drop across the source impedance. The CRT display therefore may show an amplitude which is not a faithful reproduction of the original signal amplitude. At the higher frequencies, the shunt capacitance of the input circuit plays a role and introduces both *phase shift* and a reduction in signal *amplitude* of the displayed waveform.



9-29
effect of oscilloscope input impedance on source impedance.

Most CROs have listed among their accessories various types of probes and connecting cables. The principal types of probes used are the following:

(a) The direct probe.
(b) The high-impedance probe.
(c) The high-voltage probe.
(d) The detector probe.

The *direct probe* consists of a coaxial cable with a probe tip. The connection from the CRO to the test point is a direct one, where only the capacitance of the cable must be taken into consideration. Cable capacitance can add up to 50 pF to the vertical input shunt capacitance, depending on the length and type of coaxial cable. The shunt capacitance reduces the over-all high-frequency response of the scope, and direct probes should therefore only be used in those low-frequency applications where the added capacitance does not affect the measurement.

The *high-impedance probe* prevents undesired circuit loading by decreasing the total input capacitance and increasing the input resistance. The simple circuit of a high-impedance passive probe is shown in Fig. 9-30(a) with its equivalent circuit in Fig. 9-30(b). The probe consists of a voltage divider network R_1 and R_2, the values of which are generally chosen to give an attenuation ratio of 10/1. In general, CROs with an input resistance of 1 MΩ are provided with a 9-MΩ probe resistance, giving an over-all input resistance of 10 MΩ. The 10/1 attenuation ratio is maintained over the entire frequency range by the variable capacitor $C1$, which provides frequency compensation. (Attenuator frequency compensation has been dealt with in Sec. 9-9.)

It is necessary, before using a high-impedance probe, to make sure that the probe is correctly compensated. This is most easily done by applying a square-wave input signal (rich in high-frequency harmonics) to the probe tip

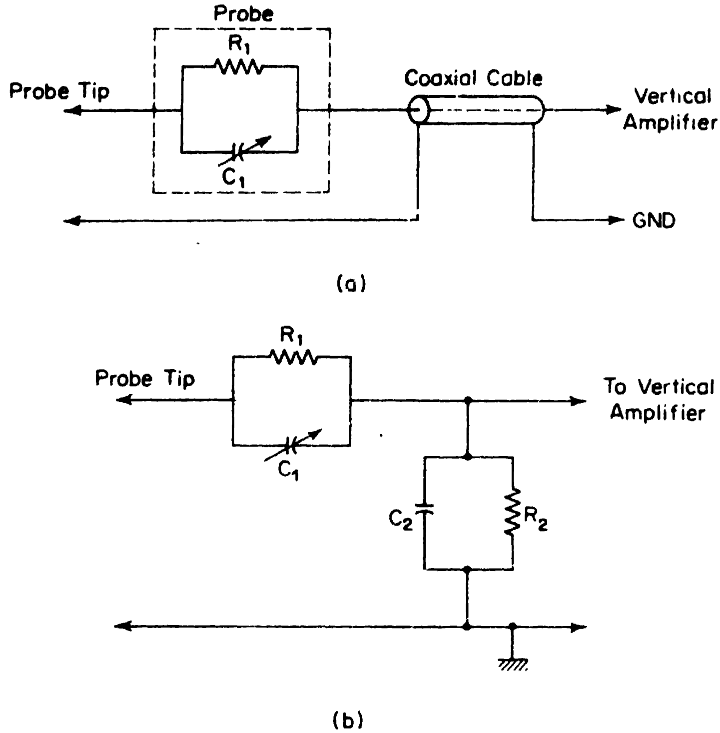


Figure 9-30
A high-impedance passive probe and its equivalent circuit, showing frequency compensation by adjustment of capacitor C_1.

and adjusting the variable capacitor C_1 until optimum display response is obtained. Most CROs are provided with an internal square-wave calibrator voltage which can be used for this purpose. Remember that the input impedance may vary between CROs and that a probe which was properly adjusted for one CRO may be out of adjustment when used with a different instrument.

A high-impedance probe of the *cathode follower* type is shown in Fig. 9-31 and is classified as an *active proble*. When viewing waveforms of low amplitude the passive high-impedance probe may attenuate the signal too much. The cathode-follower probe still presents a high impedance to the circuit under test, but since it has approximately unity gain, the attenuation is negligible. The active probe is essentially a small-signal, wide-band device and is not recommended for faithful reproduction of input signals exceeding 3 V p-p.

The *high-voltage probe* is used in applications where the input voltage is greater than the allowable voltage ratings of the instrument as specified by the CRO manufacturer. The probe consists of a resistive or capacitive voltage divider, which allows only a small fraction of the input signal to reach the

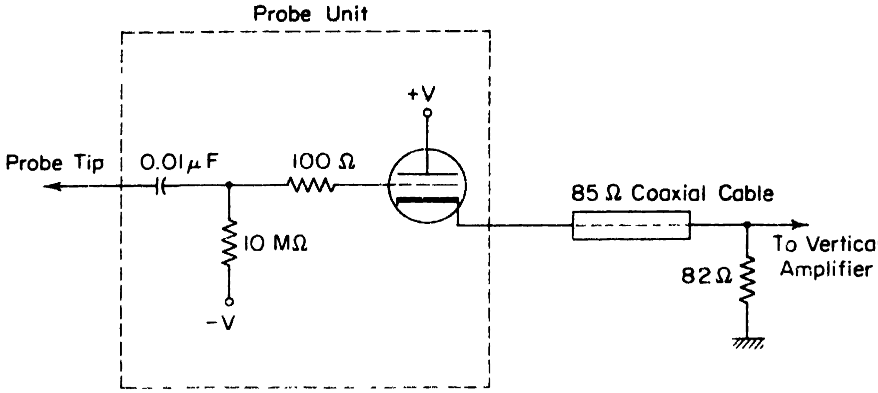


Figure 9-31
The active probe: a cathode-follower circuit.

vertical input terminals. Typical dc ratings are 12 kV with an input shunt resistance in the order of 100 MΩ and a shunt capacitance of 3 pf. For voltages above 12 kV, a capacitive voltage divider is recommended. This type of circuit can safely handle voltages up to 50 kV.

The *detector probe* is used in the response analysis of tuned circuits, where only the shape of the response curve is desired and not the high-frequency carrier. The circuit of a detector probe is shown in Fig. 9-32. Basically the probe

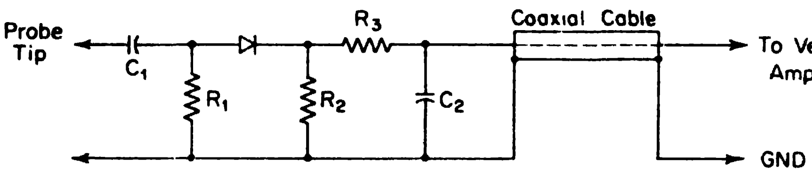


Figure 9-32
The detector or demodulator probe.

circuit consists of a diode detector which can be of either the series or the parallel type. The diode provides rectification of the modulated input signal and the shunt capacitor bypasses the RF component leaving the low-frequency envelope to be applied to the CRO input terminals. The resulting display of the low-frequency envelope can be used effectively for alignment of tuned circuits or for further analysis of the over-all tuned circuit response.

9-16 Lissajous Patterns

Lissajous patterns result when sine waves are applied simultaneously to both the horizontal and the vertical deflection plates of the CRO. The construction of a Lissajous figure is shown graphically in Fig. 9-33. Sine wave A represents the vertical deflection; sine wave B, the horizontal deflection voltage. The frequency of the vertical deflection signal is twice that of the horizontal signal, indicating that the CRT spot travels two complete cycles in the vertical direction against one cycle in the horizontal direction. Figure 9-33 suggests that numbers 1 to 24 on both waveforms represent points of corresponding time intervals. Assuming that the spot starts in the center of the CRT screen (point 0), the travel of the spot can be reconstructed in the manner indicated in Fig. 9-33, and the resultant figure is called a *Lissajous pattern*.

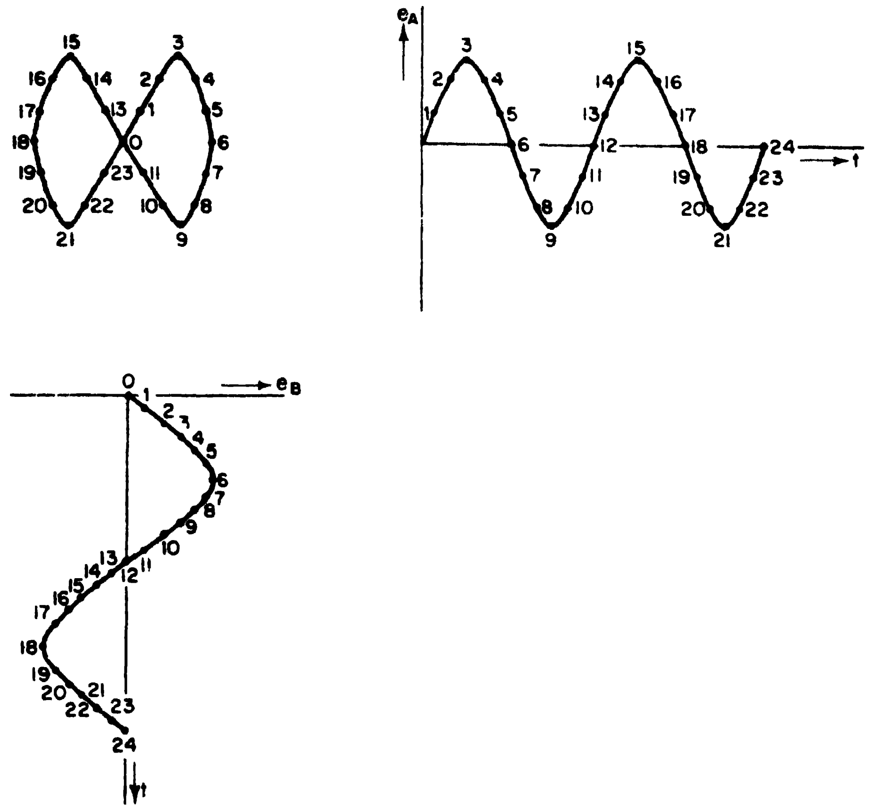


Figure 9-33
Graphical construction of a Lissajous pattern.

To determine the frequency of an unknown signal by using a Lissajous pattern, it is necessary to know one of the frequencies. This frequency is usually supplied by an external oscillator. The two numbers comprising the ratio of the two frequencies must be integers (1/1, 1/2, 1/5, etc.) in order for the pattern to be stationary. When the ratio is 1/1, or unity, the pattern is a straight line,

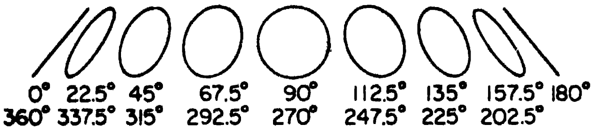


Figure 9-34 Lissajous patterns for a 1:1 frequency ratio.

an ellipse, or a circle, depending on the phase relationship of the two signals and also on their relative amplitudes. Figure 9-34, in which the two signals are assumed to be equal in amplitude, illustrates the various possibilities. In patterns of this type there is frequently a slow drift from one pattern shape to the other. This condition is due to a slight frequency instability in at least one of the signals.

There are many possible configurations for any ratio of applied signals. One consideration is whether the higher or the lower frequency is applied to the horizontal deflection plates. The most significant consideration, however, is the *phase* of the high-frequency signal with respect to the low-frequency signal. The pattern of Fig. 9-33 shows a figure eight, resting on its side, which results when both signals start out together. A tangent drawn against the top edge of the pattern would make contact at two places; a tangent drawn against a vertical side would be tangent at one place. Evidently, the horizontal tangencies correspond to the vertical deflection voltage and the vertical tangencies to the horizontal deflection voltage. Hence, the ratio of the vertical deflection frequency to the horizontal deflection frequency is 2/1.

Interesting patterns result when the high-frequency signal and the low-frequency signal do not start at the same time but are shifted in phase. Figure 9-35(b) shows the situation where the high-frequency signal is shifted ahead by 90°. Here the high-frequency signal is at its maximum when the low-frequency signal is just starting its cycle. When this condition occurs, the resulting pattern forms an inverted parabola. A pattern of this type is commonly referred to as a *double image,* since the electron beam, after reversing its direction, traces out exactly the same path. Figure 9-35 shows the intermediate phase relationships between the signals.

When a double image, such as the parabola, is developed, a different method of evaluating the frequency ratio must be used. In this case, a tangent drawn against an open end of the pattern is counted as a half tangency. For example, in Fig. 9-35(d), a tangent drawn against the top makes two contacts which are open ended and therefore count as one half tangency each, giving a total of one point. Against the vertical side there is only one open contact, giving a count of one-half. The ratio of vertical frequency to horizontal frequency therefore is still 2/1. There are some restrictions on the frequencies which can be applied to the deflection plates. One, obviously, is that the CRO must have the bandwidth required for these frequencies. The other restriction

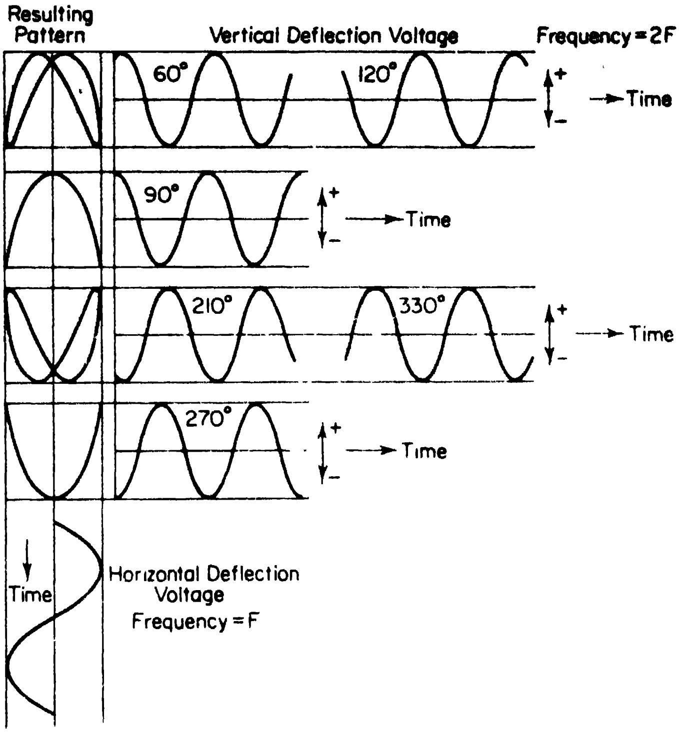


Figure 9-35
Lissajous patterns for various phase relationships.

is that the relationship between the two frequencies should not result in a pattern which would be too involved for an accurate determination of the frequency ratio. As a rule, ratios as high as 10/1 and as low 10/9 can be determined comfortably.

In addition to the patterns for integral ratios of the two frequencies there are many patterns for which the numerator and denominator of the ratio are whole numbers. For example, Fig. 9-36 shows the patterns for a 3/2 ratio and a

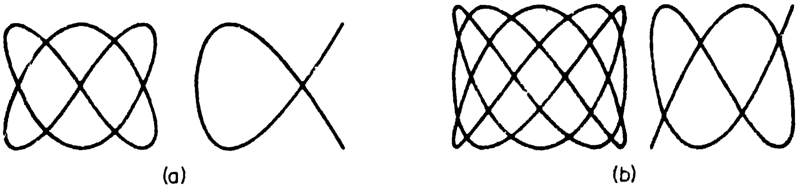


Figure 9-36
(a) 3: 2 Lissajous patterns. (b) 5: 3 Lissajous patterns.

5/3 ratio. In every case, the method for determining the ratio of the applied frequencies is the same as desribed earlier.

Two sine waves of the same frequency produce a Lissajous pattern which may be a straight line, an ellipse, or a circle, depending on the phase and amplitude of the two signals. A circle can be formed only when the amplitudes of both signals are equal. If they are not equal and/or out of phase an ellipse is formed, the axes of which are the horizontal and the vertical plane (assuming normal positioning of the CRT). Excluding the consideration of signal amplitude, the property which determines the type of pattern formed when two signals of the same frequency are applied to the deflection plates, is the *phase difference* between these signals. Figure 9-37 shows the phase relationships necessary for each of the patterns produced.

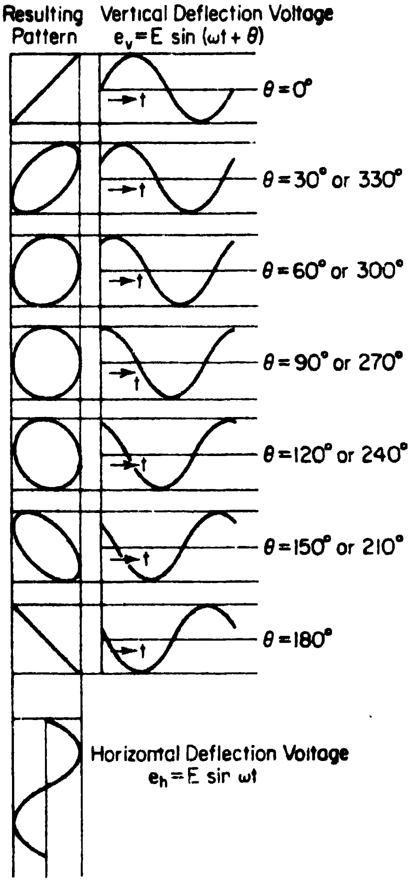


Figure 9-37
1 : 1 Lissajous patterns showing the effect of phase relationships.

A number of conclusions can be drawn from a study of these patterns. For example, a *straight line* results when the two signals are either in phase or 180° out of phase with each other. The angle formed with the horizontal will be exactly 45° when the amplitudes of the signals are equal. An increase in the vertical deflection voltage causes the line to have an angle greater than 45° with the horizontal. Similarly, a reduction in vertical amplifier gain results in a line with an angle smaller than 45° with the horizontal. A *circle* is displayed when the phase difference between the two signals is exactly 90° or 270°, assuming that the two signals are equal in amplitude. If the vertical signal has a larger amplitude, an *ellipse* with a vertical major axis is formed. When the horizontal signal is larger, the main axis of the ellipse will lie along the horizontal axis. In the case of ellipses resulting from phase differences other than 90°, a change in relation between the deflection voltages has the same effect.

Regardless of the relative amplitudes of the applied voltages, the ellipse

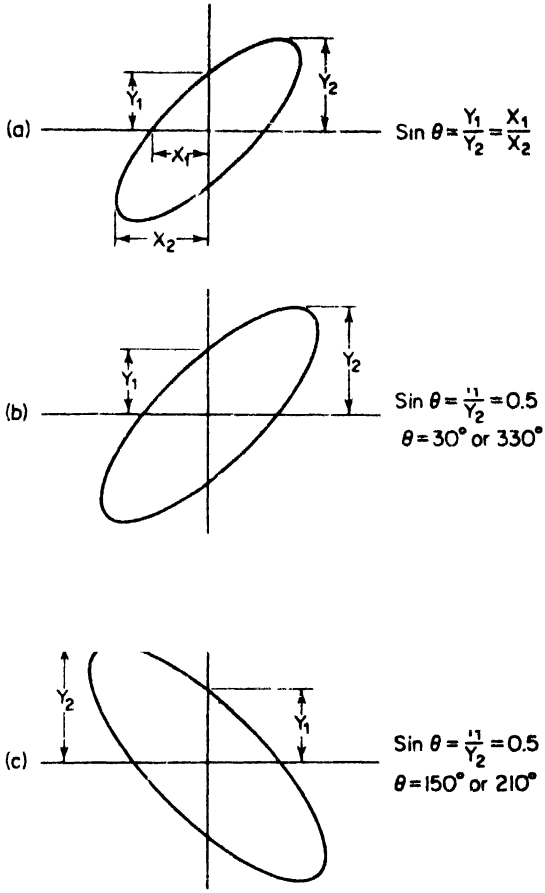


Figure 9-38
Computation of phase angles.

provides a simple means of finding the phase difference between the two signals. The method is illustrated in Fig. 9-38. The sine of the phase angle between the two signals is equal to the ratio between the Y-axis intercept represented by Y_1 to the maximum vertical deflection represented by Y_2. We can write

$$\sin\theta = \frac{Y_1}{Y_2} \tag{9-32}$$

For convenience, the gains of the vertical and horizontal amplifiers are adjusted so that the ellipse fits exactly within a square as marked by the coordinate lines on the graticule. Figure 9-38 shows how to interpret the phase angle corresponding to the orientation of the ellipse. If the major axis lies in the first and third quadrants, as shown in Fig. 9-38(b), the phase angle is either between 0° and 90° or between 270° and 360°. When the major axis passes through the second and fourth quadrants, the phase angle is between 90° and 180° or between 180° and 270°. In Fig. 9-38, the sine of the phase angle equals 0.5, corresponding to the different values of phase angles indicated.

Frequency comparison, using Lissajous patterns, is often performed on the CRO. With a small difference in the two frequencies the pattern appears to drift slowly, according to the phase difference between them. From a beginning, where the pattern is a straight line, the line opens to an ellipse, then to a circle, closes to an ellipse, and then to a straight line with inclination opposite to the original. This sequence occurs in a drift of the one-half cycle. At the completion of one cycle of difference, the pattern has returned to its original starting position. For example, if two oscillators are being compared, one with a frequency of 1,000 Hz and the other with a frequency of 1,001 Hz, the picture on the CRT screen completes one cycle of change in one second. If the frequency of one oscillator is adjustable so that several seconds are needed for one complete change of the pattern, then the two frequencies are known to be within a fraction of a cycle within each other, which is a very small percentage in terms of the 1,000-Hz frequency of oscillation. The situation where one frequency drifts slightly with respect to the other frequency causes the pattern to rotate or *barrel* from the integral ratio.

9-17 High-Frequency CRO Considerations

When the vertical deflection signal increases in frequency, the writing speed of the electron beam increases. The immediate result of higher writing speed is a reduction in image intensity on the CRT screen. In order to obtain sufficient image brilliance, the electron beam must be accelerated to a higher velocity so that more kinetic energy is available for transfer to the screen and normal image brightness is maintained. An increase in electron beam velocity is easily achieved by raising the voltage on the accelerating anodes. A beam with higher velocity also needs a greater deflection potential to maintain the deflection sensitivity, and this immediately places higher demands on the vertical amplifier.

One way to improve the deflection system at higher frequencies has been successfully developed by the CRO manufacturers and consists of the *traveling-wave type* CRT. Figure 9-39 shows such a CRT, where a series of deflection plates is mounted inside the CRT. These deflection plates are so shaped and spaced that an electron traveling between them will receive from each plate an additional deflecting force in the proper time sequence. The vertical deflection signal is applied to each plate through a delay line which is so designed that the time delays correspond exactly to the transit times of the electrons traveling down the CRT towards the screen. The electron speed must be con-

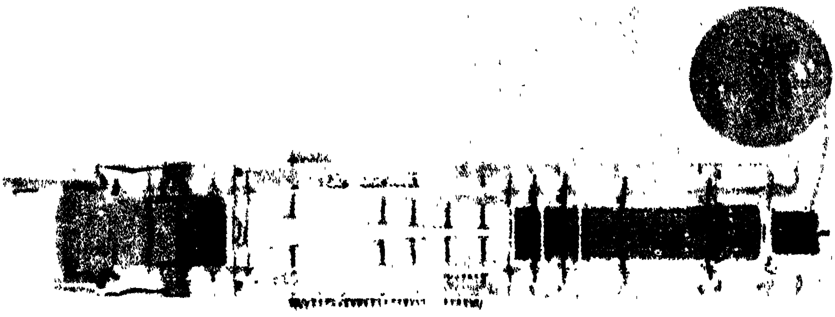


Figure 9-38
The traveling-wave type CRT. A special high-frequency CRT with a series of vertical deflection plates. (*Courtesy Tektronix, Inc.*)

trolled very accurately to avoid distortion in the trace. In addition to this special high-frequency CRT, new fluorescent materials have been developed to increase the image brightness at the higher frequencies. Further improvement in the vertical deflection system must be found in the vertical amplifiers themselves

In a conventional amplifier stage, increased bandwidth may be obtained at the expense of amplifier gain. In order to obtain an approximate relationship between gain, bandwidth, and number of stages in a cascaded amplifier, consider the following calculations:

For an RC-coupled vacuum tube amplifier, the voltage gain of a single stage may be expressed as

$$\mathbf{A}_V = \frac{\mathbf{E}_{\text{out}}}{\mathbf{E}_{\text{in}}} = -g_m R_{\text{sh}} \tag{9-33}$$

where g_m = the transconductance of the tube
R_{sh} = the total shunt resistance contained in the equivalent circuit of the stage,* consisting of the parallel combination of plate resistance r_p, load resistance R_L, and grid resistance R_g of the next stage.

The total gain of n stages in a cascaded arrangement is found by multiplication of the gains of the individual stages. For n stages with equal gain, $\mathbf{A}_V$, therefore

$$\mathbf{A}_{\text{cascaded}} = (-g_m R_{\text{sh}})^n \tag{9-34}$$

The bandwidth characteristics of the RC-coupled amplifier are described by the lower and upper 3-dB frequencies (the corner frequencies). The upper

*Paul M. Chirlian, *Analysis and Design of Electronic Circuits* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965), p. 203

3-dB frequency, indicating the high-frequency limitations of the amplifier stage, is given by

$$f_h = \frac{1}{2\pi R_{sh} C_{sh}} \tag{9-35}$$

where f_h = the upper 3-dB frequency
R_{sh} = the total shunt resistance of the equivalent circuit
C_{sh} = the total shunt capacitance of the equivalent circuit of the stage, consisting of the parallel combination of input capacitance C_{in} and output capacitance C_{out}.

When n stages are cascaded in the RC-coupled amplifier, its bandwidth is reduced by approximately $\sqrt{n}$. The upper corner frequency of an n-stage amplifier is therefore described by the approximate expression

$$f_{h(\text{cascaded})} = \frac{1}{2\pi R_{sh} C_{sh}} \sqrt{n} \tag{9-36}$$

For the RC-coupled transistor amplifier, similar considerations are valid. The current gain of a transistor amplifier stage is given by

$$\mathbf{A}_i = \frac{\mathbf{I}_{out}}{\mathbf{I}_{in}} = \frac{h_{fe} R_{sh}}{R_i} \tag{9-37}$$

where R_{sh} = the total shunt resistance of the equivalent circuit of the stage.*

Cascading n stages gives a total current gain of

$$\mathbf{A}_{i(\text{cascaded})} = \left(\frac{h_{fe} R_{sh}}{R_i}\right)^n \tag{9-38}$$

The upper corner frequency of the n-stage transistor amplifier is again given by Eq. 9-36.

The considerations above show that the gain of an amplifier may be substantially increased by cascading a number of stages, but the bandwidth, and therefore the upper frequency limit, will automatically be reduced. In a vacuum tube amplifier, the shunt capactances of a single stage are determined by the vacuum tube itself and by the circuit wiring. Careful wiring design and selection of a tube with small interelectrode capacitances may reduce the shunt capacitance, but once this limit is reached, the high-frequency response can only be increased by reducing the plate load resistance (see Eq. 9-36). To maintain sufficient gain, a reduction in plate load requires an increase in plate current and the size of the tube must then be increased.

One approach, used by some manufacturers, to solve the problem of maintaining adequate gain by supplying large currents to a small load resistance,

*Chirlian, op. cit., p. 211

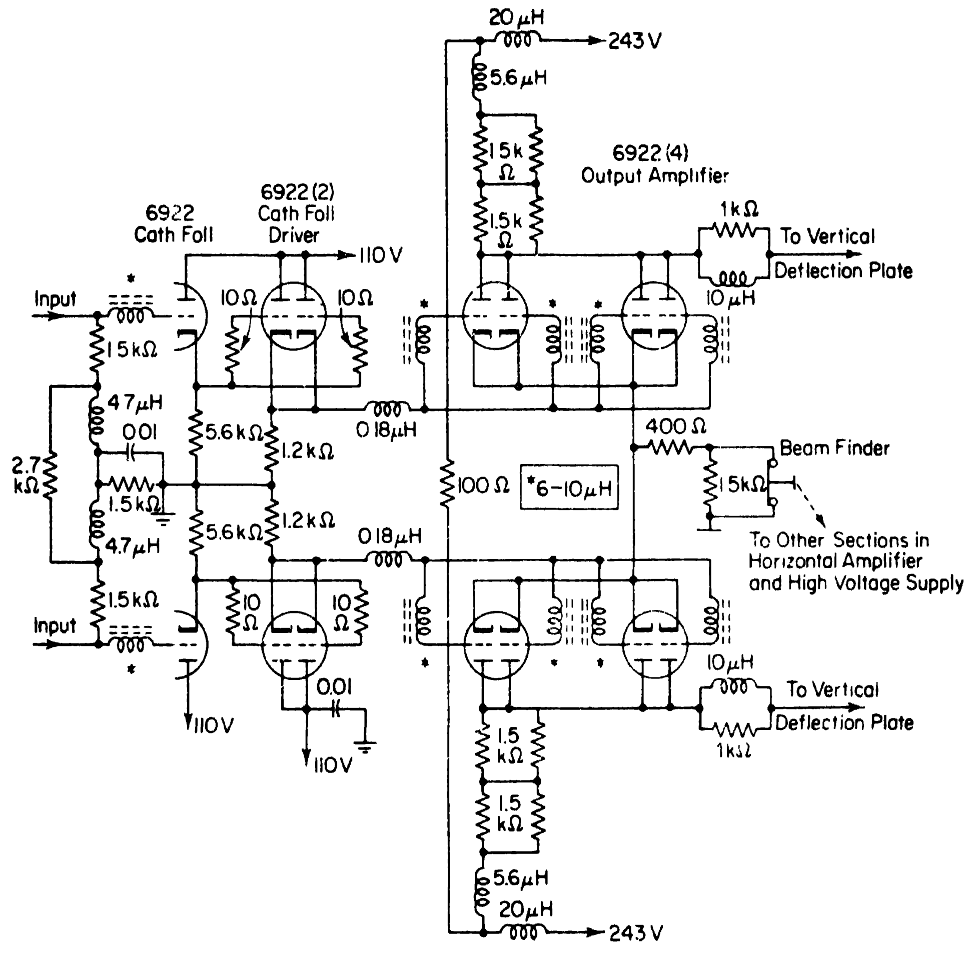


Figure 9-40

Partial circuit of the main vertical amplifier of a high-frequency laboratory oscilloscope, illustrating the use of parallel amplifiers to obtain sufficient deflection voltage and high bandpass. (*Courtesy Hewlett-Packard Company.*)

is found in the use of *parallel output tubes* in the vertical amplifier. This is shown in Fig. 9-40.

Some high-performance oscilloscopes use this method to advantage, achieving sufficient deflection sensitivity with high bandpass. In the circuit diagram of Fig. 9-40, several output tubes are connected in parallel, supplying a common plate load. The frequency response of this particular amplifier has been increased to 30 MHz with sufficient gain to drive the vertical deflection plates. For frequencies above approximately 30 MHz, even the parallel output tubes do not provide sufficient gain. The basic limitation of cascading n stages in an amplifier (namely, the shunt capacitances) is overcome by combining amplifier stages in the manner shown in Fig. 9-41.

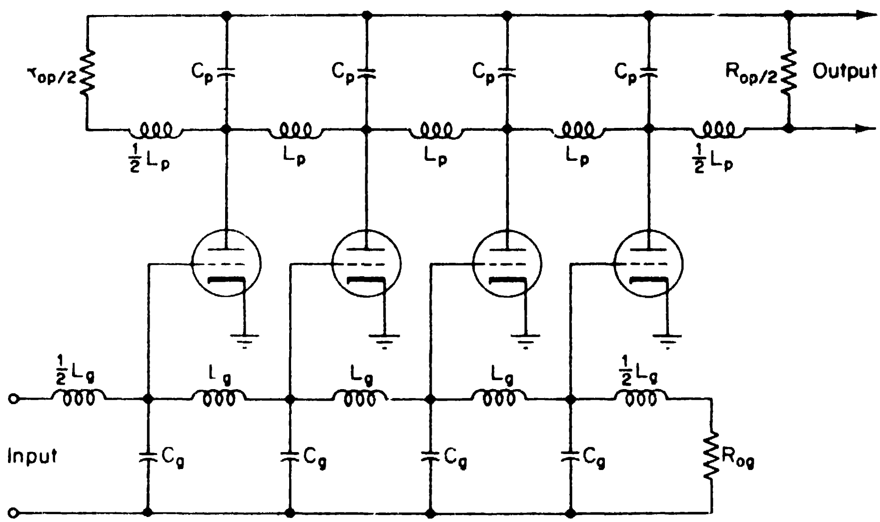


Figure 9-41
The basic distributed amplifier, here consisting of four sections.

This arrangement is called a *distributed amplifier.* Capacitances C_g and C_p represent the input and output capacitances of the vacuum tubes, together with the stray capacitances present. A signal, applied at the input terminals, travels down the *grid transmission line,* which consists of inductances L_g and capacitances C_g. The signal reaches the grid of each tube in turn, delayed by the time constant of each LC section and is finally absorbed in the matched termination resistance R_{og}. Each tube delivers plate current to the *plate transmission line,* which consists of inductances L_p and capacitances C_p and is terminated at both ends by the characteristic impedance of the transmission line (resistor $R_{op}/2$ at each end of the line).

The delays per section of grid line and plate line are carefully adjusted to be identical, so that all the current components reaching the plate line output termination in response to an input signal, will arrive at this termination at exactly the same time. The plate signals therefore add algebraically and the gain of this amplifier, having n sections, is

$$A = \tfrac{1}{2} n g_m R_{op} \tag{9-39}$$

where R_{op} is the characteristic impedance of the plate line. The upper frequency limit of the amplifier is determined only by the cut-off frequencies of the delay lines, and any improvement in the design of the delay lines will improve the performance of the amplifier.

The following distinctive features of the distributed amplifier may be observed:

(a) The gain of a distributed amplifier is computed by adding the gains provided by each stage. Therefore, even if the gain per tube is less than unity, the over-all gain still increases with an increasing number of tubes.

(b) At low frequencies, where the reactances of the elements of the transmission line are negligible, the amplifier simply appears as n tubes in parallel, feeding a load of $R_{op}/2$. Hence the gain is again given by Eq. (9-39).

(c) The distributed amplifier effectively parallels the tubes as far as the transconductances are concerned, but keeps the capacitances separate.

Figure 9-42 gives a practical circuit arrangement of the distributed amplifier as used in a modern laboratory-type CRO with a frequency response of 100 MHz. Notice that the special CRT with the series of deflection plates is also used in this instrument in order further to improve its sensitivity. The main deflection amplifier uses ten triode sections and a dual-pentode final stage. The sensitivity of the instrument is specified as 0.25 V/in. with a bandpass of dc to 100 MHz and a rise time of 3.5 ns.

The *sampling CRO* uses a different approach to improve the instrument's high-frequency performance. Here the input waveform is reconstructed from many *samples* taken during recurrent cycles of the input waveform and so circumvents the bandwidth limitations of conventional CRTs and amplifiers. The technique is illustrated by the waveforms indicated in Fig. 9-43.

In reconstructing the waveform, the sampling pulse *turns on* the sampling circuit for an extremely short time interval and the waveform voltage at that instant is measured. The CRT spot is then positioned vertically to the corresponding voltage input. The next sample is taken during a subsequent cycle of the input waveform at a slightly later position. The CRT spot is moved horizontally over a very short distance and is re-positioned vertically to the new value of the input voltage. In this way the CRO *plots* the waveform *point by point,* using as many as 1,000 samples to reconstruct the original waveform. The sample frequency may be as low as one-hundredth of the input signal frequency. If the input signal has a frequency of 1,000 MHz, the bandwidth of the amplifier required would be only 10 MHz, a very reasonable figure.

A simplified block diagram of the sampling circuitry is given in Fig. 9-44.

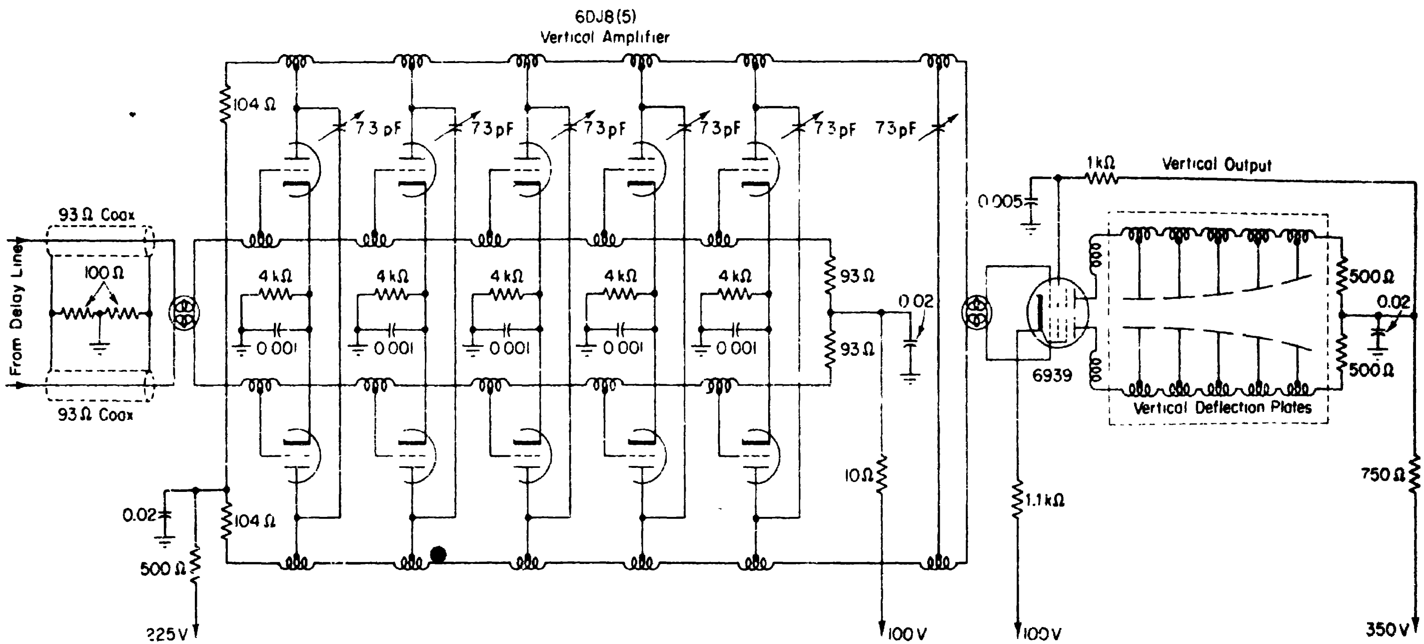


Figure 9-42. Schematic diagram of a distributed amplifier, also including the traveling-wave type CRT. (*Courtesy Tektronix Inc.*)

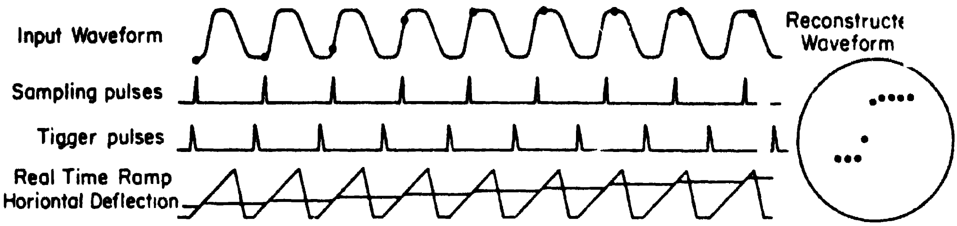


Figure 9-43
Waveforms pertinent to the operation of the sampling oscilloscope.

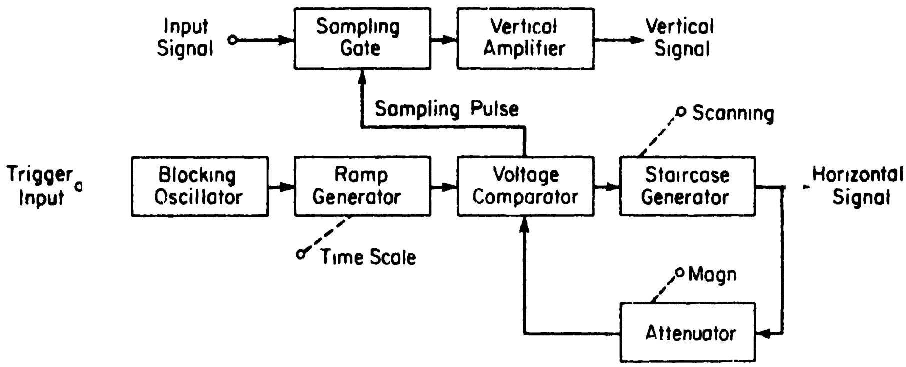


Figure 9-44
Simplified block diagram of the sampling circuitry. (*Courtesy Hewlett-Packard Company.*)

The input waveform, which must be of a repetitive type, is applied to the sampling gate. Sampling pulses momentarily bias the diodes of the balanced sampling gate in the forward direction, thereby briefly connecting the gate input capacitance to the test point. These capacitances are slightly charged toward the voltage level of the input circuit. The capacitor voltage is amplified by the vertical amplifier and applied to the vertical deflection plates. Since the sampling must be synchronized with the input signal frequency, the signal is delayed in the vertical amplifier, allowing the sweep triggering to be done by the input signal. When a trigger pulse is received, the avalanche blocking oscillator (so called because it uses avalanche transistors) starts an exactly linear ramp voltage, which is applied to a voltage comparator. The voltage comparator compares the ramp voltage to the output voltage of a staircase generator. When the two voltages are equal in amplitude, the staircase generator is allowed to advance one step and simultaneously a sampling pulse is applied to the sampling gate. At this moment, a sample of the input voltage is taken, amplified, and applied to the vertical deflection plates.

The real-time horizontal sweep is shown in Fig. 9-43, indicating the horizontal deflection rate of the beam. Notice that the horizontal displacement of the beam is *synchronized* with the trigger pulses which also determine the moment of sampling. The resolution of the final image on the CRT screen is determined by the size of the steps of the staircase generator. The larger these steps, the greater the horizontal distance between the CRT spots that reconstitute the trace.

9-19 Advanced Laboratory CROs

This section offers a brief discussion of some common advanced-type CROs, such as may be found in research laboratories, in development work, or in specialized applications.

The *dual-beam CRO* is built around a special CRT, which produces two completely *independent beams*. This CRT has two electron guns and two independent sets of vertical and horizontal deflection plates. Unusual display versatility is available through the many different combinations of vertical amplifiers and time bases. For example, one channel may be used to show phase shift between two sine-wave signals, while a related function is displayed versus time on the other channel. Slow and fast signals may be observed simultaneously at different sweep speeds or the same signal may be studied at different sweep rates.

A representative example of a dual-beam CRO is shown in Fig. 9-45. This instrument uses two identical vertical amplifiers, each with a sensitivity

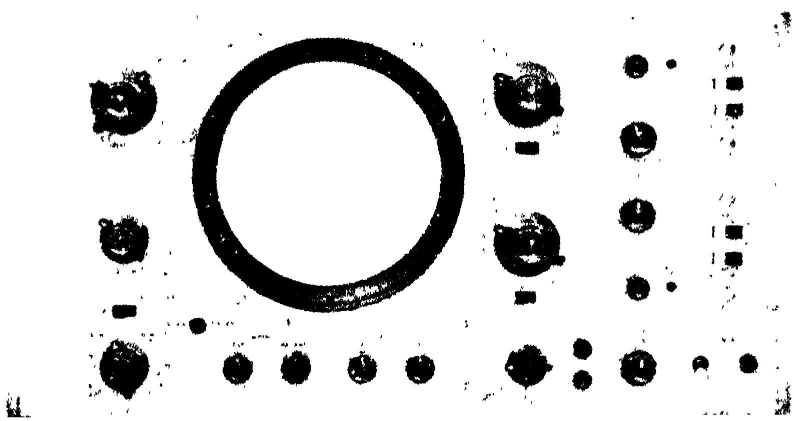


Figure 9-45
Front-panel view of a dual-beam oscilloscope, illustrating the duplication of the two vertical channels. (*Courtesy Hewlett-Packard Company.*)

of 100 μV/cm and a bandwidth of 500 kHz. These two amplifiers may be used to make X-Y plots on single-beam operation by connecting one amplifier to the vertical deflection plates and the other amplifier to the horizontal deflection plates. In this case the unused beam is positioned off the screen and does not interfere with the display.

One time-base generator is available in this CRO, with a range of 1 μs/cm to 5 s/cm and magnified sweep capabilities. Triggering of the sweep generator may be selected from an external source or internally from either vertical channel or from the line voltage. Provision is made for connection of an external horizontal amplifier which may be used on both beams simultaneously or on one beam only, while the other beam is swept by the internal sweep generator.

Figure 9-46 shows the simplified block diagram of the dual beam oscil-

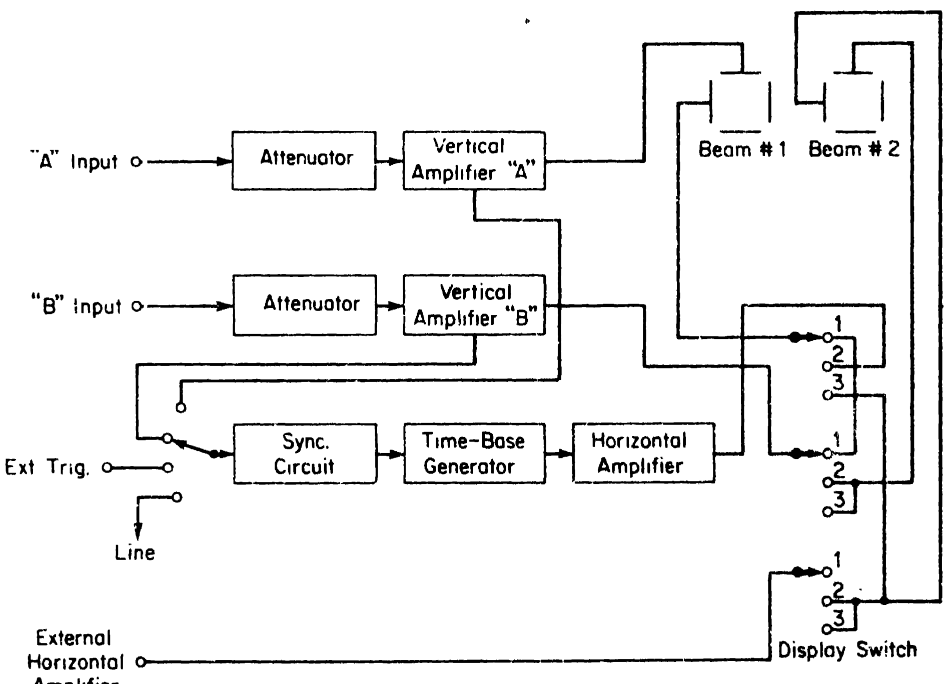
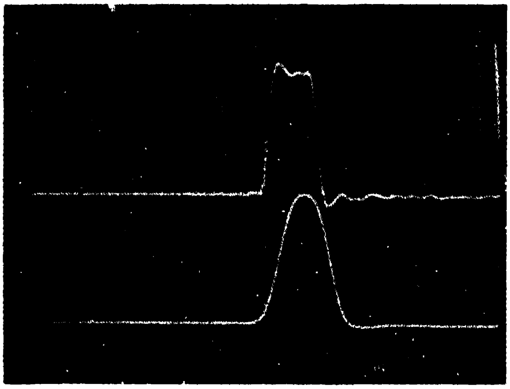


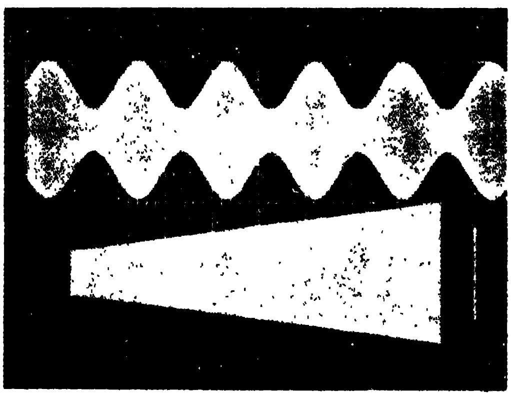
Figure 9-46
Simplified block diagram of a dual-beam oscilloscope showing the display-switching arrangement.

loscope. The figure shows that a number of displays is available through selection of the display-control switch. When the switch is in position 1, the instrument is used for single-beam operation and produces X-Y plots. With the display switch in position 2, both beams are used: one beam with an external horizontal amplifier to produce X-Y plots and the other beam with the internal sweep. In the third position of the display switch, dual plots can be made by using an external horizontal amplifier to drive both beams simultaneously.

Figure 9-47 shows two examples of dual-beam CRO usage. These photographs show the versatility of the dual-beam CRO in presenting simultaneous displays.



(a)



(b)

Figure 9-47
Typical dual-beam displays. (a) Input and output waveform of a pulse-shaping circuit. (b) Modulation patterns. (*Courtesy Tektronix, Inc.*)

The *dual-trace CRO* expands the display capability of the ordinary single-beam CRT to simultaneous portrayal of two separate vertical input signals. The instrument essentially consists of a single-beam CRT, a single time-base generator, and two identical vertical amplifiers (or *channels*) with an *electronic* *switch.* The outputs of the vertical channels are connected to the electric switch via a *mode control switch,* generally allowing two modes of operation. In the *alternate* mode of operation the electronic switch alternately connects each vertical channel to the CRT *after each sweep.* The switching rate of the electronic switch is controlled by the sweep rate of the time-base generator, and the CRT display, therefore, consists of alternating displays of each vertical channel. Each vertical amplifier has its own calibrated input attenuator and positioning controls so that the amplitude of each input signal can be individually adjusted and the two traces positioned on the CRT screen. In the *chopped* mode of operation, the electronic switch is free running with a *chopping rate* of approximately 100 kHz and alternately connects *successive segments* of each vertical channel to the CRT for display, independent of the time-base frequency. The chopped mode of operation is especially useful for simultaneous observation of two low-frequency input signals where the chopping rate of the electronic switch is very much higher than the sweep rate. When the sweep rate approaches the chopping rate, the individual segments of the chopped traces become visible on the CRT screen and the alternate mode of operation would then be preferred.

The *storage CRO* is rapidly becoming one of the most useful tools in the presentation of very slowly swept signals and finds many applications in the mechanical and biomedical fields. In the conventional CRT the persistence of the phosphor ranges from microseconds to perhaps seconds. In applications where the persistence of the screen is smaller than the rate at which the signal sweeps across the screen, the start of the display will have disappeared before the end of the display is written.

With the *variable-persistence* or *storage* CRO, the slowly swept trace can be kept on display continuously by adjusting the persistence of the CRT screen to match the sweep time. Persistence times much greater than a few seconds or even hours, are available, making it possible to *store* events on a CRT screen. The storage CRO uses a special CRT, called the *storage tube.* This special CRT contains all the elements of a conventional CRT, such as the electron gun, the deflection plates, and a phosphor screen, but in addition holds a number of special electrodes. A schematic representation of one type of storage tube is given in Fig. 9-48.

The *storage mesh* or *storage target,* mounted just behind the phosphor screen, is a conductive mesh covered with a highly resistive coating of magnesium fluoride. The *write gun* is a high-energy electron gun, similar to the conventional gun, giving a narrow focused beam which can be deflected and used to write the information to be stored. The write gun etches a positively charged pattern on the storage target by knocking off secondary-emission electrons. Because of the excellent insulating properties of the magnesium fluoride

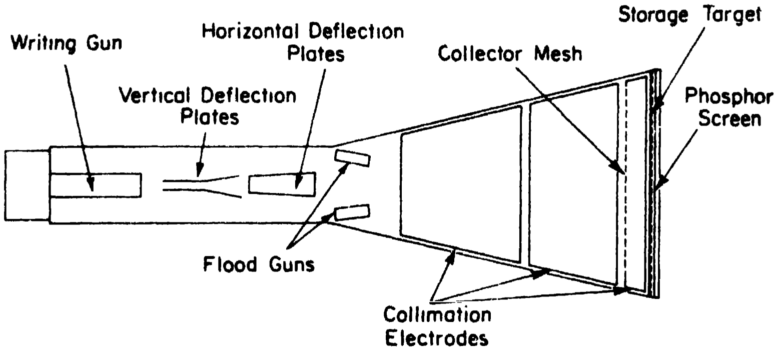


Figure 9-48
Schematic representation of the storage CRT. (*Courtesy Tektronix, Inc.*)

coating, this positively charged pattern remains exactly in the position on the storage target where it was first deposited. The electron beam, which is deflected in the conventional manner both in the horizontal and the vertical direction, therefore traces out the waveform pattern on the storage target.

The stored pattern may be made available for viewing at a later time by the use of two special electron guns, called *flood guns*. The flood guns are placed inside the CRT in a position between the deflection plates and the storage target and they emit low-velocity electrons over a large area toward the entire screen. When the flood guns are switched on (*the viewing mode*), low-energy electrons are sprayed toward the screen. The electron trajectories are adjusted by the collimation electrodes which constitute a low-voltage electrostatic lens system, so that the flood-electrons cover the entire screen area. Most of the flood-electrons are collected by the collector mesh and therefore never reach the phosphor screen. In the area near the stored positive charge on the storage target, the positive field pulls some of the flood-electrons through the storage mesh and these electrons continue on to hit the phosphor. The CRT display therefore will be an exact copy of the pattern which was initially stored on the target and the display will be visible as long as the flood guns continue emission of low-energy electons. To *erase* the pattern which is etched on the storage mesh, a negative voltage is applied to the storage target, neutralizing the stored positive charge.

To obtain *variable persistence,* the erase voltage is applied in the form of pulses instead of a steady dc voltage. By varying the width of these pulses, the rate of erasure is varied. The variable-persistence control on the front panel of the scope is then the width control of the erase-pulse generator.

A typical storage CRO is shown in Fig. 9-49. This instrument does not have the variable-persistence feature, but only the storage mode of operation.

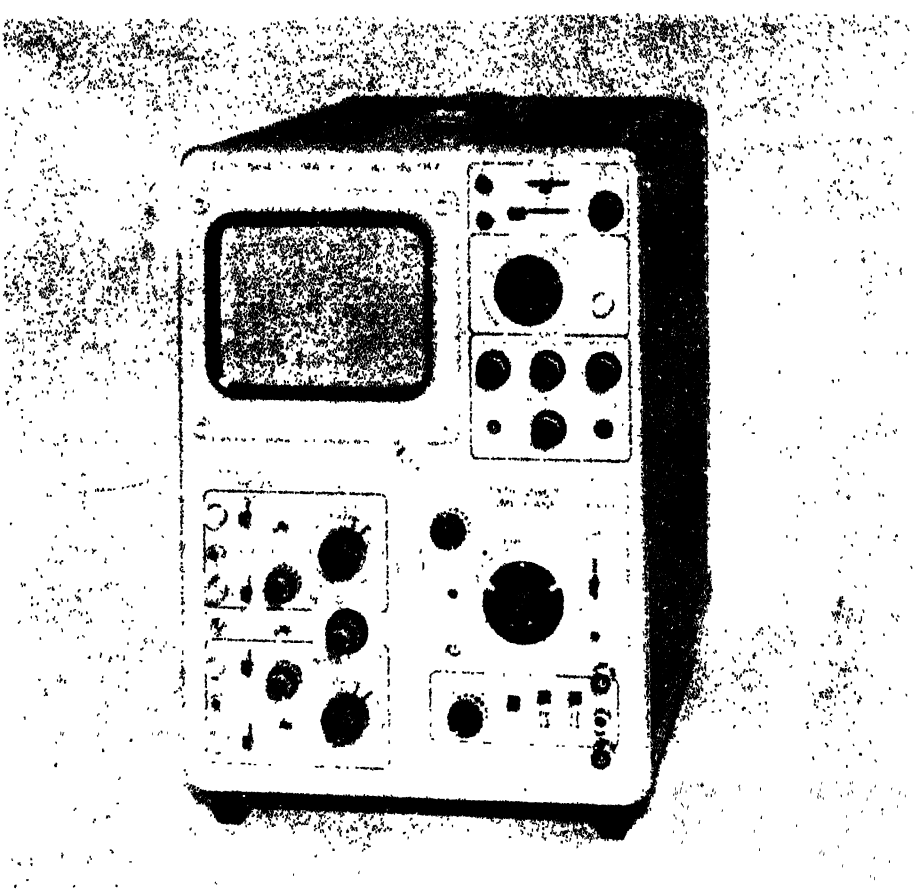


Figure 9-49
A storage oscilloscope using general-purpose plug-in units for the vertical amplifier and time base. The storage controls are shown to the right of the CRT and enable the operator to use the scope as a normal nonstore instrument or as a storage scope. (*Courtesy Tektronix, Inc.*)

The *digital readout CRO* introduces the concept of providing digital readout of signal information, such as voltage or time, in addition to the conventional CRT display. The digital readout CRO basically consists of a conventional high-speed laboratory CRO plus an electronic counter, both contained in one cabinet. The circuitry of both units is connected by means of a display-logic control, allowing measurements to be made with great speed and accuracy. The digital readout CRO gives readout of rise-time, amplitude, time-difference, depending on the positions of the various operating controls, such as TIME/DIV, AMPLITUDE/DIV, and PROGRAM. (A photograph of a representative digital readout CRO is shown in Fig. 9-50.) The input wave-

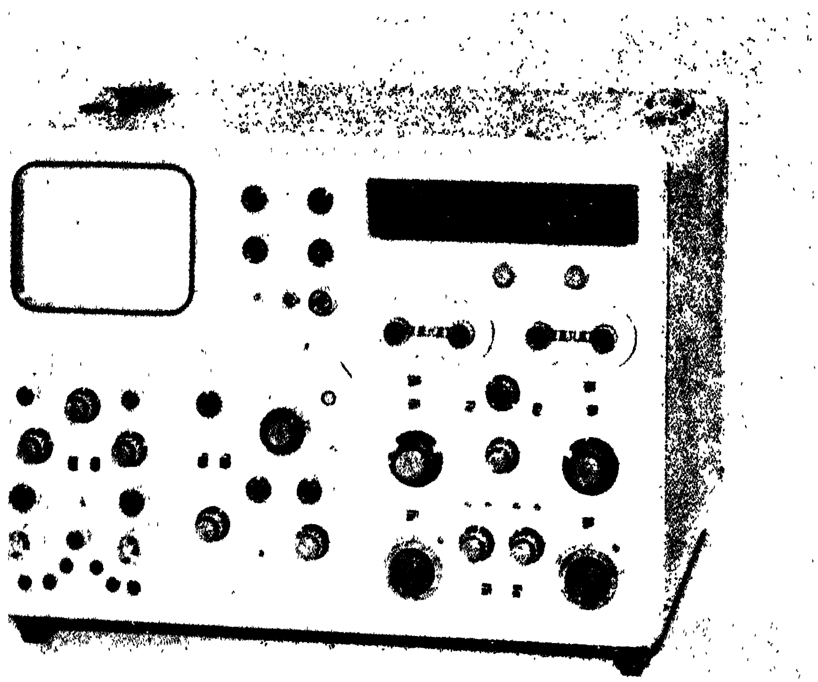


Figure 9-50
A digital readout oscilloscope. (*Courtesy Tektronix, Inc.*)

form is sampled by means of a sampling unit. With each repetition of the input signal, the sampling unit strobes one point at a time, a little later than the previous sample. (The process of advancing the sampling instant in fixed increments is called *strobing*). A reconstructed replica, much slower than the original input waveform, is reproduced on the CRT as a point-by-point plot of amplitude versus time.

The equivalent time between each sample depends on the number of samples taken per centimeter of the displayed waveform and on the sweep time per centimeter. For example, a sweep rate of 1 ns/cm and a sampling rate of 100 samples/cm, gives a time of 10 ps per sample. By counting the number of samples taken between two selected points on the waveform, the time between these points can be determined. Counter-circuits, driving nixie tubes, display the measured time interval. Figure 9-51 shows, in block diagram form, the operation of the digital readout scope when measuring time. The input waveform, strobed by the sampling unit, is applied to the CRT and is displayed in point-by-point form. Two intensified portions of the CRT trace identify 0 per cent and 10 per cent zones. Each zone can be positioned to cover any

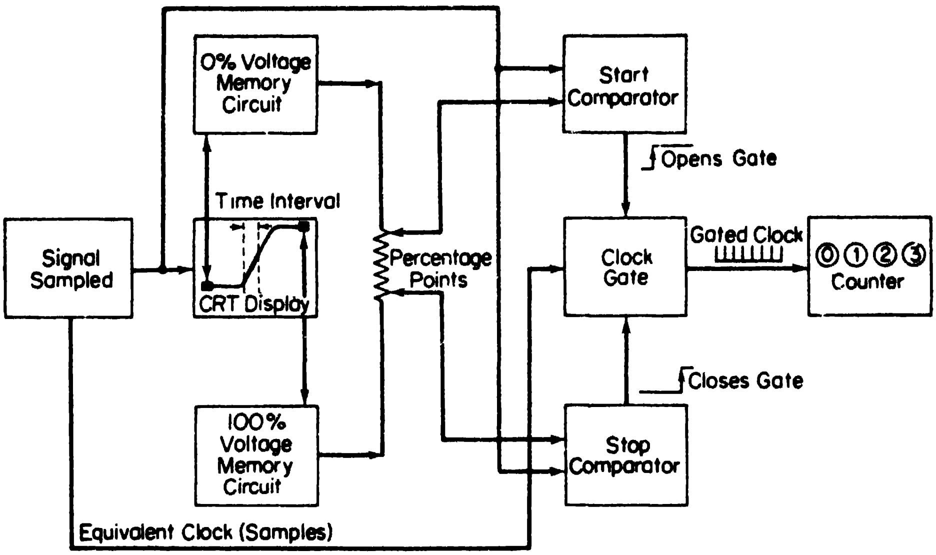


Figure 9-51
Block diagram of the digital readout oscilloscope when it is measuring time.

portion of the CRT display. The input waveform amplitudes, corresponding to the intensified zones on the display, are stored in voltage-memory circuits. Voltage divider taps between the 0 per cent and 100 per cent memory voltages are set for start-and-stop timing at selected percentage points of the waveform to be measured. Coincidence of the input waveform amplitude with the selected reference amplitude (percentage selected) is sensed by voltage comparators which open and close the clock-gate to the digital counter. The number of clock pulses are read out digitally in nanoseconds, microseconds, milliseconds, or seconds on the nixie display tubes. In the time-measurement application, the clock pulses consist of the actual samples taken.

Figure 9-52 shows in block diagram form the operation of the digital readout CRO when measuring voltage. The input waveform again is strobed by the sampling unit and displayed on the CRT. The voltage memory circuits provide intensified portions on the CRT display and identify the 0 per cent and 100 per cent reference voltages. The 0 per cent reference voltage is applied to the start comparator and the 100 per cent reference voltage is applied to the stop comparator. A linear ramp generator is connected to both comparators. For the period of time that the linear ramp voltage is at values between the 0 per cent and 100 per cent amplitudes, as set by the voltage memory circuits, 1-MHz clock pulses are gated to the digital counter-circuit. The number of clock pulses is directly proportional to the voltage between the selected

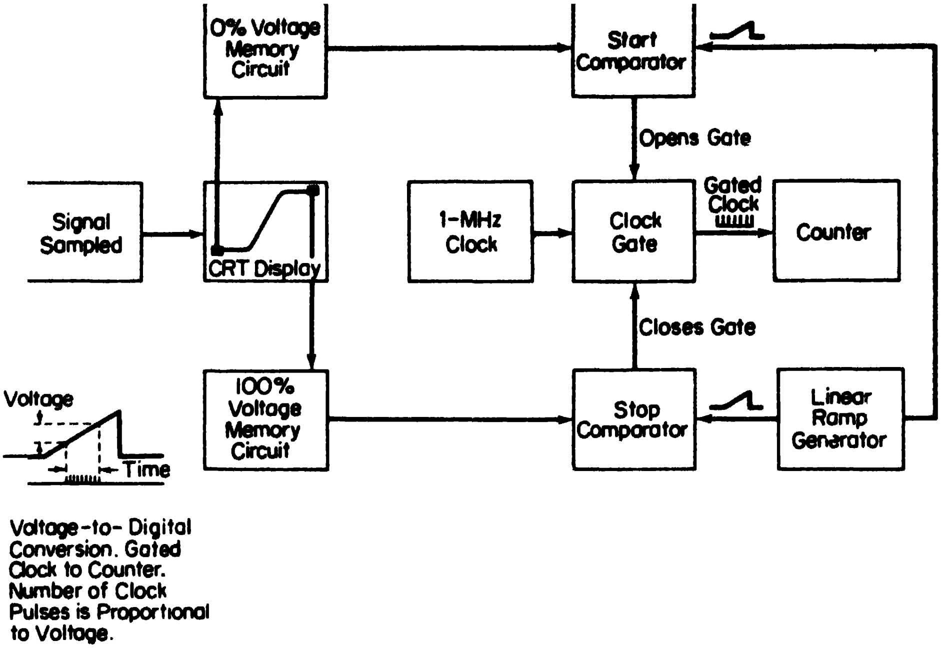


Figure 9-52
Block diagram of the digital readout oscilloscope when it is measuring voltage.

measurement points and is read out in millivolts or volts by the nixie tube display.

Manufacturers of laboratory type CROs tend to supply special features in the form of a number of different arrangements which can be plugged into the main CRO frame. The *main frame* of the plug-in type CRO consists of the CRT with its associated circuitry for focusing, intensity, etc., the low-voltage and high-voltage power supplies, and the final driver stages for the horizontal and vertical deflection plates. Different *plug-in units* are inserted into the main frame and are connected to the vertical and horizontal driver stages, respectively. The main frame, together with a selection of various plug-in units, can therefore be used to perform a variety of special measurement functions, providing extreme versatility at moderate investment cost. Manufacturers of laboratory-type CROs of the plug-in variety make many different types of plug-in units available. The vertical channel units are usually vertical deflection amplifiers, sometimes with special performance characteristics, although many applications now include the use of sophisticated systems for driving the vertical deflection plates. Some of the more common vertical channel plug-in units are

(a) Dual-trace dc unit.
(b) Four-trace unit.
(c) High-gain differential unit.
(d) Fast-rise high-gain unit.
(e) Strain-gage unit.
(f) Operational-amplifier unit.
(g) Spectrum-analyzer unit.

The last three units are examples of specially developed vertical channel units, designed with specific applications in mind.

The horizontal channel plug-in units are usually time-base generators with different capabilities, although here again different systems have been developed for specific applications. The more common horizontal channel units are (a) single-sweep unit, (b) sweep-delay unit, (c) sampling-sweep unit.

Details about the performance characteristics of the various plug-in units may be found in the manufacturer's manuals and instruction material.

2. A laboratory-type CRO generally uses an "unblanking cathode follower" in its CRT circuit. Explain its function and describe its operation.

3. Discuss the relation between wide-band performance and high sensitivity of a general-purpose CRO. Make suggestions about steps to be taken to improve the gain-bandwidth performance of a CRO.

4. A simple RC time-base generator generally delivers a nonlinear ramp voltage which may be unsuitable for the time base of a laboratory-type CRO. Suggest several methods which may be used to improve the time-base linearity and explain the principles involved in these methods of linearization.

5. What is the reason for using a delay line in the vertical-deflection system of a laboratory-type CRO?

6. The input attenuator in the vertical amplifier of a general-purpose CRO is generally followed by an emitter-follower or cathode-follower circuit. Suggest three possible reasons for using this circuit.

7. Construct a block diagram of a general-purpose CRO. Label all the blocks and show the waveforms entering and leaving each block (when applicable) assuming a sinusoidal voltage applied to the input of the vertical amplifier.

8. The gain, frequency response, and phase shift of a 10-W audio amplifier (frequency range 20 Hz–20 kHz) are to be measured, using a CRO as the basic measuring tool. An audio oscillator and several types of ac and dc voltage and current meters are available. Suggest a measurement technique, indicating the equipment required to perform each measurement. The result of each measurement is to be presented in graphical form. Suggest a suitable way of presenting the measurement results and approximately sketch the expected shape of each graph.

9. The input attenuator of Fig. 9-14 is used in a general-purpose CRO requiring a time constant $\tau = C_2R_2$ of 4 μs. Calculate the values of C_1, C_2, R_1, and R_2, if the sum of R_1 and R_2 is 2 MΩ.

10. The calibrated time base of a laboratory-type CRO is set at 0.2 mV/cm. The horizontal display switch is in the "5× magnified" position. A sinusoidal waveform of unknown frequency is applied to the input terminals of the vertical amplifier and produces $3\frac{1}{2}$ cycles over a sweep width of 10 cm. Determine the frequency of the input voltage.

11. A Lissajous pattern is produced by applying sinusoidal voltages to the vertical and horizontal input terminals of a CRO. The pattern makes five tangencies with the vertical and three with the horizontal. Calculate the frequency of the signal applied to the vertical amplifier if the frequency of the horizontal input voltage is 3 kHz.

12. Voltage V_1 is applied to the vertical input and voltage V_2 to the hori-

zontal input of a CRO. The Lissajous pattern is symmetrical about the vertical and horizontal axes, with V_1 and V_2 having the same frequency. The slope of the major axis is positive, with a maximum vertical value of 2.5 divisions. The point where the figure crosses the vertical axis is 1.2 divisions high. Calculate the possible phase angles of V_2 with respect to V_1.

13. The transit time of an electron through the deflection plates is one of the factors determining the frequency limits of a CRO. Assuming that this transit time should be kept below 0.1 cycle, determine the upper frequency limit of an electrostatic deflection system with plates of 1-cm length if electrons enter with velocities corresponding to a kinetic energy of 1,000 eV.

14. Calculate the deflection sensitivity, S, for the CRO of Problem 13 if $L = 20$ cm and the distance d between the plates is 5 mm.

15. Which factor(s) can be changed if the upper frequency limit of the CRO of Problem 13 is to be doubled, without affecting the deflection sensitivity calculated in Problem 14?

16. The accelerating voltage of a CRT is 1,000 V. A sinusoidal voltage is applied to a set of deflecting plates whose axial length is 1 cm. Calculate (a) the maximum frequency of the sinusoidal voltage if the electrons are not to remain between the plates for more than one half cycle, (b) the time, in μs, that an electron remains in the region of the deflection plates if the frequency of the applied voltage is 60 Hz.

17. The calibrated time base of a laboratory-type CRO is set to 0.1 ms/cm. The sweep width is 10 cm. Assuming that the sweep voltage is an ideal ramp with zero retrace time, sketch the waveform patterns resulting from applying the following signals to the vertical amplifier input terminals:

(a) sine wave with a frequency of 5 kHz
(b) sine wave with a period of 0.5 ms
(c) cosine wave with a period of 2 ms
(d) square wave with a frequency of 10 kHz
(e) pulse with a repetition rate of 2,000 cycles per second and a duty cycle of 25 per cent

18. Explain the function of each of the following CRO controls and indicate in which section of the CRO circuitry these controls are found:

(a) focus
(b) horizontal position
(c) sweep vernier
(d) external horizontal input
(e) Z-axis modulation

19. The horizontal amplifier section of a CRO usually provides for several connections to trigger the time base. Give one application example for each of the following trigger input settings:

(a) internal
(b) line
(c) external

20. With the aid of simple waveform sketches, explain the effect of excessive sync amplitude on the displayed waveform pattern.

TEN

ELECTRONIC INSTRUMENTS FOR THE MEASUREMENT OF VOLTAGE, CURRENT, RESISTANCE, AND OTHER CIRCUIT PARAMETERS

10-1 Introduction

Electronic voltmeters, ammeters, and ohmmeters use amplifiers, rectifiers, and other circuits to generate a current proportional to the quantity being measured. This current then drives a conventional meter movement of the type discussed in Chapter 4. It is interesting to note that many modern electronic voltmeters use taut-band suspension movements (Sec. 4-3) instead of the more conventional pivot-and-jewel mechanisms. Instruments which use meter movements to indicate the magnitude of the quantity under measurement on a continuous scale are sometimes called *analog instruments*.

When the result of the measurement is displayed in *discrete* intervals or numerals (instead of by a pointer deflection on a continuous scale), it is termed a *digital* indication. Direct numeral readout reduces human error and tedium; it eliminates parallax and other reading errors, and it increases reading speed. Additional features in modern digital instruments, such as automatic polarity and range-changing facilities, further reduce measurement error and possible instrument damage due to accidental overloading.

Digital instruments are available to measure both dc and ac voltage, current, and resistance. Other physical variables can be measured by using suitable transducers. Many digital instruments have auxiliary output provisions to make permanent records of measurement results using printers, card or tape punches, or magnetic tape equipment. With data already in digital form, they may then be processed without loss of accuracy.

In this chapter some of the more common *analog* instruments are first discussed; then some of the major digital devices are analyzed in block-diagram form. The last part of the chapter deals with some of the special-purpose instruments used for measuring circuit or component parameters.

10-2 Electronic DC Voltmeters

The dc electronic voltmeter represents a straightforward application of electronics to measuring instruments. This instrument usually consists of a normal dc meter movement preceded by a dc amplifier consisting of one or more stages. The dc amplifiers used in electronic voltmeters may be classified into two main groups: (a) *direct-coupled dc amplifiers;* (b) *chopper-type dc amplifiers.*

Direct-coupled dc amplifiers are attractive because they are economical; they are commonly found in lower-priced dc voltmeters. Figure 10-1 shows a

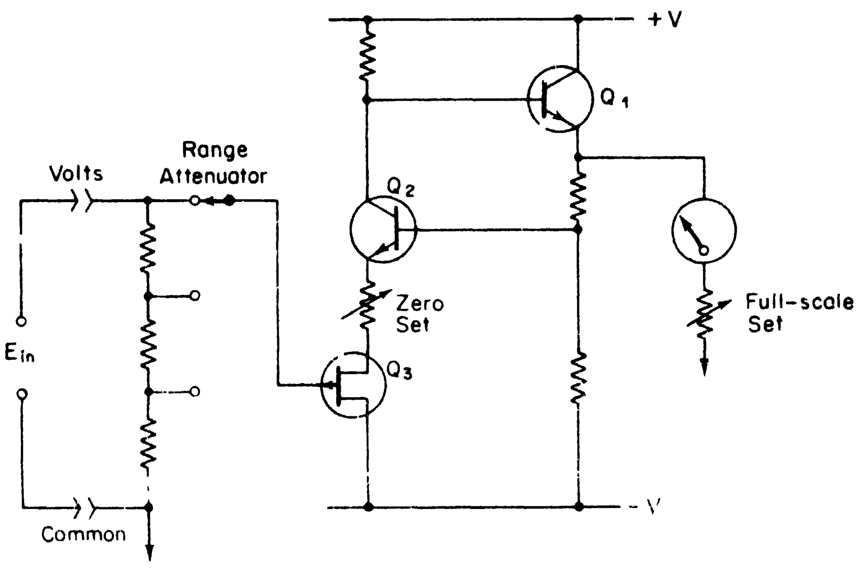


Figure 10-1
A basic dc voltmeter circuit.

simplified partial schematic of a direct-coupled dc amplifier with an indicating meter.

The dc input voltage is applied to a RANGE attenuator which is a calibrated front-panel control. The attenuator is necessary to provide input voltage levels which can be accommodated by the dc amplifier. The input stage of the amplifier consists of a field-effect transistor (FET). The FET is a popular choice because of its high input impedance which effectively isolates the meter circuit from the circuit under measurement.

The two transistors, Q_1 and Q_2, form a direct-coupled dc amplifier driving the meter movement. Provided that the amplifier operates within the limits of its dynamic range, the deflection of the meter movement is directly proportional to the magnitude of the applied input voltage. The gain of the dc amplifier allows the instruments to be used for the measurement of dc voltages in the millivolt range. The amplifier has the added advantage that accidental input overloads do not burn out the meter because the amplifier saturates, which limits the maximum current though the meter. The input impedance of this dc voltmeter is high enough to make it unnecessary to correct for the loading effect on the circuit under measurement.

Instruments in the microvolt range of measurement require a high-gain dc amplifier in order to supply sufficient current for driving the meter movement. To avoid the drift problems usually associated with direct-coupled dc amplifiers, the *chopper-type dc amplifier* is commonly found in high-sensitivity voltmeters. In the chopper amplifier, the dc input voltage is converted into an ac voltage before it is applied to the amplifier input terminals. The amplified ac voltage is then converted back into a dc voltage, proportional to the original input signal.

The block diagram of Fig. 10-2 illustrates the operation of the chopper-type amplifier used in a millivolt dc voltmeter. Photodiodes are used as nonmechanical choppers for modulation (conversion from dc into ac) and demodulation (conversion from ac back into dc). A photoconductor has a low resistance, from a few hundred to a few thousand ohms, when it is illuminated by a neon or incandescent lamp. The photoconductor resistance increases sharply, usually to several megohms, when it is not illuminated. In the circuit of Fig. 10-2, an oscillator drives two neon lamps into illumination on alternate half cycles of oscillation. Each neon lamp illuminates one photoconductor in the input circuit of the amplifier and one in the output circuit. The two photodiodes in the input circuit form a series-shunt half-wave *modulator* or *chopper*. Together they act like a switch across the input of the amplifier, alternately opening and closing at a rate determined by the frequency of the neon oscillator.

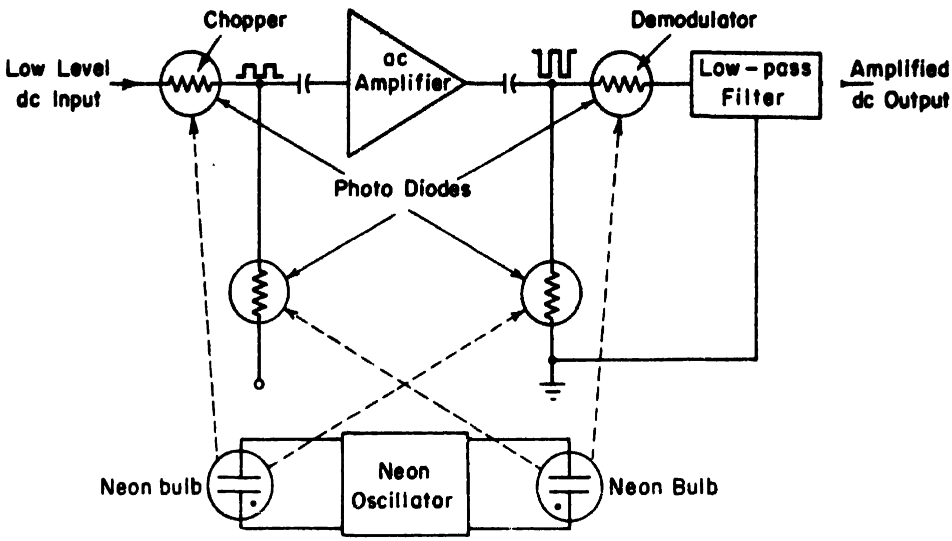


Figure 10-2
Nonmechanical photoconductive chopper.

The input signal to the amplifier is a square-wave voltage of amplitude proportional to the input voltage with frequency equal to the oscillator frequency. (This frequency is limited to a few hundred hertz because the transition time between the high- and low-resistance conditions of the photodiodes limits the chopping rate.) The ac amplifier thus delivers an amplified square wave at its output terminals. The two photodiodes in the amplifier output circuit, operating in antisynchronism with the input chopper, recover the dc signal by their demodulating action. The dc output signal is then passed through a low-pass filter to remove any residual ac components and is finally applied to the meter movement. Notice that the amplifier is not stabilized by the chopper diodes. The choppers only eliminate the need for a high-gain dc amplifier with its inherent drift and stability problems.

The input impedance of chopper-amplifier dc voltmeter is usually in the order of 10 MΩ or higher, except on the very low input ranges. To eliminate measurement errors caused by high source impedances, a *nulling* feature is sometimes included in the meter circuitry. This extremely useful addition places a *bucking* voltage in series with the input. A front panel control allows the user to null the input voltage with the bucking voltage. When a null is indicated on the meter, the bucking voltage equals the input voltage and no current is drawn from the source. The meter therefore represents an infinite input impedance and eliminates any loading effect. The function switch then allows the input to be disconnected from the meter circuit, and the bucking voltage (which is equal to the input voltage) is displayed on the meter.

10-3 Electronic AC Voltmeters

Electronic ac voltmeters are essentially identical to dc voltmeters except that the ac signal may be rectified before it can be applied to the meter movement. When the ac input voltage is rectified before amplification, a dc amplifier is needed; when rectification takes place after amplification, an ac amplifier can be used. *Both* systems find application in ac voltmeters. (Since the exact circuit configuration varies from one instrument to another, see the manufacturer's manual supplied with the electronic voltmeter.)

AC voltmeters generally fall into three broad categories: (a) average-responding voltmeters; (b) peak-responding voltmeters; (c) true rms-responding voltmeters. The only difference between an average-responding and a peak-responding voltmeter is in the recitfier circuit. The two circuits are given in Fig. 10-3.

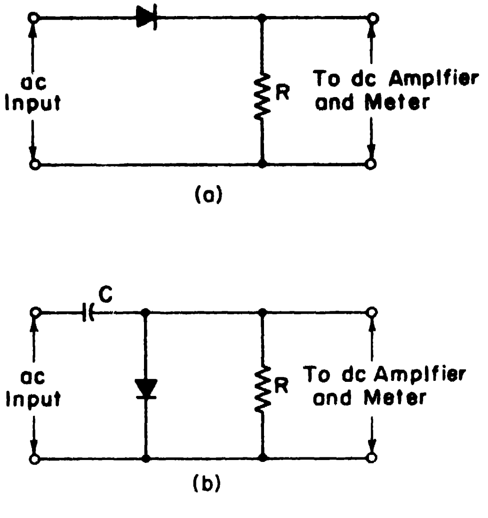


Figure 10-3
Rectifier arrangements used in ac voltmeters. (a) Series-connected diode, providing half-wave rectification for an average-reading voltmeter. (b) Shunt-connected diode for a peak-reading voltmeter. Either circuit is usually contained in the instrument probe.

The series-connected diode of Fig. 10-3(a) provides half-wave rectification of the ac input signal; the average value of the half-wave voltage is developed across the resistor and applied to the meter circuit. In the peak-reading circuit of Fig. 10-3(b) the rectifier charges the small capacitor to the peak value of the applied input voltage. Provided that no current is drawn from the input source, the peak voltage is applied to the meter circuit. The meter scale is calibrated in terms of the average value or the peak value of the applied input waveform.

Although the rms voltage is usually of principal interest, ac voltmeters are generally of the average-responding type, with the meter scale *calibrated* in terms of the rms value of a sine wave. Since so many waveforms in electronics are sinusoidal, the average-responding meter with an rms-calibrated scale is an entirely satisfactory instrument. In point of fact, it is much less expensive for the same accuracy and bandwidth than a *true* rms-responding instrument.

The effective, or rms, value of a voltage wave which has equal positive and negative excursions is related to the average value by the *form factor*. The form factor is the ratio of the rms to the average value of a waveform. It can be expressed as

$$\text{form factor} = k = \frac{\sqrt{\dfrac{1}{T}\displaystyle\int_0^T e^2\,dt}}{\dfrac{2}{T}\displaystyle\int_0^{T/2} e\,dt} \tag{10-1}$$

For a sinusoidal waveform the form factor equals

$$k_{\text{sinusoidal}} = \frac{E_{\text{rms}}}{E_{\text{av}}} = \frac{0.707E_m}{0.636E_m} = 1.11 \tag{10-2}$$

Therefore, when an average responding voltmeter has scale markings corresponding to the rms value of the applied sinusoidal input waveform, those markings are actually corrected by a factor of 1.11 from the true (average) value of applied voltage.

Nonsinusoidal waveforms, when applied to this voltmeter, will cause the meter to read either high or low, depending on the form factor of the waveform. An illustration of the effect of nonsinusoidal waveforms on ac voltmeters is given in Examples 10-1 and 10-2:

Example 10-1: A symmetrical square-wave voltage is applied to an average-responding ac voltmeter having a scale calibrated in terms of the rms value of a sine wave. Calculate: (a) the form factor of the square-wave voltage, (b) the error in the meter indication.

SOLUTION:

(a)

$$E_{\text{rms}} = \sqrt{\frac{1}{T}\int_0^T e^2\,dt} = E_m$$

$$E_{\text{av}} = \frac{2}{T}\int_0^{T/2} e\,dt = E_m$$

$$k = \frac{E_m}{E_m} = 1$$

(b) The meter deflection is caused by the average value of the applied voltage. The scale markings represent the rms values of a sine-wave voltage, where $E_{rms} = k \times E_{av} = 1.11\,E_{av}$. For the square-wave voltage, $E_{rms} = E_{av}$, since $k = 1$. Therefore, the meter indication for the square-wave voltage is *high* by a factor $k_{\text{sine wave}}/k_{\text{square wave}} = 1.11$. The percentage error equals

$$\frac{1.11 - 1}{1} \times 100\% = 11\%$$

Example 10-2: Repeat Example 10-1 if the voltage applied to the meter consists of a sawtooth waveform with a peak value of 150 V and a period of 3 s.

SOLUTION: (a) The analytical expression for the sawtooth waveform between the limits of $t = 0$ and $t = T = 3$ s is $e = 50t$ V. Therefore,

$$E_{rms} = \sqrt{\frac{1}{T}\int_0^T e^2\,dt} = \sqrt{\frac{1}{3}\int_0^3 (50\,t)^2\,dt} = 50\sqrt{3}\text{ V}$$

$$E_{av} = \frac{1}{T}\int_0^T e\,dt = \frac{1}{3}\int_0^3 50\,t\,dt = 75\text{ V}$$

$$\text{Form factor, } k = \frac{50\sqrt{3}}{75} = 1.155$$

(b) The ratio of the two form factors is

$$\frac{k_{\text{sine wave}}}{k_{\text{sawtooth}}} = \frac{1.11}{1.155} = 0.961$$

The meter indication is *low* by a factor of 0.961. The percentage error equals:

$$\frac{0.961 - 1}{1} \times 100\% = -3.9\%$$

Examples 10-1 and 10-2 point out that any departure from a true sinusoidal waveform may cause an appreciable error in the result of the measurement. Complex waveforms are therefore most accurately measured with an *rms-responding voltmeter*. This instrument produces an 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 feeding an amplified version of the input waveform to the heater element of a *thermocouple* whose output voltage is then proportional to E_{rms}^2.

One difficulty with this technique is that the thermocouple is often nonlinear in its behavior. This difficulty is overcome in some instruments by using two couples mounted in the same thermal environment. The block diagram of Fig. 10-4 shows one form of a true rms-responding voltmeter using two thermocouples. The effect of the nonlinear behavior of the couple in the input circuit (*the measuring thermocouple*) is canceled by similar nonlinear effects of the couple in the feedback circuit (*the balancing thermocouple*). The two couple elements form part of a bridge circuit connected to the input of a dc amplifier.

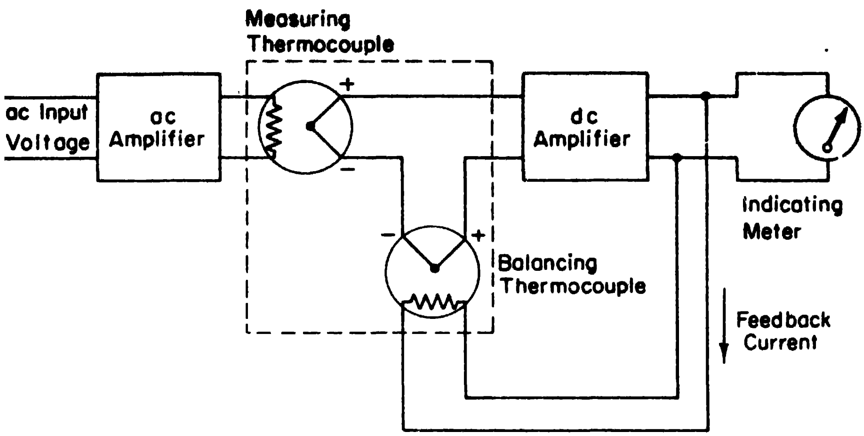


Figure 10-4
Block diagram of a true rms-reading voltmeter. The measuring and balancing thermocouples are located in the same thermal environment.

The unknown ac input voltage is amplified and applied to the heating element of the measuring thermocouple. The application of heat produces an output voltage which upsets the balance of the bridge. The unbalance voltage is amplified by the dc amplifier and fed back to the heating element of the balancing thermocouple. Bridge balance will be reestablished when the feedback current delivers sufficient heat to the balancing thermocouple, so that the voltage outputs of both couples are the same. At this point the dc current in the heating element of the feedback couple is equal to the ac current in the input couple. This dc current is therefore directly proportional to the effective, or rms, value of the input voltage and is indicated on the meter movement in the output circuit of the dc amplifier. The true rms value is measured quite independently of the waveform of the ac signal, provided that the peak excursions of the waveform do not exceed the dynamic range of the ac amplifier.

A typical laboratory-type rms-responding voltmeter provides accurate rms readings of complex waveforms having a *crest factor* (ratio of peak value to rms value) of 10/1. At 10 per cent of full-scale meter deflection, where there is less chance of amplifier saturation, waveforms with crest factors as high as 100/1 could be accommodated. Voltages throughout a range of 100 μV to 300 V within a frequency range of 10 Hz to 10 MHz may be measured with most good instruments.

10-4 The Balanced-Bridge Voltmeter (VTVM or TVM)

One of the basic laboratory and general-purpose instruments used to measure *both* dc and ac voltages is the vacuum tube voltmeter (VTVM) or its transistorized version (TVM). This instrument usually contains additional circuitry for measuring resistance, making it a truly versatile meter.

Although the actual circuits used in this type of meter may vary from one instrument to the next, the VTVM or TVM generally comprises a number of major operational blocks. These are (a) the balanced-bridge dc amplifier, together with a dc meter circuit; (b) the rectifier section to convert an ac input signal into an equivalent dc value; (c) an internal battery and additional circuitry for measuring resistance; (d) switching arrangements to select the function of the meter and the range of measurement. In addition, each instrument contains a built-in power supply of conventional design.

Figure 10-5 shows a functional schematic of a balanced-bridge dc amplifier,

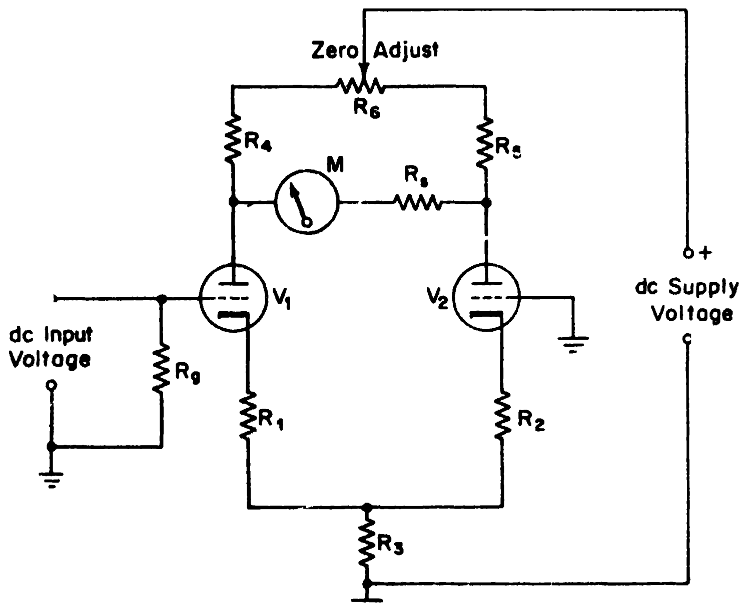


Figure 10-5
Functional schematic of a vacuum tube voltmeter.

using vacuum tubes; the TVM uses essentially the same circuit configuration, with transistors replacing the vacuum tubes.

The circuit shown here consists of two triodes, V_1 and V_2, which, together with their cathode resistors, R_3 and R_4, form the lower arms of the bridge circuit. The upper arms of the bridge consist of anode resistors, R_1 and R_2, together with a portion of the *zero adjust* control, R_6. The wiper of this control is connected to the positive terminal of the dc power supply. Resistor R_3 is placed in series with the cathode resistors and is returned to the negative side of the power supply. R_6 has a comparatively high value so that it controls the cathode current to a large extent. The dc meter is connected to the anodes of the triodes, at opposite corners of the bridge. The control grid of V_2 is connected to ground.

The dc voltage under measurement is applied to the control grid of the *input tube*, V_1. When a positive voltage is applied to the grid of the input tube, its plate current increases and raises the junction of R_3, R_4, and R_5 to a higher positive value. Since the grid of the *balancing tube*, V_2, is connected to ground, the rise in voltage at the junction of its cathode resistor, R_4, and the other two resistors, R_3 and R_5, effectively increases the negative bias of V_2, and its plate current decreases. Since the common portion of the cathode resistance, R_5, carries the total anode current of V_1 and V_2, the voltage drop across it remains approximately constant. Its only function is to limit the current through the bridge arms.

The increase in plate current of input tube, V_1, causes a drop in its anode potential. Simultaneously, the decrease in the plate current of the balancing tube, V_2, causes its anode voltage to rise. The difference in potential between the two anodes is registered by the meter movement connected between the two anodes. The meter deflection is linear, provided that the two triodes operate in the linear region of their characteristics.

The circuit of Fig. 10-5 is extremely sensitive, responding to small input voltages. This occurs mainly because both triodes operate as dc amplifiers, where a small change in grid potential causes a large change in anode potential. The second reason for the high sensitivity is that the increase in plate current of one tube is accompanied by a simultaneous decrease in plate current of the second tube, thereby doubling the potential difference between the two anodes.

When the input to the bridge circuit is zero volts (grid 1 at ground potential), the two triodes operate under identical conditions, and their anode currents should be the same, causing the meter to indicate zero deflection. In practice, however, small differences in the operating characteristics of the vacuum tubes and tolerances in the resistors of the bridge circuit may cause a slight unbalance between the two anode currents, and the meter will indicate a small initial deflection. The meter is initially set to zero by the *zero adjust* resistor, R_6, which allows a limited adjustment of the relative values of the two anode resistors.

The meter is rather insensitive to excess input voltages since the unbalance caused by such a condition rapidly cuts off one or the other of the tubes, limiting the maximum voltage appearing across the meter. The maximum input voltage to the bridge circuit is limited by the operating range of the vacuum tubes. For linear operation this range usually extends to about 3–5 V. The basic

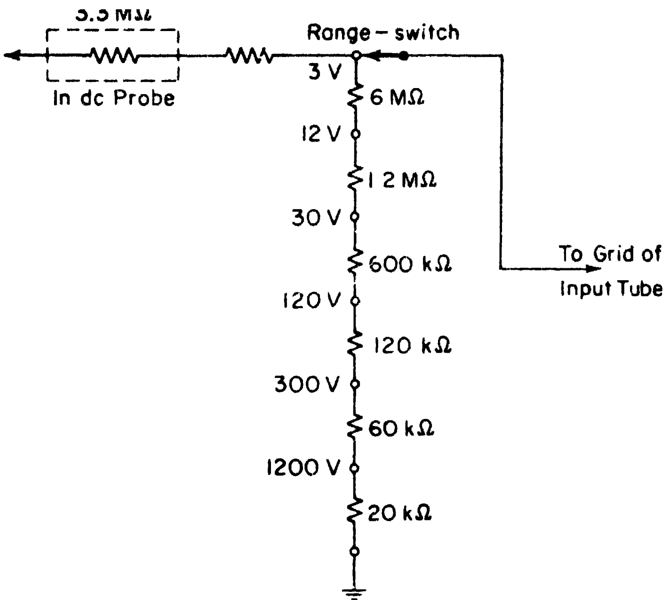


Figure 10-6
Voltage divider for the dc volts function of a VTVM. Selection of the dc volt ranges is accomplished by setting the *range* switch on the front panel of the instrument.

input range can easily be expanded by using an input attenuator or *range switch*, such as the one shown in Fig. 10-6.

The unknown dc voltage is applied through the resistor in the probe body directly to the divider network as shown. Thus, with the range selector of the VTVM in the 3-V position, the input to the grid of the input tube is taken across 8 MΩ of the total resistance of 12 MΩ and the sensitivity of the circuit is designed to indicate 3 V full scale with this arrangement. When the range selector is set to the 30-V position, the input to the grid of the bridge is across a smaller portion of the total resistance (800 kΩ instead of 8 MΩ). For full-scale deflection, the meter circuit now requires an input voltage of 30 V, thereby allowing the same original voltage at the input to the bridge circuit.

A typical circuit for measuring ac voltage is shown in Fig. 10-7. When the *function switch* of the VTVM is in the normal position, the ac voltage under measurement is rectified by one-half of the twin-triode and the resultant dc voltage is applied to the input attenuator. In the *normal position* of the function switch, the meter is calibrated to read in terms of the rms value of the applied ac input voltage.

When the function switch is in the *peak-to-peak* position, both sections of the twin-diode are used. The ac input voltage is rectified by the voltage doubler

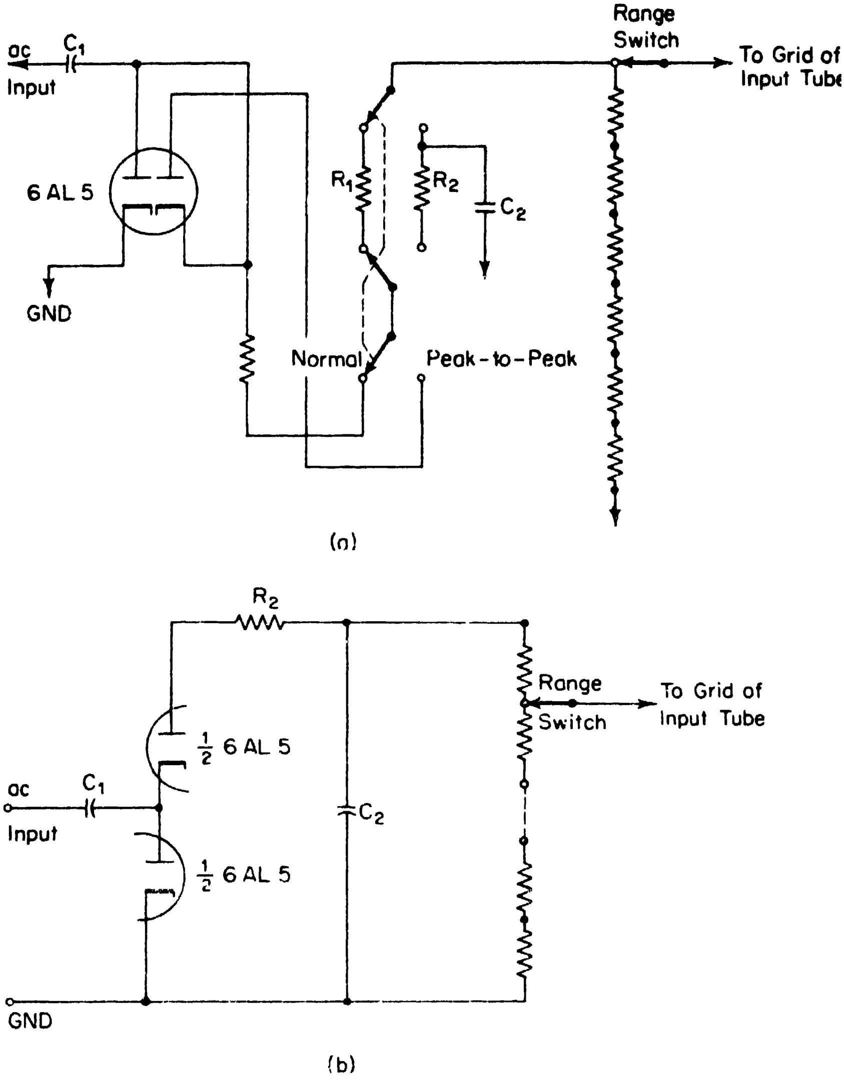


Figure 10-7

The ac volt function of the VTVM. (a) One half of the diode operates as a rectifier in the *normal* position of the mode switch. (b) In the *peak-to-peak* position, capacitor C_2 and the other half of the diode produce a rectified voltage proportional to the peak-to-peak value by voltage-doubling action.

action of the input capacitor, C_1, the two diodes and the output capacitor, C_2. The resultant peak-to-peak voltage is applied to the input attenuator of Fig. 10-6. Note that the same input attenuator is used for measurement of dc voltage, ac rms voltage, and ac peak-to-peak voltage.

The VTVM measures resistance by placing the unknown resistor in series with an internal battery and a new divider network, used only when the function switch is set to the *ohms* position. The voltage drop across the unknown resistance is applied to the control grid of the input tube, and the resulting bridge unbalance is an indication of the magnitude of the unknown resistance.

The simplified circuit for operation of the VTVM in the ohms function is shown in Fig. 10-8. When the unknown resistance is connected between the

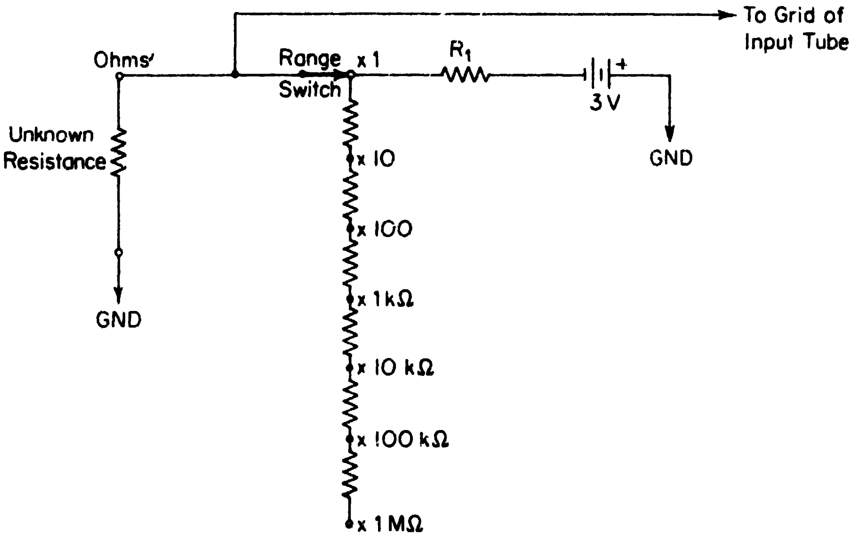


Figure 10-8
The ohms function of the VTVM. The voltage drop across the unknown resistor, produced by the meter's internal battery, is applied to the grid of the balanced bridge.

ohms and *ground* terminals, the battery supplies current through the unknown resistor and that part of the divider selected by the range switch. The voltage drop across the unknown is applied to the input of the bridge circuit, causing an indication on the meter. The meter is calibrated in terms of resistance.

Note that the resistance scale on a VTVM reads increasing resistance from left to right, which is opposite to the way resistance scales read on nonelectronic multimeters. This is to be expected, since the VTVM reads a higher resistance as a higher voltage, whereas a normal multimeter indicates a higher resistance as a smaller current. The successive positions on the range selector of the VTVM provide resistance ranges increasing in decade steps.

The complete circuit diagram of a commercial VTVM is given in Fig. 10-9. Although the actual circuitry may vary from one instrument to the next, it is easy to recognize the various sections discussed so far,

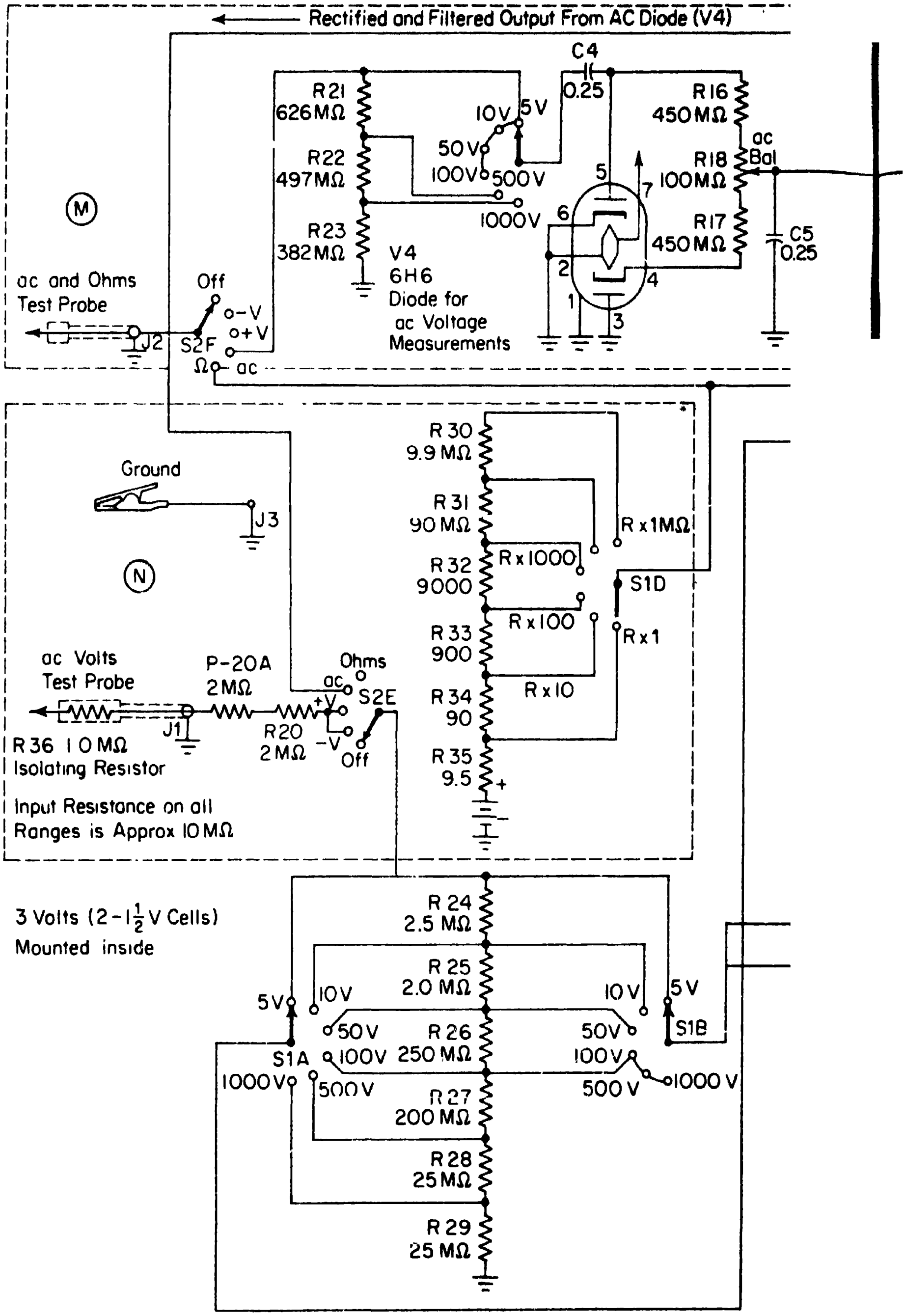
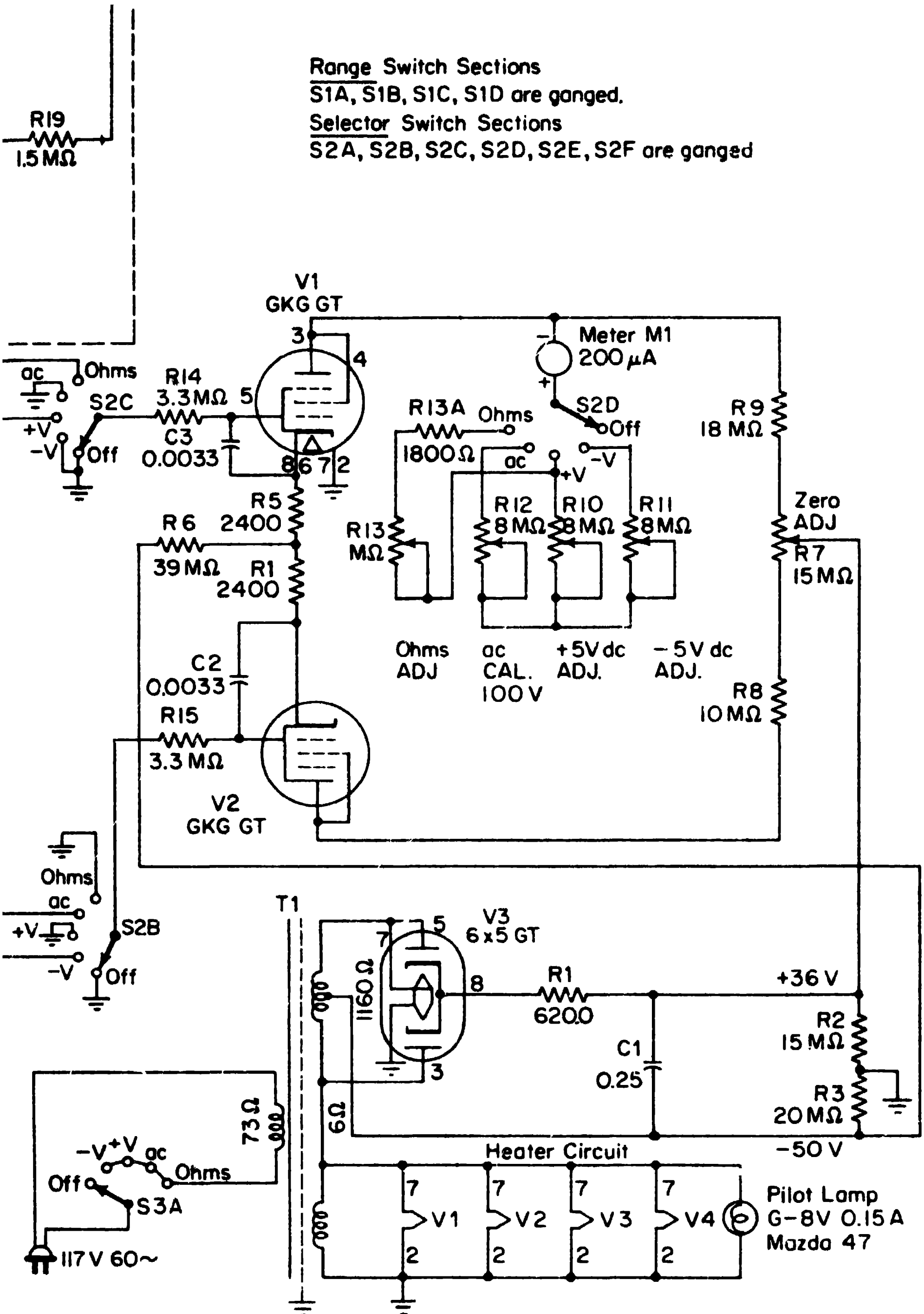


Figure 10-9
Circuit diagram of a commercial VTVM. (*Courtesy of Hewlett-Packard Co.*)


Range Switch Sections
S1A, S1B, S1C, S1D are ganged.
Selector Switch Sections
S2A, S2B, S2C, S2D, S2E, S2F are ganged
R19
1.5 MΩ
V1
6K6 GT
Meter M1
200 μA
ac
Ohms
S2C
+V
-V
Off
R14
3.3 MΩ
C3
0.0033
R13A
1800 Ω
Ohms
S2D
Off
ac
+V
-V
R9
18 MΩ
R5
2400
R6
39 MΩ
R1
2400
R13
MΩ
R12
8 MΩ
R10
8 MΩ
R11
8 MΩ
Zero
ADJ
R7
15 MΩ
Ohms
ADJ
ac
CAL.
100 V
+5V dc
ADJ.
-5V dc
ADJ.
C2
0.0033
R15
3.3 MΩ
R8
10 MΩ
V2
6K6 GT
Ohms
ac
+V
S2B
-V
Off
T1
V3
6x5 GT
1160 Ω
R1
620 Ω
+36 V
R2
15 MΩ
C1
0.25
R3
20 MΩ
-50 V
73 Ω
6 Ω
-V +V ac
Off
Ohms
S3A
117 V 60~
Heater Circuit
V1
V2
V3
V4
Pilot Lamp
G-8V 0.15 A
Mazda 47

10-5 Considerations in Choosing an Analog Voltmeter

The most appropriate instrument for a particular voltage measurement application depends on the performance required in a given situation. Some of the more important considerations in choosing a voltmeter are

(a) *Circuit Loading.* The voltmeter should have a considerably higher impedance than the source or the circuit being measured. For example, measuring the low-frequency voltage across a 100-kΩ resistor with a voltmeter having a 10-MΩ input resistance results in an error of 1 per cent because the voltmeter in effect changes the test circuit.

Input impedance is usually specified by the resistive and the reactive parts. The input reactance due to the input shunt capacitance C_{in} is a reciprocal function ($X_c = 1/\omega C$); therefore, the smaller the input shunt capacitance of the meter, the smaller the loading effect at higher frequencies.

In some applications, a passive voltage-divider probe can be used to reduce the input capacitance at the point of measurement at the sacrifice of perhaps 20 dB of sensitivity. With such a probe, measurements can be made quite easily at random points without upsetting the circuit under test.

(*b*) *Ranges.* The ranges on the meter scale may be in the 1-3-10 sequence with 10 dB of separation, in the 1.5-5-15 sequence or in a single scale calibrated in decibels. In any case, the scale divisions should be compatible with the instrument accuracy. A linear meter with a 1 per cent full-scale accuracy should have one hundred divisions on the 1.0-V scale so that 1 per cent can be easily resolved. An instrument with an accuracy of 1 per cent or less should have a mirror-backed scale to reduce the parallax problem.

(*c*) *Decibels.* The decibel unit is very effective in measurements covering a wide range of voltages. The response of amplifiers and filters, for example, is usually expressed as a graph of voltage in decibels versus frequency in hertz. Most voltmeters with dB scales are calibrated in dBm, referenced to some particular impedance. The 0-dBm reference for a 600-Ω system is 0.7746 V; for a 50-Ω system it is 0.2236 V. In many applications, only a 0-dB reference is needed. In this case 0 dBv (relative to 1 V) can be used for any impedance system.

(*d*) *Sensitivity Versus Bandwidth.* Noise is a function of bandwidth. A voltmeter with a broad bandwidth will pick up and generate more noise than one operating over a narrow range of frequencies. In general, an instrument with a bandwidth of 10 Hz to 10 MHz has a sensitivity of 1 mV. On the other hand, a voltmeter whose bandwidth extends only to 5 MHz could have a sensitivity of 100 μV.

(*e*) *Battery Operation.* For field work, an instrument powered by an internal battery is necessary. A battery-operated TVM is a very popular choice. If an area contains troublesome groundloops, a battery-powered instrument is to be preferred over a mains-powered voltmeter to remove the groundpaths.

(*f*) *AC Current Measurements.* Current measurements can be made by using a sensitive ac voltmeter and a series resistance. The usual practice, however, employs an ac current probe which enables an ac current to be measured without disturbing the circuit under test. The current probe clips around the wire carrying the current and in effect makes the wire the one-turn primary of a transformer formed by a ferrite core and a many-turn secondary within the current-probe body. The signal induced in the secondary is amplified and the amplifier's output voltage can be applied to any suitable ac voltmeter for measurement. Normally, the amplifier is designed so that 1 mA in the wire being measured produces 1 mV at the amplifier output. The current is then read directly on the voltmeter, using the same scale as for voltage measurements.

Summarizing the preceding considerations, the following guidelines are stated

(a) For measurements involving dc applications, select the meter with the broadest capability meeting the circuit's requirements.

(b) For ac measurements involving sine waves with only modest amounts of distortion ($<$ 10 per cent), the average-responding voltmeter provides the best accuracy and most sensitivity per dollar investment.

(c) For high-frequency measurements ($>$ 10 MHz), the peak-responding voltmeter with a diode-probe input is the most economical choice. Peak-responding circuits are acceptable if the inaccuracies caused by distortion in the input waveform can be tolerated.

(d) For measurements where it is important to determine the effective power of waveforms which depart from the true sinusoidal form, the rms-responding voltmeter is the appropriate choice.

10-6 Differential Voltmeters

One of the most accurate means of measuring an unknown voltage is found in the *differential voltmeter technique,* where the unknown voltage is compared to a known voltage. The principle of operation of the differential voltmeter is similar to that of the potentiometer discussed in Chapter 6, and for this reason, the instrument is sometimes called a *potentiometric voltmeter.*

The classic differential volt^ge measurement is shown in its basic form in the circuit of Fig. 10-10. A precision resistive divider network is used to divide

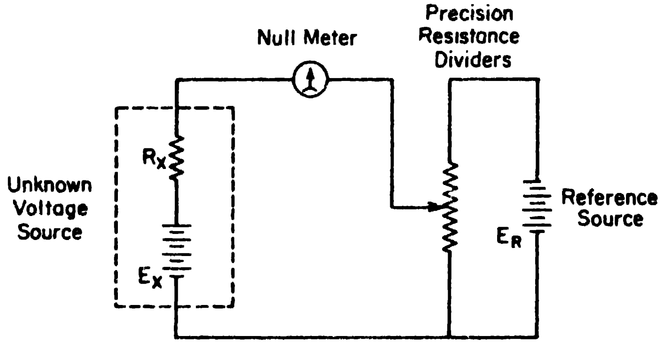


Figure 10-10
Classic differential voltage measurement.

down an accurately known *reference voltage.* The divider is adjusted until its output voltage equals the unknown voltage. The null meter, which is connected between the unknown source and the divider-output terminals, indicates zero volts when the two voltages are equal. In this *null condition* neither the source nor the reference supply delivers current to the meter, and the differential voltmeter presents an *infinite impedance* to the source under test. Note that the null meter serves to indicate only the residual differential between the known voltage and the unknown voltage. To detect small differences in unbalance potentials, a sensitive meter movement is required; accuracy of the meter is of secondary importance since the meter is not used to indicate the absolute value of the unknown voltage.

The reference source usually consists of a low-voltage dc standard, such as a 1-V dc laboratory reference standard or a low-voltage Zener-controlled precision supply.

In order to measure high voltages, a high-voltage reference supply may be used. The usual practice, however, employs a divider across the unknown source to reduce the voltage to a sufficiently low value for direct comparison against the usual low-voltage dc standard. The main drawback of this system is that a differential voltmeter with input divider has a relatively low input resistance, especially for unknown voltages much higher than the reference standard. This low input resistance is undesirable because of its loading effect. A differential voltmeter offers an input resistance approaching infinity only at the null condition and then *only* if an input divider is *not* used.

The *ac differential voltmeter* is a modification of the dc instrument and involves use of a precision *rectifier* circuit. The unknown ac voltage is applied to the rectifier for conversion to a dc voltage equivalent to the average value of the ac. The resulting dc is then applied to the potentiometric voltmeter in the usual manner. The simplified block diagram of an ac differential voltmeter shown in Fig. 10-11 is self-explanatory.

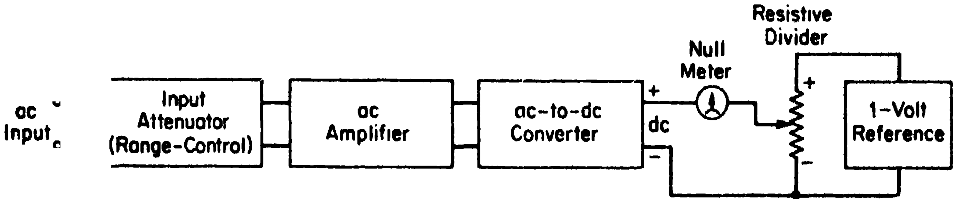


Figure 10-11
Simplified block diagram of an ac differential voltmeter.

Since the differential voltmeter requires a reference source in order to make a valid measurement and also a meter circuit to detect an unbalance between the unknown and the known voltage, some manufacturers combine the various elements into a multifunction laboratory instrument. An example of this is found in the *dc standard/differential voltmeter,* which has three modes of operation: as a dc voltage standard, as a dc differential voltmeter, and as a dc voltmeter. Function switching allows many of the same basic circuits to be used in each mode of operation.

The block diagram of Fig. 10-12 illustrates the "standard" mode of operation, where the instrument generates precision output voltages from 0 V to 1,000 V as a reference source for many laboratory applications. An oven-controlled reference supply generates a very stable +1 V dc, which is applied to a decimal divider network. The divider ratios are controlled by a set of front-panel switches allowing the reference supply to be adjusted in 1-μV steps from 0 V to 1 V. This reference output voltage is applied to a high-gain dc amplifier with degenerative voltage feedback to obtain precisely controlled gain characteristics.

The dc amplifier, consisting of several stages in cascade, provides an open-loop gain of 10^8 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 can be expressed by the relationship:

$$G = \frac{A}{1 + \beta A} \tag{10-3}$$

where G = closed-loop gain (voltage gain with feedback)

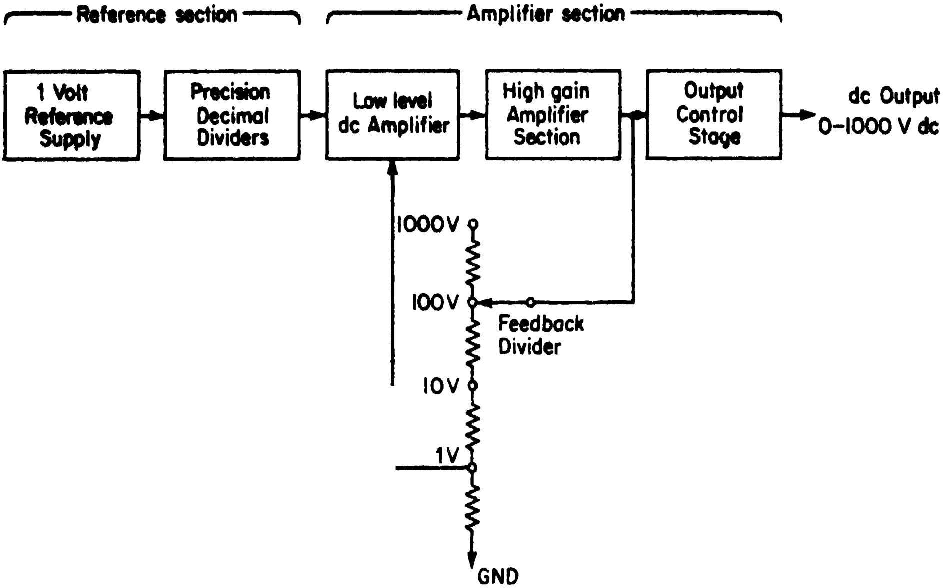


Figure 10-12
Simplified block diagram of the differential dc voltmeter in the *standard* mode of operation. The reference section in conjunction with the dc amplifier section provides precision dc output voltages from 0 V to 1,000 V.

A = open-loop gain (voltage gain without feedback)
β = fraction of the output voltage used as degenerative feedback.

If the open-loop gain is very high (ideally infinite) expression (10-3) reduces to

$$G = \frac{1}{\beta} \tag{10-4}$$

showing that the gain of the amplifier depends only on the amount of degenerative feedback. The accuracy of the closed-loop gain therefore depends only upon the accuracy of the voltage divider which determines β. The feedback divider, shown in the block diagram of Fig. 10-12, is made from stable, wire-wound precision resistors, enabling the amplifier to have precisely controlled closed-loop gain characteristics.

For the circuit illustrated in Fig. 10-12, the output terminals of the instrument in the "standard" mode of operation provide the following range of precision voltages:

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)
0–1,000 V in 1-mV steps (1,000-V range)

In the "differential" mode of operation, the same building blocks are used, together with a metering circuit. This is illustrated in the block diagram of Fig. 10-13.

The unknown voltage is applied to the input terminals of the dc amplifier. Feedback voltage from the output stage of the dc amplifier to its input stage again controls the closed-loop gain. A fraction of the feedback voltage, proportional to the unknown input voltage, is applied in series opposition to the reference-divider output voltage. The meter circuit monitors the difference between the feedback voltage and the reference voltage and indicates a null when the two voltages are equal. The range selector on the front panel of the instrument controls both the feedback voltage and the voltage which is applied in opposition to the reference divider output in such a way that the 1-V capability of the reference supply is never exceeded.

In the "voltmeter" mode of operation, the dc amplifier acts as a buffer stage to provide high input impedance to the unknown voltage source. The input voltage is amplified and feedback voltage from the dc amplifier output stage is applied directly to the meter circuit. The meter circuit, which incorporates a feedback-controlled amplifier, allows selection of its sensitivity by adjustment of the feedback loop through a front-panel control, marked *sensitivity*. This feature provides for extreme sensitivity of the meter circuit, often in the order of 1-μV full-scale deflection. Meaningful measurements at the very high sensitivities, however, are difficult to make because of the problems of noise generation and pickup.

An ac-to-dc converter may be incorporated in the instrument to provide the capability of ac voltage measurement by potentiometric methods.

10-7 Digital Voltmeters

The digital voltmeter (DVM) displays measurements of dc or ac voltages as discrete numerals in the decimal number system, rather than as the pointer deflection on a continuous scale commonly used in analog devices. Numerical readout is advantageous in many applications since it reduces human reading and interpolation errors, eliminates parallax error, increases reading speed, and often provides outputs in digital form suitable for further processing or recording.

The DVM is a versatile and accurate instrument used in many laboratory measurement applications. Since the development and perfection of integrated circuit (IC) modules, the size, power requirements, and cost of the DVM

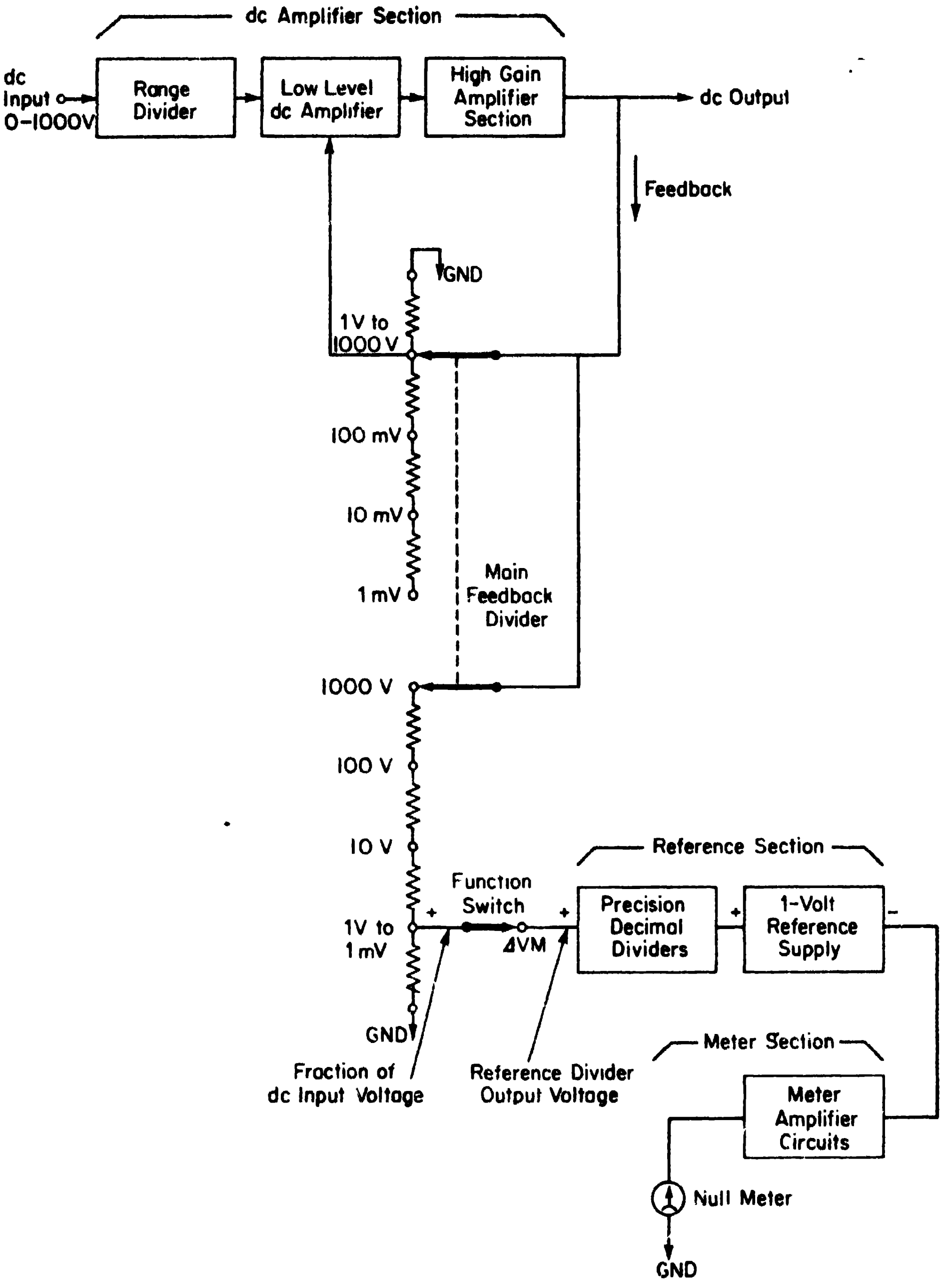


Figure 10-13
Simplified block diagram of the dc differential voltmeter in the *differential* mode of operation. The meter section indicates the voltage balance between the reference section and the dc amplifier section.

have been drastically reduced so that some simple DVMs now actively compete with conventional analog instruments, both in portability and price.

The DVM's outstanding qualities can best be illustrated by quoting some typical operating and performance characteristics. The following specifications do not all apply to one particular instrument but they do represent valid information on the present state of the art:

(1) Input range: from ± 1.000000 V to $\pm 1{,}000.000$ V, with automatic range selection and overload indication.

(2) Absolute accuracy: as high as 0.005 per cent ($\pm$) of the reading.

(3) Stability: short-term: 0.002 per cent of the reading for a 24-hr period. long-term: 0.008 per cent of the reading for a 6-month period.

(4) Resolution: 1 part in 10^6 (1 μV can be read on the 1-V input range).

(5) Input characteristics: input resistance typically 10 MΩ; input capacitance typically 40 pF.

(6) Calibration: internal calibration standard allows calibration independent of the measuring circuit; derived from stabilized reference source.

(7) Output signals: print command allows output to printer; BCD (binary-coded-decimal) output for digital processing or recording.

Optional features may include additional circuitry to measure current, resistance, and voltage ratios. Other physical variables may be measured by using suitable transducers.

Most digital voltmeters on the market can be classified according to one of the following categories:

(1) Ramp-type DVM.
(2) Integrating DVM.
(3) Potentiometric DVM.
(4) Successive-approximation type DVM.
(5) Continuous-balance DVM.

Ramp-Type DVM. The operating principle of the ramp-type DVM is measurement of the time it takes for a linear ramp voltage to change from 0 V to the level of the input voltage (or vice versa). This time interval is then measured with an electronic time-interval counter and the count is displayed as a number of digits on electronic indicating tubes.

Conversion from a voltage to a time interval is illustrated by the waveform diagram of Fig. 10-14. At the start of a measurement cycle, a ramp voltage is initiated; this voltage can be positive- or negative-going. The negative-going ramp, shown in the example of Fig. 10-14, is compared continuously with the

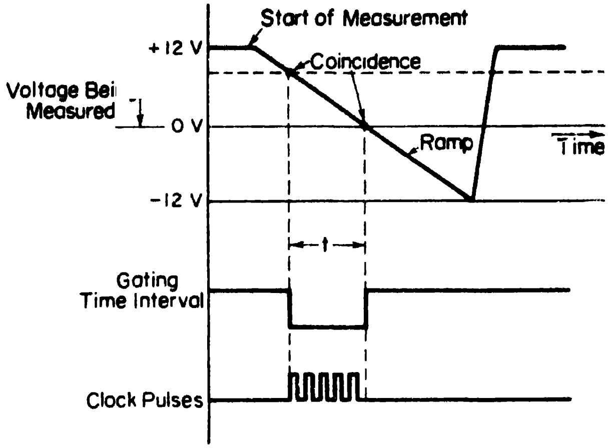


Figure 10-14
Illustrating a voltage-to-time conversion by using gated clock pulses.

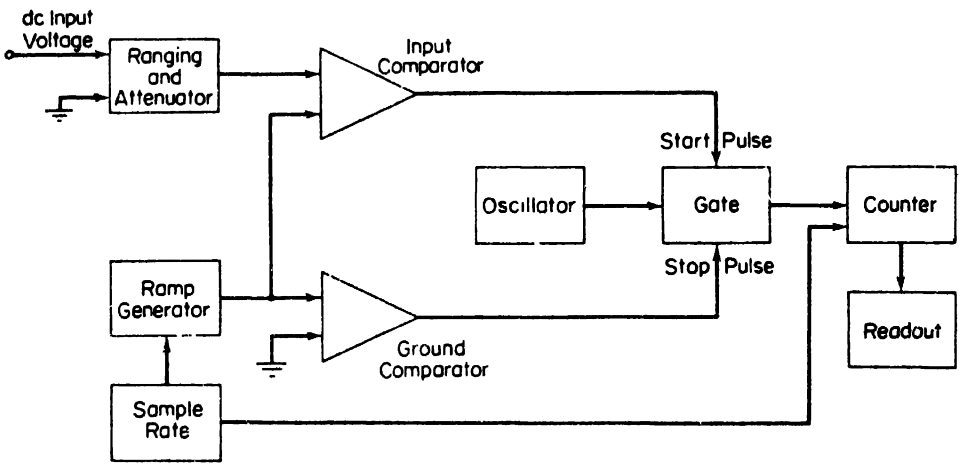


Figure 10-15
Block diagram of a ramp-type digital voltmeter.

unknown input voltage. At the instant that the ramp voltage equals the unknown voltage, a coincidence circuit or *comparator* generates a pulse which opens a gate (see Fig. 10-15). The ramp voltage continues to decrease with time. When it finally reaches 0 V (or ground potential), a second comparator generates an output pulse which closes the gate.

An oscillator generates *clock pulses* which are allowed to pass through the gate to a number of decade counting units (DCUs) which totalize the number of pulses passed through the gate. The decimal number, displayed by the indicator tubes associated with the DCUs, is a measure of the magnitude of the input voltage.

The *sample-rate* multivibrator determines the rate at which the measurement cycles are initiated. The oscillation of this multivibrator can be adjusted by a front panel control, marked *rate*. In some instruments the rate is fixed, a typical value being five measuring cycles per second. 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 returns all the DCUs to their 0 state, removing the display momentarily from the indicator tubes.

Staircase-Ramp DVM. This instrument is given in block diagram form in Fig. 10-16, and is a variation of the ramp-type DVM. It is somewhat simpler in over-all design, resulting in a moderately priced general-purpose instrument, which can be used in the laboratory, on production test-stands, in repair shops, and at inspection stations.

This DVM again makes voltage measurements by comparing the input voltage to an internally generated "staircase ramp" voltage. The instrument shown in Fig. 10-16 contains a 10-MΩ input attenuator, providing five input ranges from 100 mV to 1,000 V full scale. The dc amplifier, with a fixed gain of 100, delivers 10 V to the comparator at any of the full-scale voltage settings of the input divider. The comparator senses coincidence between the amplified input voltage and a staircase ramp voltage. This ramp voltage is generated as the measurement proceeds through its cycle.

When the measurement cycle is first initiated, the clock (a 4.5-kHz relaxation oscillator) provides pulses to three DCUs in cascade. The *units* counter provides a *carry* pulse to the *tens* decade at every tenth input pulse. The *tens* decade counts the carry pulses from the *units* decade and provides its own carry pulse after it has counted ten carry pulses. This carry pulse is fed to the *hundreds* decade which provides a carry pulse to an *overrange* circuit. The overrange circuit causes a front panel indicator to light up, warning the operator that the input capacity of the instrument has been exceeded. The operator should then switch to the next higher setting on the input attenuator.

Each decade counter unit is connected to a *digital-to-analog converter* (D/A converter). The outputs of the D/A converters are connected in parallel and provide an output current proportional to the current count of the DCUs. The *staircase amplifier* converts the D/A current into a staircase voltage which is applied to the comparator. When the comparator senses coincidence of the input voltage and the staircase voltage, it provides a trigger pulse to stop the oscillator. The current content of the counter is then proportional to the magnitude of the input voltage.

The *sample rate* is controlled by a simple relaxation oscillator. This oscil-

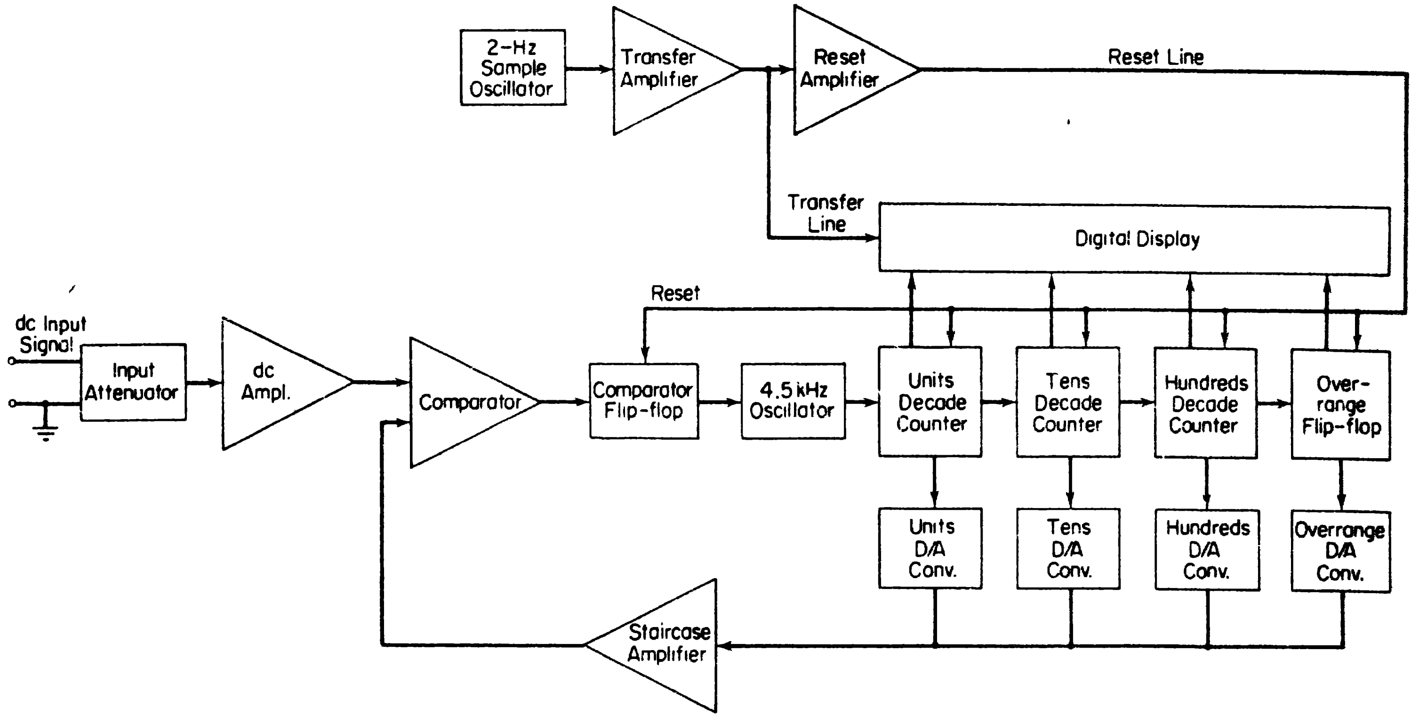


Figure 10-16. Block diagram of a staircase ramp digital voltmeter

lator triggers and resets the *transfer amplifier* at a rate of two samples per second. The transfer amplifier provides a pulse which transfers the information stored in the decade counters to the front panel display unit. The trailing edge of this pulse triggers the *reset amplifier* which sets the three decade counters to zero and initiates a new measurement cycle by starting the master oscillator or clock.

The display circuits store each reading until a new reading is completed, eliminating any blinking or counting during the computation.

The Integrating-Type Digital Voltmeters. This voltmeter measures the *true average* of the input voltage over a fixed measuring period, in contrast to the ramp-type DVM which *samples* the voltage at the end of a measuring cycle. A widely used technique to accomplish integration employs a voltage-to-frequency (V/F) converter. The V/F converter functions as a feedback control system which governs the rate of pulse generation in proportion to the magnitude of the input voltage.

The simplified block diagram of an integrating DVM is given in Fig. 10-17. The dc voltage under test is applied to the input stage which isolates the meter circuitry from the test circuit and provides the necessary input attenuation. The attenuated input signal is applied to the V/F converter. This circuit consists of an integrating amplifier, a level detector (*comparator circuit*), and a pulse generator. The integrating amplifier produces an output voltage proportional to the input voltage and related to the input and feedback elements by the equation

$$\begin{aligned} V_{\text{out}} &= -\frac{1}{C}\int i\,dt \\ &= -\frac{1}{RC}\int V_{\text{in}}\,dt \end{aligned} \tag{10-5}$$

If the input voltage is constant, the output is a linear ramp following the equation

$$V_{\text{out}} = -V_{\text{in}}\frac{t}{RC} \tag{10-6}$$

When the ramp reaches a certain negative voltage level, the level detector triggers the pulse generator, which applies a negative voltage step to the summing junction of the integrating amplifier. The sum of the input voltage and the pulse voltage is negative, causing the ramp to reverse its direction. This "retrace" is very rapid since the pulse is large in amplitude compared to the input voltage. When the now positive-going ramp reaches 0 V, the level detector generates a reset trigger to the pulse generator. The negative pulse is removed from the summing junction of the integrating amplifier and only the original input voltage is left. The amplifier then produces a negative-going ramp again and the procedure repeats itself.

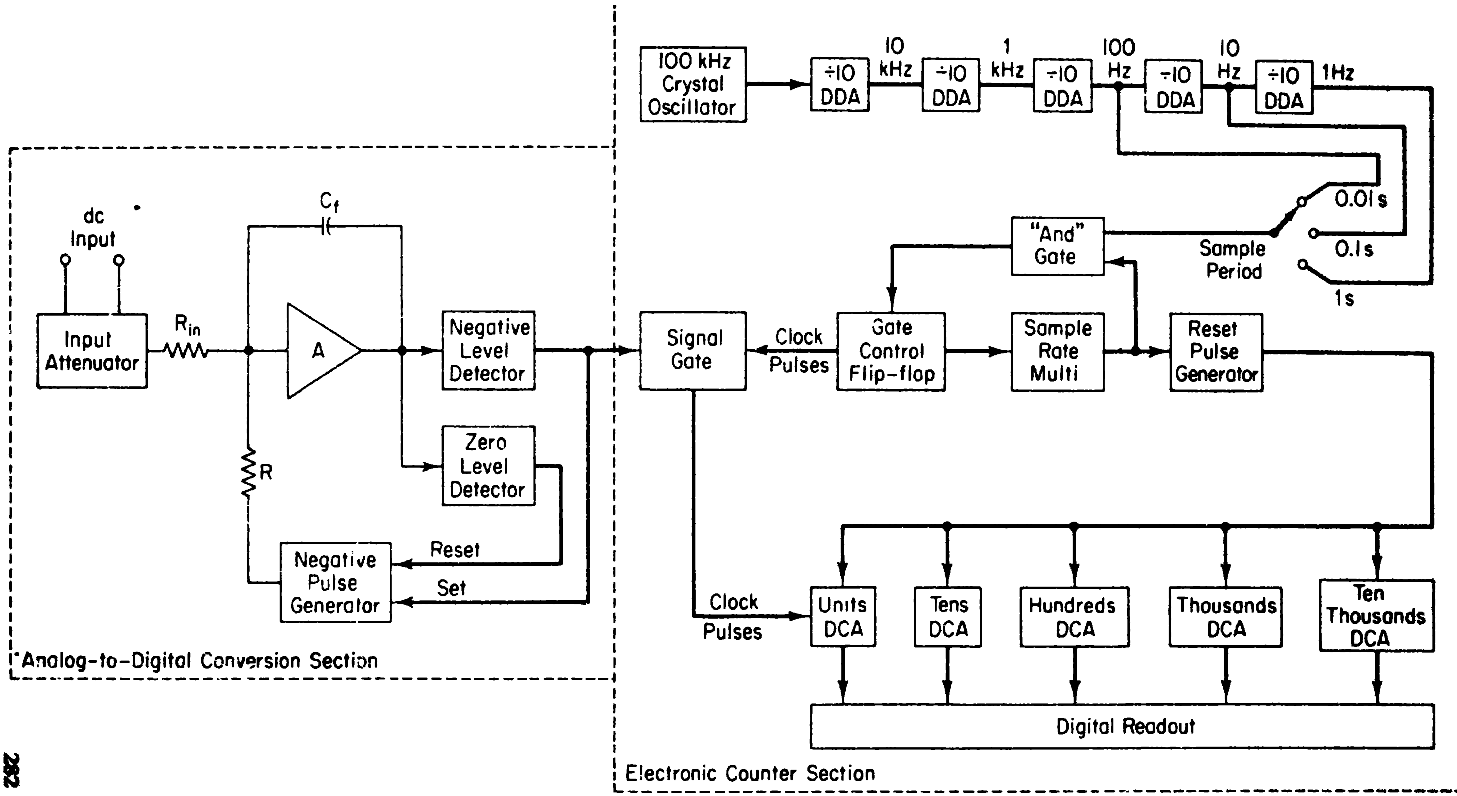


Figure 10-17. Block diagram of the intergrating digital voltmeter.

The rate of pulse generation is governed by the magnitude of the dc input voltage. A larger input voltage causes a steeper ramp and therefore a higher pulse repetition rate (PRR).

The major advantage of this system of A/D conversion is its ability to measure accurately in the presence of large amounts of superimposed noise, since the input is integrated.

The level-detector output pulse controls the signal gate allowing the decimal counters to accumulate a count provided by the crystal oscillator circuitry. The remainder of the circuit is essentially identical to any conventional counter and needs no further elaboration.

The Servo-Balancing Potentiometer-Type DVM. This is a low-cost instrument providing excellent performance. The accuracy of this meter is usually in the order of 0.1 per cent of its input range. It has an input impedance of about 10 MΩ, and acceptable resolution.

The block diagram of this DVM is given in Fig. 10-18. The dc input volt-

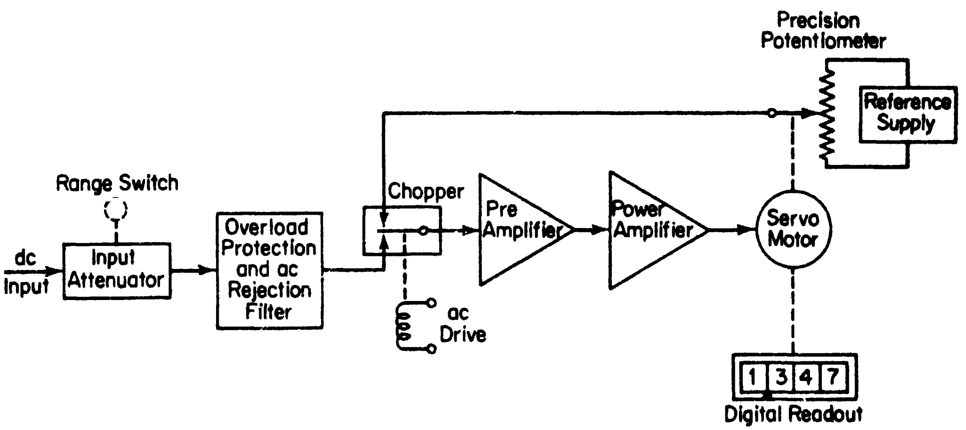


Figure 10-18
Functional block diagram of a servo-balancing potentiometer-type digital voltmeter.

age is applied to an *input attenuator* providing suitable range switching. This is a front panel control, which also causes a decimal point indicator to move on the display area in accordance with the input range selected. After passing through an overvoltage protection circuit and ac rejection filter, the input voltage is applied to one side of a mechanical *chopper comparator*. The opposite side of the comparator is connected to the wiper arm of the motor-driven precision potentiometer. The potentiometer is connected across a precision reference supply. The output of the chopper comparator, which is driven by the line voltage and vibrates at the line frequency rate, is a square-wave signal. The amplitude of the square wave is a function of the difference in magnitude and polarity of the dc voltages connected to the opposite sides of the chopper. The square-wave signal is amplified by a high-impedance, low-noise preamplifier and fed to a power amplifier. This amplifier has special damping to minimize overshoot and hunting at the null position. The servo motor, upon receiving the amplified square-wave difference signal, drives the arm of the precision potentiometer in the direction required to *cancel* the difference voltage across the chopper comparator. The servo motor also drives a drum-type mechanical indicator which has the digits from 0 to 9 imprinted about the periphery of its drum segments. The position of the servo motor shaft corresponds to the amount of feedback voltage required to null the chopper input, and this position is indicated by the drum-type indicator. The position of the shaft therefore is an indication of the magnitude of the input voltage.

It is clear that this instrument does not "sample" the unknown dc voltage at regular intervals, as is the case with more sophisticated instruments, but it continuously seeks to *balance* the input voltage against the internally generated reference. Due to the different mechanical movements involved in the mechanism, such as the positioning of the potentiometer arm and the rotation of the indicator mechanism, the average reading time is approximately 2 s. Simplicity of design and low cost, however, make this instrument a very attractive choice when extreme accuracy is not required.

10-8 The Vector-Impedance Meter

Impedance measurements are concerned with both the magnitude (Z) and the phase angle (θ) of a component. In many applications it is not only necessary to determine the amount of opposition to current, but it may also be important to find the ratio of the reactance to the resistance and whether the reactance is inductive or capacitive.

At frequencies below 100 MHz, measurement of voltage and current is usually sufficient to determine the magnitude of the impedance. The phase difference between the voltage waveform and the current waveform indicates whether the component is inductive or capacitive. If the phase angle can be determined, for example, by using a CRO displaying a Lissajous pattern, the reactance can be computed. If a component must be fully specified, its properties should be determined at several different frequencies, and many measurements may be required. Especially at the higher frequencies, these measurements become rather elaborate and time consuming and many steps may be required to obtain the desired information.

The development of instruments, such as the *vector impedance meter*, greatly simplifies impedance measurements over a wide frequency range. This instrument makes it possible to obtain *sweep-frequency plots* of impedance and phase angle versus frequency, providing complete coverage within the frequency band of interest.

The vector-impedance meter, shown in the photograph of Fig. 10-19,

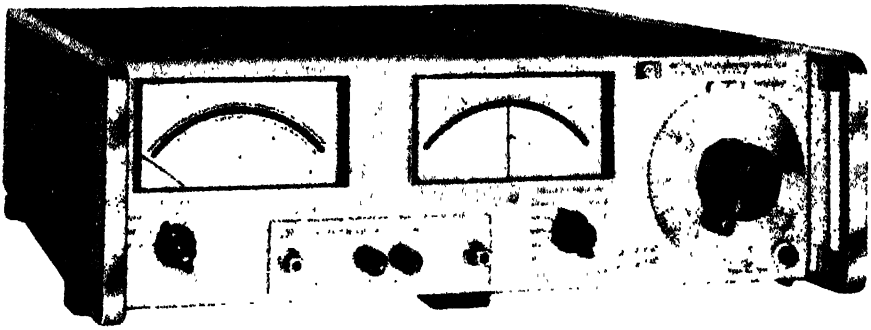


Figure 10-19
A vector impedance meter. (*Courtesy Hewlett-Packard Co.*)

makes simultaneous measurements of impedance and phase angle over a frequency range of from 5 Hz to 500 kHz. The unknown component is simply connected across the input terminals of the instrument, the desired frequency is selected by turning the front panel controls, and the two front panel meters indicate the magnitude of the impedance and the phase angle.

The operation of the vector-impedance meter is best understood by referring to the block diagram of Fig. 10-20, which relates to the instrument shown in the photograph of Fig. 10-19. Two measurements take place: (1) the magnitude of the impedance is determined by measuring the current through the unknown component when a known voltage is applied across it, or, by measuring the voltage across the component when a known current is passed through it; (2) the phase angle is found by determining the phase difference between the voltage across the component and the current through the component.

Inspection of the block diagram of Fig. 10-20 shows that the meter contains a *signal source* (a Wien bridge oscillator) with front panel controls for selecting the frequency range and continuous adjustment of the frequency selected. The output of the Wien bridge oscillator is fed to an *AGC amplifier* which allows accurate gain adjustment of the oscillator output by means of its feedback voltage. This gain adjustment is an internal control actuated by the setting of the *impedance range switch,* to which the AGC amplifier output is connected. The impedance range switch is a precision attenuator network controlling the

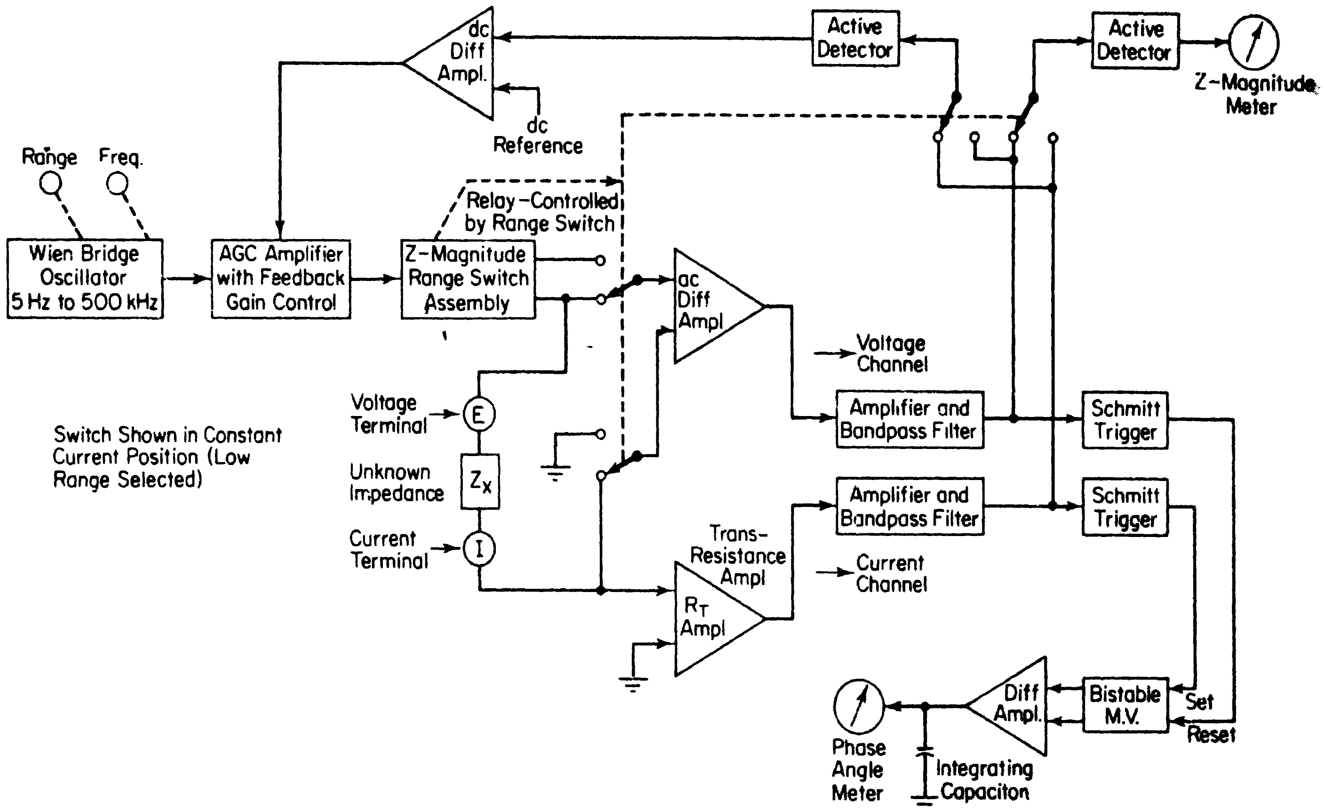


Figure 10-20. Biock diagram of the vector impedance meter. (*Courtesy Hewlett-Packard Co.*)

oscillator output voltage and at the same time determining the manner in which the unknown component will be connected into the circuitry which follows the range switch.

The impedance range switch permits operation of the instrument in two modes: the *constant-current mode* and the *constant-voltage mode*. The three lower ranges ($\times 1$, $\times 10$, and $\times 100$) operate in the constant-current mode and the four higher ranges ($\times 1$k, $\times 10$k, $\times 100$k, and $\times 1$M) operate in the constant-voltage mode.

In the constant-current mode of operation, the unknown component is connected across the input of the ac differential amplifier. The current supplied to the unknown depends on the setting of the impedance range switch. This current is held constant by the action of the transresistance amplifier (R_T) which converts the current through the unknown to a voltage output equal to the current times its feedback resistance. The transresistance amplifier is an operational amplifier where the output voltage is proportional to the input current. The output of the R_T amplifier is fed to a detector circuit, compared to a dc reference voltage and the resulting control voltage regulates the gain of the AGC amplifier and hence the voltage applied to the impedance range switch. The output of the ac differential amplifier is applied to an amplifier and filter section consisting of high- and low-band filters which are changed with the frequency range to restrict the amplifier bandwidth. The output of the bandpass filter is connected, when selected, to a detector which drives the *Z-magnitude meter*. Since the current through the unknown is held constant by the R_T amplifier, the Z-magnitude meter, which measures the voltage across the unknown, deflects in proportion to the magnitude of the unknown impedance and is calibrated accordingly.

In the constant-voltage mode of operation, the two inputs to the differential amplifier are switched. The terminal which was connected to the input of the transresistance amplifier in the constant-current mode is now grounded. The other input of the differential amplifier which was connected to the voltage terminal of the unknown component, is now connected to a point on the Z-magnitude range switch which is held at a constant potential. The voltage terminal of the unknown is connected to this same point of constant potential, or, depending on the setting of the Z-magnitude range switch, to a decimal fraction of this voltage. In any case, the voltage across the unknown is held at a constant level. The current through the unknown is applied to the transresistance amplifier which again produces an output voltage proportional to its input current.

The roles of the ac differential amplifier and the transresistance amplifier are now reversed. The voltage output of the R_T amplifier is applied to the detector and then to the Z-magnitude meter. The output voltage of the differential amplifier controls the gain of the AGC amplifier in the same manner that the R_T amplifier did in the constant-current mode.

Phase-angle measurements are carried out simultaneously. The outputs of both the voltage channel and the current channel are amplified and each output is connected to a Schmitt trigger circuit. The Schmitt trigger circuits produce a positive-going spike every time that the input sine wave goes through a zero crossing. These positive spikes are applied to a *binary phase detector circuit.* The phase detector consists of a bistable multivibrator, a differential amplifier, and an integrating capacitor. The positive-going pulse from the constant current channel sets the multivibrator, and the pulse from the constant voltage channel resets the multivibrator. The "set" time of the MV is therefore determined by the zero crossings of the voltage waveform and the current waveform. The "set" and "reset" outputs of the MV are applied to the differential amplifier, which applies the difference voltage to the integrating capacitor. The capacitor voltage is directly proportional to the zero-crossing time interval and is applied to the *phase-angle meter* which then indicates the phase difference, in degrees, between the voltage and current waveforms.

Calibration of the vector-impedance meter is usually performed by connecting standard components to the input terminals. There components may be standard resistors or capacitors. An electronic counter is needed to accurately determine the period of the applied test frequency. When the value of the component under test and the frequency of the test signal are both known accurately, the impedance or reactance can be calculated and compared to the indication on the Z-magnitude meter. With the standard resistor connected to the input terminals, the phase-angle meter should read 0°.

10-9 The Vector Voltmeter

A *vector voltmeter* measures the amplitude of a signal at two points in a circuit and simultaneously measures the phase difference between the voltage waveforms at these two points. This instrument can be used in a wide variety of applications, especially in situations where other methods are very difficult or time consuming. The vector voltmeter is extremely useful in very *high-frequency measurement situations* and is capable of accurate phase determinations at frequencies up to several GHz. The vector voltmeter may be used successfully in the following measurements:

amplifier gain and phase shift
complex insertion loss
filter transfer functions
two-port network parameters

The vector voltmeter converts two RF signals of the same fundamental frequency in the range from 1 MHz to 1 GHz to two IF signals with 20-kHz fundamental frequencies. The IF signals have the same amplitudes, waveforms, and phase relationships as the RF signals. Consequently, the fundamental components of the IF signals have the same amplitude and phase relationships as the fundamental components of the RF signals. These fundamental components are filtered from the IF signals and measured by a voltmeter and a phase meter.

The instrument consists of five major sections, incicated in the block diagram of Fig. 10-21 by the dashed outlines. They are identical channel A and channel B RF-to-IF converters, an automatic phase control section, a phase meter, and a voltmetor. The RF-to-IF converters and the phase control section produce two 20-kHz sine waves which have the same amplitudes and phase relationship as the fundamental components of the RF signals applied to channels A and B. The phase meter section continuously monitors these two 20-kHz sine waves and provides a meter display of the phase angle between them. The voltmeter section is manually switched to channel A or channel B (20-kHz sine wave) and provides a meter display of the amplitude.

Each RF-to-IF converter consists of a *sampler* and a *tuned amplifier*. The sampler produces a 20-kHz replica of the RF input waveform and the tuned amplifier extracts the 20-kHz fundamental component from this waveform replica.

Sampling is a time-stretching process with which a high-frequency repetitive signal is duplicated at a much lower frequency. The process is illustrated by the diagram of Fig. 10-22. An electronic switch is connected between the RF input waveform and a storage capacitor. Each time the switch is momentarily closed, the capacitor is charged to the instantaneous value of the input voltage and holds this voltage until the switch closes again. With appropriate timing, samples are taken at progressively later points on the RF waveform. Provided that the RF waveform is repetitive, the samples make up a reconstructed waveform which is a reproduction of the RF signal at a much slower rate.

Each input channel has a sampler consisting of a *sampling gate* and a storage capacitor. The sampling gates are controlled by pulses from the same pulse generator. Samples are taken in each channel at exactly the same instant, and the phase relationship of the input signals is therefore preserved in the IF signals.

The *phase control unit* is a rather sophisticated circuit which generates the sampling pulses for both RF-to-IF converters and automatically controls the pulse rate to produce 20-kHz IF signals which have the same phase relationship as the RF input signals. The sampling pulse rate is controlled by a voltage tuned oscillator (VTO) for which the tuning voltage is supplied by the auto-

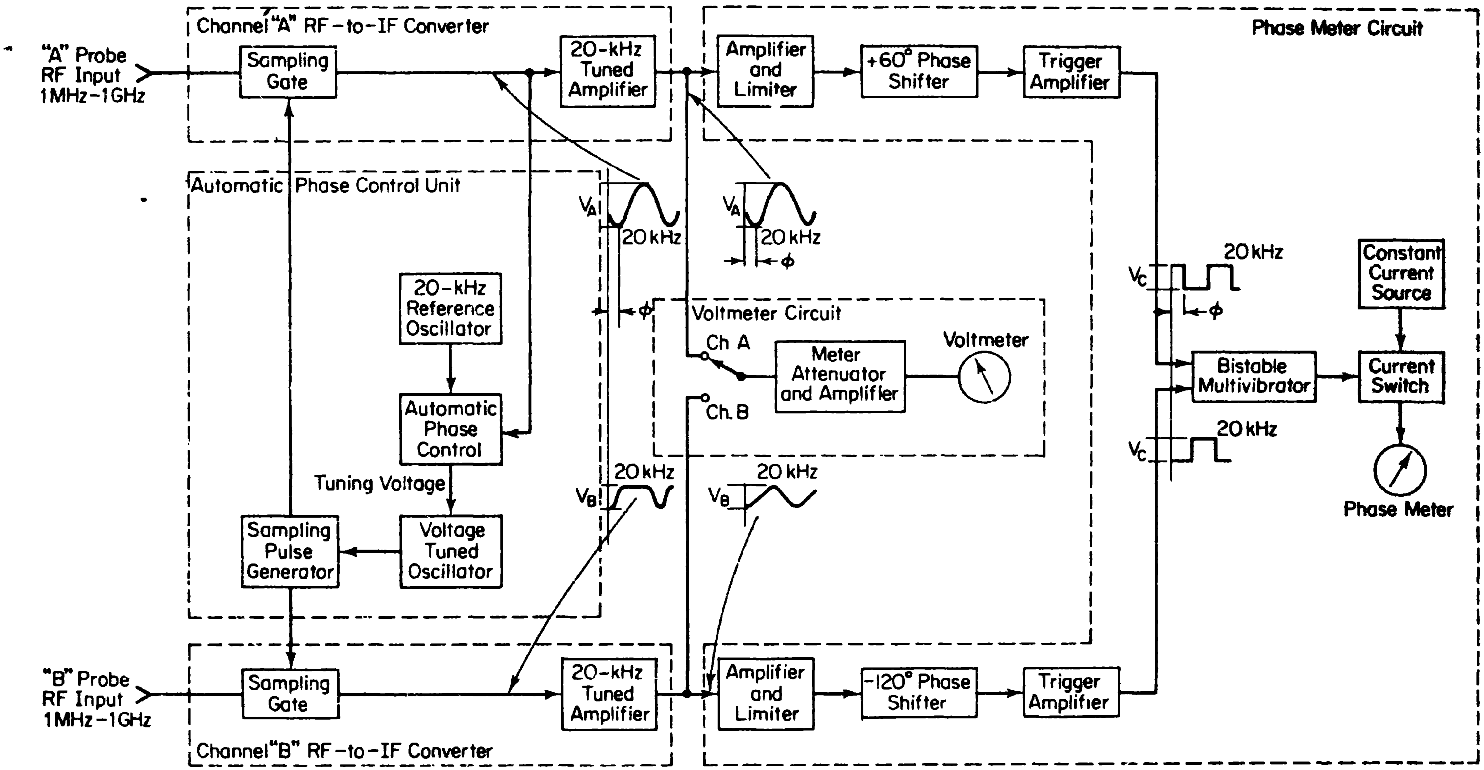


Figure 10-21. Block diagram of the vector voltmeter. Modified from HP Model 8405A Vector voltmeter. (*Courtesy Hewlett-Packard Co.*)

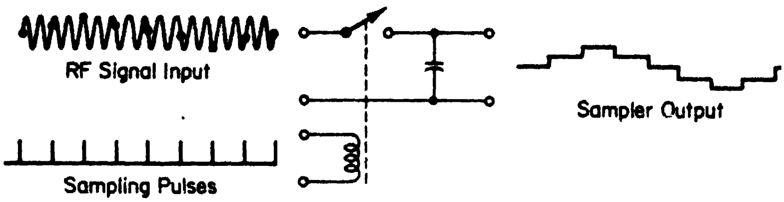


Figure 10-22
Simplified diagram of a sampler.

matic phase control section. This section frequency and phase locks the channel A IF signal to a 20-kHz reference oscillator. To get initial locking, the phase control section applies a ramp voltage to the VTO. This ramp voltage sweeps the sampling rate until channel A IF is 20 kHz and in phase with the reference oscillator. Then the sweep stops and channel A IF is held in phase with the reference oscillator.

The *tuned amplifier* passes only the 20-kHz fundamental component of the IF signal of each channel. The output of each tuned amplifier then consists of a signal which has retained its original phase relationship with respect to the signal in the other channel and also its correct amplitude relationship. The two filtered IF signals are made available to the *voltmeter* circuit by means of a front-panel controlled manual switch, marked *channel A* and *channel B*.

The voltmeter circuit contains an input attenuator, providing a selection of the appropriate input range. This attenuator is again a front panel control, marked *amplitude range*. The meter amplifier consists of a stable fixed-gain feedback amplifier, followed by a rectifier and a filter section. The rectified signal is applied to a dc voltmeter.

To determine the phase difference between the two IF signals, the tuned amplifiers are followed by the *phase meter circuit*. Each channel is first amplified and then limited, resulting in square-wave signals at the inputs to the *IF phase-shifting* circuits. The circuit in channel A shifts the phase of the square-wave signal by +60°; the circuit in channel B shifts the phase of its signal by −120°. Both phase shifts are accomplished by a combination of capacitive networks and inverting and noninverting amplifiers, whose vector-sum outputs provide the desired phase shift.

The outputs of the phase-shift circuits are amplified and clipped, producing square waveforms, and applied to the *trigger* circuits. These circuits convert the square-wave input signals to positive spikes with very fast rise times. The phase relationship of the pulses in each channel is maintained throughout this process. The *bistable multivibrator* is triggered by pulses from both channels. Channel A is connected to the *set* input of the MV; channel B is connected to the *reset* input of the MV. If the initial phase shift between the RF signals at the probes was 0°, the trigger pulses into the multivibrator are 180° out of phase owing to the action of the phase-shift circuits. The MV then produces a square-wave output voltage which is symmetrical about zero. Any phase shift at the RF probes carries through the entire system and varies the trigger pulses from their 180° relationship, producing an asymmetrical waveform.

The (asymmetrical) square wave controls the *current switch,* which is a transistor switched into conduction by the negative portion of the square wave. The switch connects the *constant current supply* to the *phase meter.* At 0° phase shift at the RF input, the switch is turned off and on for equal amounts of time and the current supply is adjusted to cause the meter to read 0° or center scale. Any RF phase shift results in an asymmetrical waveform and allows either more or less current to the phase meter, depending on whether the phase shift caused the negative half cycle of the square wave to be larger or smaller. An input phase shift of 180° would cause the square wave to collapse into either a positive or a negative dc voltage and the switch would then allow no current or maximum current to the phase meter. These maximum deviations from the center reading of 0° are marked on the meter face as +180° and −180°.

The *phase range* can be selected by a front panel switch which places a shunt across the phase meter and changes its sensitivity.

The instrument contains a power supply section, which is not shown on the block diagram of Fig. 10-21. The power supply generates all the necessary supply voltages for the various sections of the instrument.

Calibration procedures and the testing of performance specifications vary from one instrument to the next. Complete descriptions of the various tests are given in the manual of the instrument and usually include the procedure and instrumentation needed for such tests.

10-10 The Q Meter

The Q meter is an instrument designed for measuring the characteristics of coils and capacitors. The measuring principle of this useful laboratory instrument is based on the familiar characteristics of a series-resonant circuit; namely, that the voltage across the coil or the capacitor is equal to the applied voltage times the Q of the circuit. If a fixed voltage is applied to the circuit, a voltmeter across the capacitor can be calibrated to read Q directly.

The voltage and current relationships of a series resonant circuit are shown in Fig. 10-23. At resonance, the following conditions are true:

$$X_C = X_L$$

$$E_C = IX_C = IX_L$$

$$E = IR$$

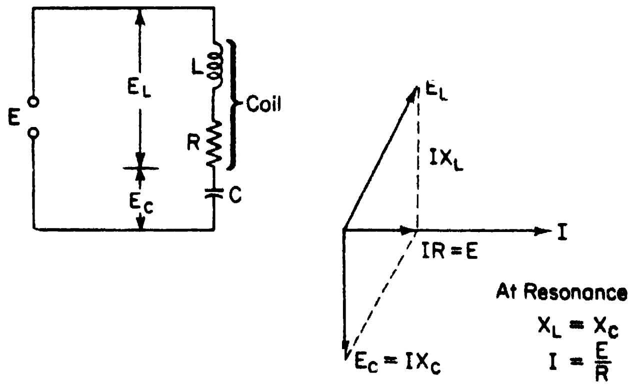


Figure 10-23
A series resonant circuit.

where X_C = capacitive reactance
X_L = inductive reactance
E_C = voltage across the capacitor
I = circuit current
R = coil resistance
E = applied voltage.

The magnification of the complete circuit, by definition, is Q, where

$$Q = \frac{X}{R} = \frac{E_C}{E} \tag{10-7}$$

Therefore, if E is maintained at a constant and known level, a voltmeter can be connected across the capacitor and can be calibrated directly in terms of the circuit Q.

A practical Q meter circuit is shown in Fig. 10-24. The self-contained oscillator has a frequency range from 50 kHz to 50 MHz and delivers current to a low-value shunt resistance. The voltage across the shunt corresponds to

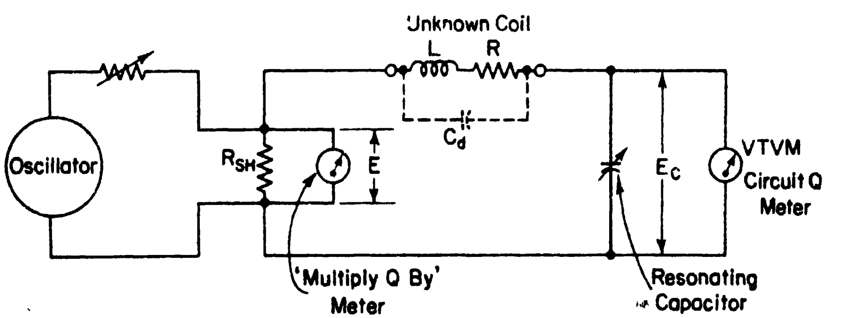


Figure 10-24
A basic Q-meter circuit.

E in Fig. 10-23 and is measured with a thermocouple meter, marked "Multiply Q by." The value of the shunt is very low, typically on the order of 0.02 Ω. It introduces almost no resistance into the oscillatory circuit and therefore represents a voltage source of magnitude, E, with a very small (in most cases negligible) internal resistance. The voltage across the variable capacitor, corresponding to E_C in Fig. 10-23, is measured by a VTVM with a scale calibrated directly in Q values.

To make a measurement, the unknown coil in Fig. 10-24 is connected to the test terminals, and the circuit is tuned to resonance either by setting the oscillator to a given frequency and varying the internal resonating capacitor or by presetting the capacitor to a desired value and adjusting the frequency controls. The Q meter is calibrated to read Q directly when the "Multiply Q by" meter is set to the unity index mark.

The *indicated Q* (which is the resonant reading on the "circuit Q" meter) is called the *circuit Q* because the losses of the resonating capacitor, voltmeter, and insertion resistor are all included in the measuring circuit. The *effective Q* of the measured coil will be somewhat greater than the *indicated Q*. This difference can generally be neglected, except in certain cases where the resistance of the coil is relatively small in comparison with the value of the insertion resistor. (A solution to this problem is discussed in Example 10-6.)

The inductance of the coil can be calculated from the known values of frequency (f) and resonating capacitance (C), since

$$X_L = X_C \qquad \text{and} \qquad L = \frac{1}{(2\pi f)^2 C}\, H \tag{10-8}$$

There are three methods for connecting unknown components to the test terminals of a Q meter. The type of component and its size determines the method of connection.

(*a*) *The Direct Connection, Fig. 10-24.* Most coils can be connected directly across the test terminals, exactly as shown in the basic Q meter circuit of Fig. 10-24. The circuit is resonated by adjusting either the oscillator frequency or the resonating capacitor. The indicated Q is read directly from the "Circuit Q" meter, modified by the setting of the "Multiply Q by" meter. When the last meter is set at the unity mark, the "Circuit Q" meter reads the correct value of Q directly.

(*b*) *The Series Connection, Fig. 10-25(a).* Low-impedance components, such as low-value resistors, small coils, and large capacitors, are measured in series with the measuring circuit. Figure 10-25(a) shows the connections. The component to be measured, here indicated by [Z], is placed in series with a stable *work coil* across the test terminals. (The work coil is usually supplied with the instrument.) Two measurements are made: the unknown is first short-circuited by a small *shorting strap* and the circuit is resonated, establishing a reference condition. The values of the tuning capacitor (C_1) and the indicated Q (Q_1) are noted. In the second measurement, the shorting strap is removed and the circuit is retuned, giving a new value for the tuning capacitor (C_2) and a change in the value of Q (ΔQ) from Q_1 to Q_2.

For the reference condition,

$$X_{C_1} = X_L \quad \text{or} \quad \frac{1}{\omega C_1} = \omega L \tag{10-9}$$

and, neglecting the measuring circuit resistance,

$$Q_1 = \frac{\omega L}{R} = \frac{1}{\omega C_1 R} \tag{10-10}$$

For the second measurement, the unknown reactance (X_S) can be expressed in terms of the new value of the tuning capacitor (C_2) and the in-circuit value of the inductor (L) and

$$X_S = X_{C_2} - X_L \quad \text{or} \quad X_S = \frac{1}{\omega C_2} - \frac{1}{\omega C_1} \tag{10-11}$$

and

$$X_S = \frac{C_1 - C_2}{\omega C_1 C_2} \tag{10-12}$$

(inductive if $C_1 > C_2$)

(capacitive if $C_1 < C_2$)

The resistive component of the unknown reactance may be found in terms of the computed reactance (X_S) and the indicated values of circuit Q, since

$$R_1 = \frac{X_1}{Q_1} \quad \text{and} \quad R_2 = \frac{X_2}{Q_2} \quad \text{or}$$

$$R_S = R_2 - R_1 = \frac{1}{\omega C_2 Q_2} - \frac{1}{\omega C_1 Q_1} \quad \text{and}$$

$$R_S = \frac{C_1 Q_1 - C_2 Q_2}{\omega C_1 C_2 Q_1 Q_2} \tag{10-13}$$

If the unknown is *purely resistive*, the setting of the tuning capacitor would not have changed in the measuring process, and $C_1 = C_2$. The equation for resistance reduces to

$$R_S = \frac{Q_1 - Q_2}{\omega C_1 Q_1 Q_2} = \frac{\Delta Q}{\omega C_1 Q_1 Q_2} \tag{10-14}$$

If the unknown is a *small inductor*, the value of the inductance is found from Eq. (10-12) and equals

$$L_S = \frac{C_1 - C_2}{\omega^2 C_1 C_2} \tag{10-15}$$

The Q of the coil is found from Eq. (10-12) and Eq. (10-13) since, by definition

$$Q_S = \frac{X_S}{R_S} \quad \text{and}$$

$$Q_s = \frac{(C_1 - C_2)(Q_1 Q_2)}{C_1 Q_1 - C_2 Q_2} \tag{10-16}$$

If the unknown is a *large capacitor,* its value is determined from Eq. (10-12), and

$$C_S = \frac{C_1 C_2}{C_2 - C_1} \tag{10-17}$$

The Q of the capacitor may be found using Eq. (10-16).

(*c*) *The Parallel Connection, Fig. 10-25(b).* High-impedance components, such as high-value resistors, certain inductors, and small capacitors, are measured by connecting them in parallel with the measuring circuit. Figure 10-25(b) shows the connections. Before the unknown is connected, the circuit is resonated, using a suitable work coil, to establish reference values for Q and C (Q_1 and C_1). Then, when the component under test is connected to the circuit, the capacitor is readjusted for resonance, and one gets new values for the tuning capacitance (C_2) and a change in the value of the circuit Q (ΔQ) from Q_1 to Q_2.

In a parallel circuit, computation of the unknown impedance is best approached in terms of its parallel components X_p and R_p, as indicated in Fig. 10-25(b). At the initial resonance condition, when the unknown is not yet connected into the circuit, the working coil (L) is tuned by the capacitor (C_1). Therefore,

$$\omega L = \frac{1}{\omega C_1} \quad \text{and} \tag{10-18}$$

$$Q_1 = \frac{\omega L}{R} = \frac{1}{\omega C_1 R} \tag{10-19}$$

When the unknown impedance is now connected into the circuit and the capacitor is tuned for resonance, the reactance of the working coil (X_L) equals the parallel reactances of the tuning capacitor (X_{C_2}) and the unknown (X_p). Therefore,

$$X_L = \frac{(X_{C_2})(X_p)}{X_{C_2} + X_p}$$

which reduces to

$$X_p = \frac{1}{\omega(C_1 - C_2)} \tag{10-20}$$

If the unknown is inductive, $X_p = \omega L_p$, and Eq. (10-20) yields the value of the unknown impedance:

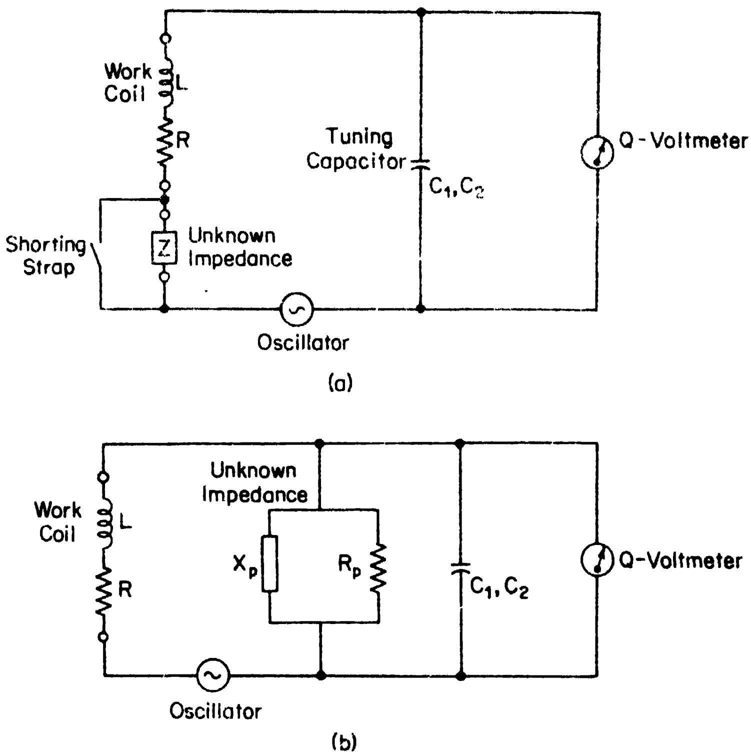


Figure 10-25
Measurement of an unknown impedance with the Q meter. (a) Series connection, used to measure low-impedance components. (b) Parallel connection, used to measure high-impedance components. C_1 is the value of the tuning capacitor for the initial balance and C_2 is the capacitor value after the unknown has been connected.

$$L_p = \frac{1}{\omega^2(C_1 - C_2)} \tag{10-21}$$

If the unknown is capacitive, $X_p = 1/\omega C_p$ and Eq. (10-20) yields the value of the unknown capacitor:

$$C_p = C_1 - C_2 \tag{10-22}$$

In a parallel resonant circuit, the total resistance at resonance is equal to the product of the circuit Q and the reactance of the coil. Therefore,

$$R_T = Q_2 X_L$$

or, by substitution of Eq. (10-18),

$$R_T = Q_2 X_{C_1} = \frac{Q_2}{\omega C_1} \tag{10-23}$$

The resistance (R_p) of the unknown impedance is most easily found by computing the *conductances* in the circuit of Fig. 10-25(b).

Let G_T = total conductance of the resonant circuit
G_p = conductance of the unknown impedance
G_L = conductance of the working coil.

Then,

$$G_T = G_p + G_L \qquad \text{or} \qquad G_p = G_T - G_L \tag{10-24}$$

From Eq. (10-23),

$$G_T = \frac{1}{R_T} = \frac{\omega C_1}{Q_2}$$

Therefore,

$$\frac{1}{R_p} = \frac{\omega C_1}{Q_2} - \frac{R}{R^2 + \omega^2 L^2}$$

$$\frac{1}{R_p} = \frac{\omega C_1}{Q_2} - \left(\frac{1}{R}\right)\left(\frac{1}{1 + \omega^2 L^2/R^2}\right)$$

$$\frac{1}{R_p} = \frac{\omega C_1}{Q_2} - \frac{1}{RQ_1^2}$$

Substituting Eq. (10-19) in the foregoing expression,

$$\frac{1}{R_p} = \frac{\omega C_1}{Q_2} - \frac{\omega C_1}{Q_1}$$

and after simplifying,

$$R_p = \frac{Q_1 Q_2}{\omega C_1(Q_1 - Q_2)} = \frac{Q_1 Q_2}{\omega C_1 \, \Delta Q} \tag{10-25}$$

The Q of the unknown is then found by using Eq. (10-20) and Eq. (10-25) so that

$$Q_p = \frac{R_p}{X_p} = \frac{(C_1 - C_2)(Q_1 Q_2)}{C_1(Q_1 - Q_2)} = \frac{(C_1 - C_2)(Q_1 Q_2)}{C_1 \, \Delta Q} \tag{10-26}$$

(*d*) *Sources of Error.* Probably the most important factor affecting measurement accuracy, and the most often overlooked, is the *distributed capacitance* or *self-capacitance* of the measuring circuit. The presence of distributed capacitance in a coil modifies the *actual or effective* Q and the inductance of the coil. At the frequency at which the self-capacitance and the inductance of the coil are resonant, the circuit exhibits a purely resistive impedance. This characteristic may be used for measuring the distributed capacitance.

One rather easy method of solving for the distributed capacitance (C_d) of a coil involves making two measurements at different frequencies. The coil under test is connected directly to the test terminals of the Q meter, as shown

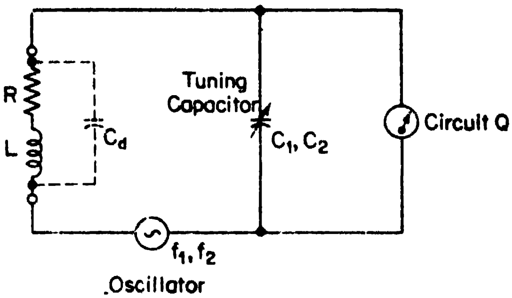


Figure 10-26 Determination of the distributed capacitance of an inductor.

in the circuit of Fig. 10-26. The tuning capacitor is set to a high value, preferably to its maximum position, and the circuit is resonated by adjustment of the oscillator frequency. Resonance is indicated by a maximum deflection on the "circuit Q" meter. The values of the tuning capacitor (C_1) and the oscillator frequency (f_1) are noted. The frequency is then increased to twice its original value ($f_2 = 2f_1$) and the circuit is retuned by adjusting the resonating capacitor (C_2).

The resonant frequency of an LC-circuit is given by the well-known equation

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{10-27}$$

At the initial resonance condition, the capacitance of the circuit equals $C_1 + C_d$ and the resonant frequency equals

$$f_1 = \frac{1}{2\pi\sqrt{L(C_1 + C_d)}} \tag{10-28}$$

After adjusting the oscillator and the tuning capacitor, the capacitance of the circuit is $C_2 + C_d$, and the resonant frequency equals

$$f_2 = \frac{1}{2\pi\sqrt{L(C_2 + C_d)}} \tag{10-29}$$

Since $f_2 \quad 2f_1$, Eqs. (10-28) and (10-29) are related so that

$$\frac{1}{2\pi\sqrt{L(C_1 + C_d)}} = \frac{2}{2\pi\sqrt{L(C_1 + C_d)}}$$

and

$$\frac{1}{C_2 + C_d} = \frac{4}{C_1 + C_d}$$

Solving for the distributed capacitance,

$$C_d = \frac{C_1 - 4C_2}{3} \tag{10-30}$$

Example 10-3: The self-capacitance of a coil is to be measured using the procedure just outlined. The first measurement is at $f_1 = 2$ MHz and $C_1 = 460$ pF. The second measurement, at $f_2 = 4$ MHz, yields a new value of tuning capacitor, $C_2 = 100$ pF. Find the distributed capacitance, C_d.

SOLUTION: Using Eq. (10-30),

$$C_d = \frac{C_1 - 4C_2}{3} = \frac{460 - 400}{3} = 20 \text{ pF}$$

Example 10-4: Compute the value of self-capacitance of a coil, when the following measurements are made. At frequency $f_1 = 2$ MHz, the tuning capacitor is set at 450 pF. When the frequency is increased to 5 MHz, the tuning capacitor is tuned at 60 pF.

SOLUTION: Since $f_2 = 2.5 f_1$, Eqs. (10-28) and (10.29) are related as follows:

$$\frac{1}{2\pi\sqrt{L(C_2 + C_d)}} = \frac{2.5}{2\pi\sqrt{L(C_1 + C_d)}}$$

This reduces to

$$\frac{1}{C_2 + C_d} = \frac{6.25}{C_1 + C_d}$$

Solving for C_d,

$$C_d = \frac{C_1 - 6.25C_2}{5.25}$$

Substituting the values for $C_1 = 450$ pF and $C_2 = 60$ pF, the value of the distributed capacitance is found to be $C_d = 14.3$ pF.

The *effective Q of a coil with distributed capacitance* is less than the true Q by a factor that depends on the value of the self-capacitance and the resonating capacitor. It can be shown that

$$\text{true } Q = Q_e\left(\frac{C + C_d}{C}\right) \tag{10-31}$$

where Q_e = effective Q of the coil
C = resonating capacitance
C_d = distributed capacitance.

The *effective Q* can usually be considered the *indicated Q*.

For many measurements, the *residual or insertion resistance* (R_{sh}) of the Q meter measuring circuit of Fig. 10-24 is sufficiently small to be considered negligible. Under certain circumstances, it can contribute an error to the measurement of Q. The effect of the insertion resistor on the measurement depends on the magnitude of the unknown impedance and, of course, on the size of the insertion resistor. For instance, the 0.02 Ω of insertion resistance may be safely neglected in comparison with a coil resistance of 10 Ω, but it assumes importance when compared to a coil resistance of 0 1 Ω. The effect of the 0.02-Ω insertion resistance is illustrated by Examples 10-5 and 10-6.

Example 10-5: A coil with a resistance of 10 Ω is connected in the "direct-measurement" mode. Resonance occurs when the oscillator frequency is 1.0 MHz and the resonating capacitor is set at 65 pF. Calculate the percentage error introduced in the calculated value of Q by the 0.02-Ω insertion resistance.

Solution: The *effective* Q of the coil equals

$$Q_e = \frac{1}{\omega CR} = \frac{1}{(2\pi)(10^6)(65 \times 10^{-12})(10)} = 245$$

The *indicated* Q of the coil equals

$$Q_i = \frac{1}{\omega C(R + 0.02)} = 244.5$$

The percentage error is then $\dfrac{245 - 244.5}{245} \times 100\% = 0.2\%$

Example 10-6: Repeat the problem of Example 10-5 for the following conditions:

The coil resistance is 0.1 Ω.
The frequency at resonance is 40 MHz.
The tuning capacitor is set at 135 pF.

Solution: The *effective* Q of the coil is

$$Q^e = \frac{1}{\omega CR} = \frac{295}{2\pi \times 40 \times 10^6 \times 135 \times 10^{-12} \times 0.1} = 259$$

The *indicated* Q of the coil is

$$Q_i = \frac{1}{\omega C(R + 0.02)} = 245$$

The percentage error equals

$$\frac{295 - 245}{295} \times 100\% = 17\%$$

Other sources of error include the *residual inductance* of the instrument, which is usually in the order of 0.015 μH and affects the measurement of only very small inductors ($< 0.5\ \mu$H); the *conductance of the* Q *voltmeter* has a slight shunting effect on the tuning capacitor at the higher frequencies, but can usually be neglected.

References

1. Thomas Harry E., and Carole A. Clarke, *Handbook of Electronic Instruments and Measurement Techniques*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1967.

2. Hewlett-Packard, Palo Alto, Calif., *Application Notes* and *Instrument Manuals*.

3. Stout, Melville B., *Basic Electrical Measurements*, 2nd ed., Chap. 10. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

4. Malvino, Albert Paul, *Electronic Instrumentation Fundamentals*, Chap. 11. New York: McGraw-Hill Book Company, 1967.

Problems

1. An average responding electronic ac voltmeter with an rms-calibrated scale is used to measure the following nonsinusoidal voltages: (a) dc voltage of 10 V; (b) square-wave voltage with an amplitude of 10 V and a duty cycle of 75%; (c) triangular voltage of symmetrical waveform and a peak value of 10 V. Calculate the percentage error indicated by the ac voltmeter for each of the voltages cited.

2. A 25-mA current meter with an internal resistance of 100 Ω is available for constructing an ac voltmeter with a voltage range of 200 V rms. Using four diodes in a bridge arrangement, where each diode has a forward resistance of 500 Ω and infinite reverse resistance, calculate the necessary series-limiting resistance for the 200-V rms voltage range.

3. In checking the distributed capacitance of a certain coil with the Q meter of Fig. 10-24, initial resonance is obtained with the resonating capacitor set to 450 pF. Resonance at twice the initial frequency is obtained with the tuning capacitor at 11 pF. Calculate the value of the distributed capacitance of the coil.

4. A coil with a resistance of 3 Ω is connected to the terminals of the Q meter of Fig. 10-24. Resonance occurs at an oscillator frequency of 5 MHz and resonating capacitance of 100 pF. Calculate the percentage error introduced by the insertion resistance $R_{sh} = 0.1\ \Omega$.

5. The following laboratory test equipment is available for calibrating

multimeters: (a) an accurate differential voltmeter; (b) a stable dc power supply; (c) a number of assorted precision resistors. With the aid of diagrams, suggest a measurement procedure for the calibration of the current ranges and the voltage ranges of a multimeter.

6. Design a range switch for the dc volt section of a balanced-bridge VTVM (see Fig. 10-6). The total resistance of the attenuator should be 11 MΩ. The attenuation should be so arranged that input voltages from 3 V to 1,000 V can be accommodated in the customary 1-3-10 sequence. The bridge circuit (see Fig. 10-5) requires 1 V at the grid of the balancing tube to cause full-scale meter deflection.

7. The differential voltage measurement of Fig. 10-10 uses a reference source with an internal resistance of 200 Ω and a terminal voltage of 3.0 V. The galvanometer has a current sensitivity of 1 mm/μA and internal resistance of 100 Ω. Calculate the emf of the unknown source, neglecting its internal resistance, if the galvanometer deflection is 250 mm.

8. For the measurement conditions given in Problem 7, calculate the resolution of the measurement set-up if the galvanometer deflection can be read to 1 mm.

9. For the differential voltage measurement of Problem 7, a second galvanometer has become available. This galvanometer has a current sensitivity of 5 mm/μA and internal resistance of 1,000 Ω. Calculate which of the two galvanometers provides the greatest sensitivity to unbalance, expressing the results in millimeters per microvolt.

ELEVEN

INSTRUMENTS FOR THE GENERATION AND ANALYSIS OF WAVEFORMS

11-1 Oscillators

(*a*) *Introduction.* Signal sources are described by several names: oscillators, test oscillators, signal generators, audio oscillators, and so on. These names vary with the design and intended use of the signal source. The *oscillator* is the one basic element common to all the sources: it generates a sine-wave signal of known frequency and known amplitude.

The name *test oscillator* generally describes an oscillator having a calibrated attenuator and an output monitor. The term *signal generator* is usually reserved for oscillators with *modulation* capability.

In selecting an oscillator to perform a certain function in a measurement situation, the user should be interested in some fundamental questions concerning the *performance characteristics* of the instrument:

(1) *Frequency range:* The oscillator should be able to supply both the the lowest and the highest frequency of interest. The frequency spectrum of laboratory instruments may range from 0.00005 Hz to 30 MHz or even higher.

(2) *Available output power or output voltage:* Some measurements

require large amounts of power; others merely require sufficient voltage output.

(3) *Output impedance:* Some oscillators have a low output impedance which can be converted to almost any desired impedance by the use of a resistive divider. Other instruments have a transformer-coupled output providing a balanced and isolated output circuit. Since many audio-range oscillators are used with 600-Ω input impedance systems, these oscillators are generally provided with a 600-Ω output attenuator.

(4) *Dial resolution and accuracy:* In the ideal case, the user should be able to set the tuning dial of the oscillator to a particular frequency with the assurance that the instrument will deliver that frequency at all times. On laboratory instruments the tuning dial may be precisely set by a vernier control. The accuracy with which the frequency tracks the tuning dial enters into the over-all accuracy figure.

(5) *Frequency stability:* The frequency stability of the oscillator determines its ability to maintain the selected frequency over a period of time. Component aging, temperature changes, and power supply variations all affect stability. Frequency stability can be improved in some instances by using large amounts of negative feedback and carefully selected components.

(6) *Amplitude stability:* Amplitude stability is important in some applications. The frequency response or amplitude variation as the frequency is changed is of particular interest when the oscillator is used for response measurements over a wide range of frequencies.

(7) *Distortion:* Distortion in the oscillator output signal is an inverse measure of the purity of the waveform. Distortion is undesirable because a harmonic of the test signal may enter the circuits under test and generate a false indication at the output. If the oscillator is used for distortion measurements, the amount of distortion that it contributes to the measurement should be far less than the distortion contributed by the circuits under test.

(*b*) *The Standard Wien Bridge* RC *Oscillator.** The Wien bridge RC oscillator has become the standard circuit for adjustable frequency test signals in the audiofrequency range. This oscillator is marked by simplicity and stability and is far less cumbersome than the LC-type oscillator or the beat-frequency oscillator. For fuller understanding of the operation of the Wien bridge oscil-

*Jacob Millman, *Vacuum Tube and Semiconductor Electronics* (New York: McGraw-Hill Book Company, Inc., 1958), pp. 479–482.

lator, one needs to define briefly the operating conditions for *feedback* oscillators:

(1) The frequency of a sinusoidal feedback oscillator is determined by the condition that the *loop phase shift* equal 0; in other words, the phase shift between the amplifier input signal and the signal which is returned through the amplifier and feedback circuit must be zero.

(2) The product of the amplifier gain (A) and the feedback factor (β) must be equal to or larger than 1.

These principles are consistent with the well-known feedback equation:

$$A_f = \frac{A}{1-\beta A} \tag{11-1}$$

where A_f = amplifier gain *with* feedback
A = amplifier gain *wtihout* feedback
β = feedback factor = $E_{\text{out}}/E_{\text{in}}$.

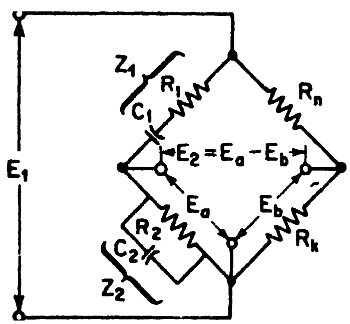


Figure 11-1
A Wien bridge circuit.

Consider now the Wien bridge circuit of Fig. 11-1, which acts as the feedback network between the amplifier output terminals and its input terminals. E_1 is the input voltage to the bridge (the output voltage of the amplifier) and E_2 is the output voltage of the bridge (the input voltage to the amplifier).

The first condition cited for oscillation indicates that the phase shift between E_1 and E_2 must be zero. This is the case *only* when the bridge is balanced, at *resonance*. The resonant frequency of the Wien bridge has been derived in Sec. 8-6 and is given by Eq. (8-41),

$$f_0 = \frac{1}{2\pi\sqrt{R_1R_2C_1C_2}}$$

In the practical case, where $R_1 = R_2$ and $C_1 = C_2$, Eq. (8-41) reduces to the more common form

$$f_0 = \frac{1}{2\pi RC} \tag{11-2}$$

Equation (11-2) indicates that bridge resonance is quite independent of the values of the other two bridge arms, R_n and R_k.

The second condition for oscillation requires that the product of the amplifier gain (A) and the feedback factor (β) be equal to or greater than 1. Since A, the amplifier gain, is a finite quantity ($A = 100$ would be a reasonable assumption), the feedback factor β must also be a finite quantity (1/100 under the assumption that $A = 100$). Since the bridge is to be used as the feedback network for an oscillator, the phase shift must be kept zero but the magnitude of β (defined as E_2/E_1) must not be zero. This means then that E_2 must not be zero.

The value of E_2 may be determined in terms of the bridge constants. At resonance, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have the same phase angle, since both R_n and R_k are purely resistive. The magnitude of $\mathbf{Z}_1$ is found by substituting Eq. (11-2) into the expression for $\mathbf{Z}_1$ where

$$\mathbf{Z}_1 = R - \frac{j}{\omega C} = (1 - j)R \tag{11-3}$$

Similarly,

$$\mathbf{Z}_2 = \frac{1}{(1/R) + j\omega C} = (1 - j)\frac{R}{2} \tag{11-4}$$

Therefore, the voltage E_a across $\mathbf{Z}_2$ equals

$$E_a = \frac{\mathbf{Z}_2}{\mathbf{Z}_1 + \mathbf{Z}_2} E_1 = \frac{1}{3} E_1 \tag{11-5}$$

and the voltage E_b cross R_k equals

$$E_b = \frac{R_k}{R_n + R_k} E_1 \tag{11-6}$$

If a null is desired, then R_n and R_k must be chosen such that $E_b = \frac{1}{3}E_1$, so that $E_2 = E_a - E_b = 0$. Then $R_k/(R_n + R_k) = \frac{1}{3}$, or $R_n = 2R_k$. In this case, however, E_2 must not be zero and, therefore, the ratio $R_k/(R_n + R_k)$ should be smaller than $\frac{1}{3}$. Let, for example,

$$\frac{E_b}{E_1} = \frac{R_k}{R_n + R_k} = \frac{1}{3} - \frac{1}{\delta} \tag{11-7}$$

where δ is a number larger than 3. Then

$$\beta = \frac{E_2}{E_1} = \frac{E_a - E_b}{E_1} = \frac{E_a}{E_1} - \left(\frac{1}{3} - \frac{1}{\delta}\right) \tag{11-8}$$

At the resonant frequency $f = f_0$; $E_a/E_1 = \frac{1}{3}$; and $\beta = 1/\delta$. The condition $A\beta = 1$ is then satisfied by making the gain of the amplifier, $A = \delta$.

Two important observations are made: The frequency of oscillation is exactly the null frequency of the balanced bridge, namely, $f_0 = 1/2\pi RC$—Eq. (11-2); at any other frequency, E_a is not in phase with E_1 and therefore $E_2 = E_a - E_b$ is not in phase with E_1 so that the condition $A\beta = 1$ is satisfied only at the one frequency f_0.

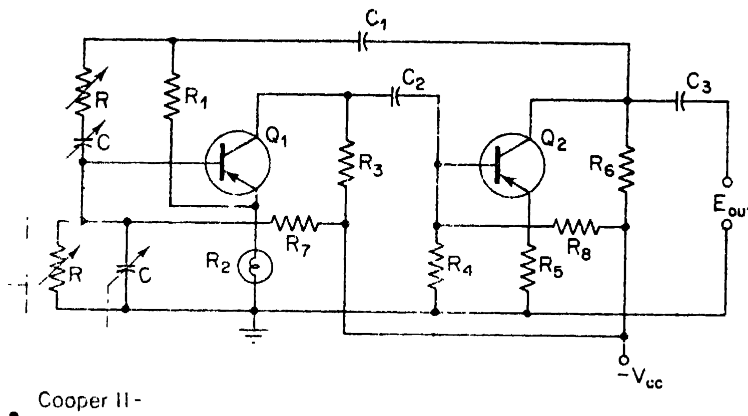


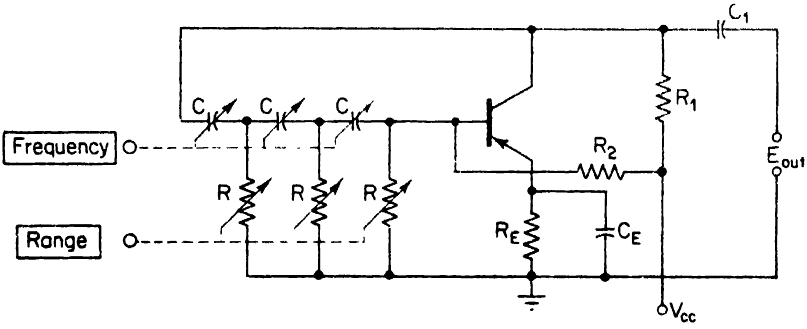
Figure 11-2
A two-stage Wien bridge oscillator. The ganged air capacitors provide variable frequency output and the ganged resistors form the range control.

The circuit diagram of a simple, but practical, Wien bridge oscillator is given in Fig. 11-2. The bridge consists of R and C in series, R and C in parallel, R_1 and R_2. Feedback is applied from the collector of Q_2 through coupling capacitor C_1 to the top of the bridge circuit. C_1 is large enough to introduce no phase shift at the lowest frequency of oscillation. Resistance R_2 serves the dual purpose of emitter resistor of Q_1 and element of the Wien bridge. Continuous variation of the frequency is accomplished by varying simultaneously the two capacitors C, which are variable air capacitors on a common shaft. Different frequency ranges are provided by switching in different values for the two *identical* resistors, R.

The amplitude of oscillation is determined by the extent to which βA is greater than unity. If β is fixed, the amplitude is determined by A, increasing as A increases until further increase is limited by the nonlinear behavior of the transistors. Regulation of the amplitude is provided by resistor R_2, which provides a variable β. Resistor R_2 is usually a *tungsten-filament* light bulb, acting as a variable resistance element. If the output of the amplifier tends to increase, the increased current through R_2 raises its temperature and increases its resistance. Then, from Eq. (11-8), β would decrease and would tend to keep the product $A\beta$ constant, thereby regulating the amplifier output to a constant level. The *thermal lag* of the filament of the tungsten lamp causes its resistance to remain almost constant during the course of an alternating cycle of output voltage or current. At very low frequencies, however, the thermal lag may not be large enough and the resistance value of the lamp may change during the cycle. In this case a *thermistor* may be used, which has sufficient bulk to provide adequate thermal lag. Since the thermal coefficient of the thermistor is negative (its resistance decreases with an increase in temperature), the thermistor would have to be placed in the other bridge arm (R_1 instead of R_2).

The Wien bridge oscillator is capable of *stable* oscillations with low distortion output. With the addition of a power amplifier to isolate the oscillator from the load, the circuit is used to provide test signals for a variety of applications. The upper frequency of the Wien bridge oscillator is limited by the amplitude and phase-shift characteristics of the amplifier and is usually in the order of 100 kHz. Above this frequency, the well-known RF oscillator circuits may be used, such as the *Colpitts* or the *Hartley* oscillator. Again because of its simplicity and purity of waveform, the *phase-shift oscillator* plays an important part in the construction of laboratory instruments with high-frequency capabilities.

(*c*) *The Phase-Shift Oscillator.** The circuit of Fig. 11-3 shows a simple



Cooper 11-3

Figure 11-3
RC phase-shift oscillator, using three cascaded RC sections to provide 180° phase shift between collector output and base input. The three ganged air capacitors provide continuous frequency adjustment and the three resistors are switched in decade steps to provide the oscillator's range control.

phase-shift oscillator, capable of generating sinusoidal output voltages at frequencies up to several hundred kilohertz. The circuit consists of a single transistor as the amplifier stage, followed by three cascaded arrangements of capacitor C and resistor R which provide the feedback voltage from the output of the amplifier back to the input.

The single transistor in the circuit shifts by 180° the phase of any voltage appearing at its base. The RC network provides the additional phase shift of 180° necessary for oscillation. At some particular frequency, the phase shift of the total RC network will be exactly 180°, and at this frequency the total phase

*Millman, *op. cit.*, pp. 476–477.

shift from the base of the transistor, around the circuit, back to the base will be exactly zero. Provided, then, that the amplification of the transistor is sufficiently large, the circuit will oscillate at that frequency.

The feedback factor β, defined as the ratio of output voltage, E_{out}, to input voltage, E_{in}, can be found by applying conventional network theory to the RC combination. This analysis yields

$$\beta = \frac{E_{out}}{E_{in}} = \frac{1}{1 - 5\alpha^2 - j(6\alpha - \alpha^3)} \tag{11-9}$$

where $\alpha = \dfrac{1}{\omega RC}$.

The phase shift between E_{out} and E_{in} will be 180° when the imaginary portion of the denominator of Eq. (11-9) equals 0, or when $\alpha^2 = 6$. The frequency corresponding to this situation then equals

$$f = \frac{1}{2\pi RC\sqrt{6}} \tag{11-10}$$

At this frequency of oscillation, the feedback factor is $\beta = -\frac{1}{29}$. Satisfying the condition for oscillation that the product βA shall not be less than unity, it is necessary for A to be *at least* 29. Therefore the transistor or vacuum tube chosen as the amplifier in this circuit should be able to supply this amplification.

The phase-shift oscillator is suited to a wide frequency range, from a few hertz to several hundred kilohertz. The upper frequency is limited because the impedance of the RC phase-shifting network may become so small that it loads the amplifier so heavily that corrective measures are required, making the circuit unnecessarily complicated. On the other hand, frequencies in the order of 1 Hz are easily obtained by using commercially available large values for R and C. The phase-shift oscillator has a decided advantage over LC-tuned oscillators in the low-frequency range since the large inductors required for the LC oscillators would be quite impractical.

(*d*) *Other Sinusoidal Oscillators.** Oscillators in the RF frequency range and above often use one of the conventional LC *tank-circuit*-type oscillators. The principle of operation of these circuits is simple and almost identical for any one of the many variations of the basic circuit and can be described as follows: The parallel LC combination is excited into oscillation and the ac voltage across the tank circuit is amplified by either a transistor or a vacuum tube. Part of the amplified ac voltage is fed back to the tank circuit by inductive or capacitive coupling to compensate for the losses in power in the tank circuit. This

*Millman, *op. cit.*, pp. 482–484.

regenerative feedback causes a constant-amplitude output voltage at the resonant frequency of the tank circuit, which is described by the equation

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{11-11}$$

These oscillators can operate at very high frequencies, up to several hundred megahertz. Specially designed tubes, such as klystrons and magnetrons, extend the frequency range in the gigahertz range.

It is instructive to analyze qualitatively the operation of one basic circuit, the *Armstrong oscillator*, given in Fig. 11-4. When the supply voltage E_{bb} is

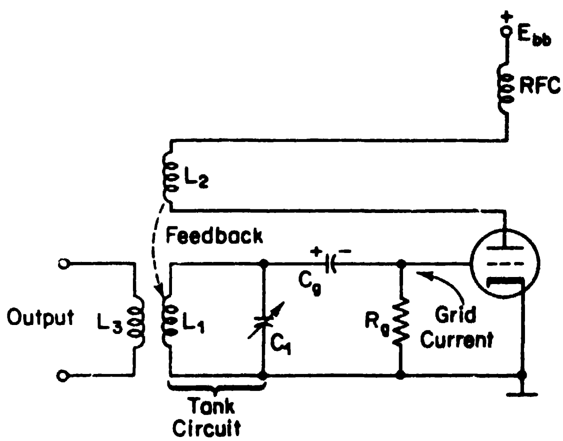


Figure 11-4
The Armstrong oscillator.

first applied, a plate current is produced. Since coil L_2 is inductively coupled to coil L_1, the increasing plate current induces a voltage across L_1 in such a direction that the grid of the tube is driven positive; the top of the coil has a positive polarity. This causes the plate current to increase at a faster rate and the induced voltage increases further. A high positive voltage is built up across the tank circuit, and capacitor C_1 charges with a positive polarity on its top plate. Since, at the same time, the grid is driven positive, grid current charges capacitor C_g to the peak value of the induced voltage with the polarity as shown. As the tube begins to approach saturation, the rate of rise in plate current decreases and therefore the induced voltage decreases. C_g must therefore discharge and does this through R_g, making the grid of the tube negative. This results in a *chain reaction*.

The plate current starts to decrease from its maximum (saturated) value and the magnetic field of L_2 collapses. This induces a *negative voltage* across coil L_1. In the meantime, C_g is still discharging and this drives the tube into cutoff so that plate current ceases entirely. The negative induced voltage across L_1 causes C_1 to discharge and then charge up again to the peak value of the negative induced voltage. Now the tube is cut off and C_1 is charged with a negative polarity on its top plate. During the next half cycle of action, the tank circuit brings the tube out of its cutoff condition. C_1 starts to discharge through L_1 and the grid of the tube is raised in potential until the tube starts to conduct again. As soon as the tube conducts, energy is transferred from the plate circuit of the tube to the tank circuit, and C_1 charges up again to the peak value of the now positive voltage. The entire cycle repeats itself in the manner described. The frequency of oscillation is governed by the charge and discharge characteristics of the tank circuit and is given by Eq. (11-11).

The *Hartley oscillator* of Fig 11-5 uses only one coil with a tap which

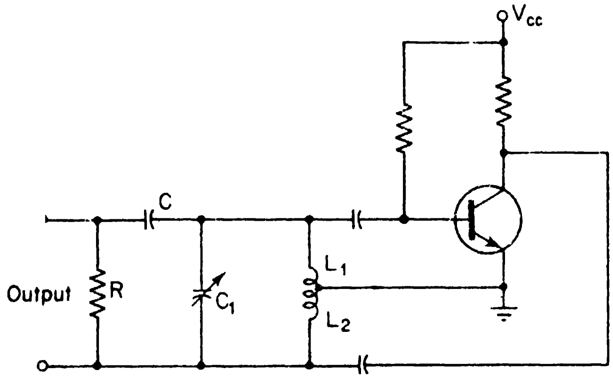


Figure 11-5
The Hartley oscillator.

corresponds to the common ac ground of the Armstrong circuit. The tuning capacitor C_1 is now shunted across the entire coil ($L_1 + L_2$). Since the coil tap remains at ground the rotor of this variable capacitor can no longer be grounded. The output voltage is available through the *RC* coupling circuit instead of by the inductive coupling as in the case of the Armstrong oscillator. This

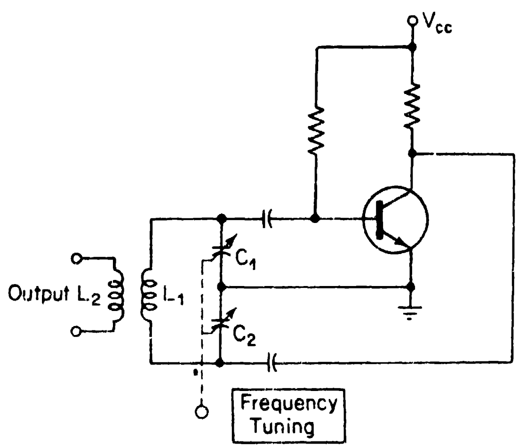


Figure 11-6
The Colpitts oscillator.

has no bearing on the operation of the circuit; either type of output coupling can be used.

The *Colpitts oscillator* of Fig. 11-6 is another variation of the basic Armstrong circuit. The tank circuit now consists of one inductor, L_1, and two capacitors in series (C_1 and C_2). Notice that, except for the method of tapping in the tank circuit, this circuit is identical to the circuit of the Hartley oscillator. The amount of feedback in the Colpitts circuit depends on the relative values of capacitors C_1 and C_2. The smaller C_1 is, the greater the feedback. As the tuning is varied, both capacitor values increase or decrease simultaneously, but the ratio of the two values remains fixed.

11-2 Pulse and Square-Wave Generators*

(*a*) *General Information.* Pulse and square-wave generators are most often used with a CRO as the measuring device. The waveforms portrayed by the CRO either at the output or at pertinent points in the system under test, provide both qualitative and quantitative information of the system or device under test.

The fundamental difference between a pulse generator and a square-wave generator concerns the *duty cycle.* Duty cycle is defined as the ratio of the average value of the pulse over one cycle to the peak value of the pulse. Since the average value and the peak value are inversely related to their time duration, the duty cycle can be defined in terms of the *pulsewidth* and the *period or pulse repetition time:*

$$\text{duty cycle} = \frac{\text{pulsewidth}}{\text{period}} \tag{11-12}$$

Square-wave generators produce an output voltage with equal *on* and *off* times, so that their duty cycle equals 0.5 or 50 per cent. The duty cycle remains at 50 per cent as the frequency of oscillation is varied.

The duty cycle of a *pulse generator,* on the other hand, may vary; very short duration pulses give a low duty cycle, so that the pulse generator generally can supply more power during its *on* period than a square-wave generator. Short duration pulses reduce the power dissipation in the component under test. For instance, measurements of transistor gain can be made with pulses short enough to prevent junction heating, and the resulting effect of heat on the gain of the transistor is then greatly minimized.

Square-wave generators are used where the low-frequency characteristics of a system are being investigated: testing audio systems, for instance. Square waves are also preferable to short-duration pulses if the transient response of a system requires some time to settle down.

(*b*) *Pulse Characteristics and Terminology*. In selecting a pulse generator or square-wave generator, the quality of the pulse is of primary importance. A test pulse of high quality insures that any degradation of the displayed pulse may be attributed to the circuit under test and not to the test instrument itself.

The pertinent characteristics of a pulse are shown in Fig. 11-7. The specifications describing these characteristics are usually given in the instrument manual or manufacturer's specification sheets.

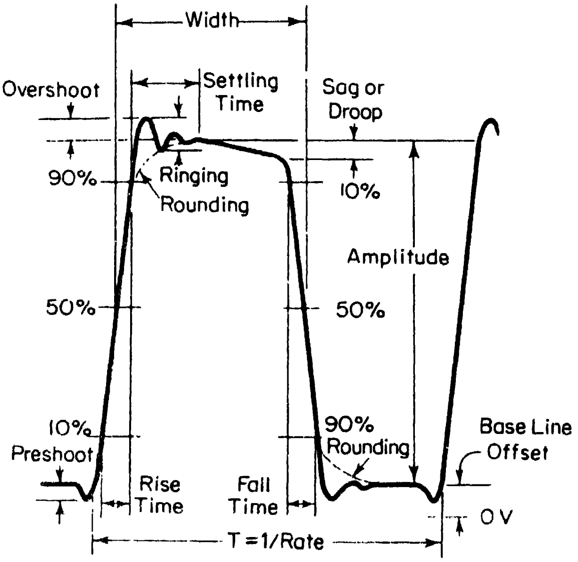


Figure 11-7
Characteristics of a pulse.

The time required for the pulse to increase from 10 per cent to 90 per cent of its normal amplitude is called the *rise time* (t_r). Similarly, the time required for the pulse to decrease from 90 per cent to 10 per cent of its maximum amplitude is called the *fall time* (t_f). In general, the rise time and the fall time of the pulse should be significantly faster than the circuit or the component under test.

When the initial amplitude rise exceeds the correct value, *overshoot* occurs. The overshoot may be visible as a single *pip*, or *ringing* may occur. When the maximum amplitude of the pulse is not constant but decreases slowly, the pulse is said to *droop or sag*. Any overshoot, ringing, or sag in the test pulse should be known to avoid confusion with similar phenomena caused by the test circuit.

The maximum pulse *amplitude* is of prime concern if appreciable input power is required by the tested circuit, such as, for example, a magnetic core memory. At the same time, the attenuation range of the instrument should be adequate to prevent overdriving the test circuit as well as to simulate actual operating conditions.

The range of the frequency control or *pulse repetition rate* (PRR) is of concern if the tested circuit can operate only within a certain range of pulse rates or if a variation in the rate is needed. Some of the more sophisticated pulse generators are capable of repetition rates of up to 100 MHz for testing "fast" circuits; others have a *pulse-burst* feature which allows a train of pulses rather than a continuous output to be used to check a system.

Some pulse generators can be *triggered* by externally applied trigger signals, very similar to the trigger features found in laboratory CROs. Conversely, the output of the pulse generator or square-wave generator may be used to provide trigger pulses for operating external circuits. The output trigger circuitry of the pulse generator then allows the trigger pulse to occur either before or after the main output pulse.

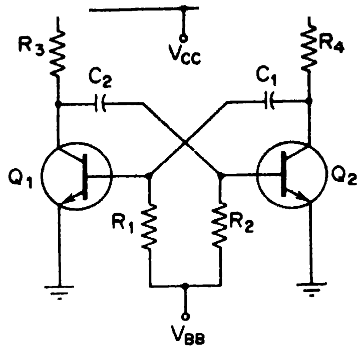
The *output impedance* of the pulse generator is another important consideration in fast pulse systems. This is so because the generator, which has a source impedance matched to the connecting cable, will absorb reflections resulting from impedance mismatches in the external circuitry. Without this generator-to-cable match, the reflections would be re-reflected by the generator, resulting in spurious pulses or perturbations on the main pulse.

DC coupling of the output circuit is necessary when retention of the dc bias levels in the test circuit is desired, in spite of variations in pulsewidth, pulse amplitude, or PRR.

(*c*) *Basic Circuits.* Circuits used in pulse generation generally fall into two categories: *passive,* or pulse-shaping, and *active,* or pulse-generating circuits. In passive-type circuits, a sine-wave oscillator is used as the basic generator and its output is passed through a pulse-shaping circuit to obtain the desired waveform. For instance, an approximate square waveform may be obtained by first amplifying and then clipping a sine wave.

Active generators are usually of the *relaxation* type. The relaxation oscillator uses the charge-and-discharge action of a capacitor to control the conduction of a vacuum tube or a transistor. Some common forms of relaxation oscillator are the *multivibrators* and the *blocking oscillators.*

The *astable* or *free-running multivibrator* is a widely used circuit for the generation of pulses. It can be made to produce either square waves or pulses, depending on the choice of circuit components. A typical free-running multivibrator is shown in Fig. 11-8(a). Essentially, the circuit consists of a two-stage RC-coupled amplifier, with the output of the second stage (Q_2) coupled back to the input of the first stage (Q_1) via a capacitor, C_1. Similarly, the output of



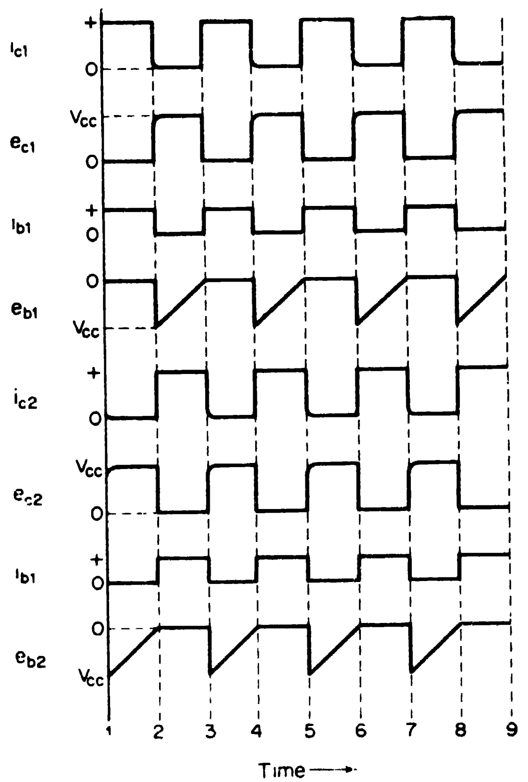


Figure 11-8
The astable multivibrator. (a) Circuit. (b) Waveforms.

Q_1 is coupled via C_2 to the input of Q_2. Since the coupling between the two transistors is taken from the collectors, this circuit is also known as a *collector-coupled* astable multivibrator.

The usual qualitative analysis of the circuit proceeds as follows: When the power is first applied to the circuit, both transistors start conducting. Owing to small differences in their operating characteristics, one of the transistors will conduct slightly more than the other. This fact starts a series of events. Assume that Q_1 initially conducts more than Q_2. This means that the collector voltage of $Q_1(e_{c1})$ drops more rapidly than the collector voltage of $Q_2(e_{c2})$. The decrease in e_{c1} is applied to the R_2C_2 network, and because the charge on C_2 cannot change instantaneously, the full negative-going change appears across R_2. This decreases the forward bias on Q_2 which in turn decreases the collector current of $Q_2(i_{c2})$, and the collector voltage of Q_2 rises. This rise in Q_2 collector voltage is applied via the R_1C_1 network to the base of Q_1, increasing its forward bias. Q_1 therefore conducts even more heavily and its collector voltage drops still more rapidly. This negative-going change is coupled to the base of Q_2, further decreasing its collector current. The entire process is cumulative until Q_2 is entirely cut off and Q_1 conducts heavily (bottoms).

With Q_2 cut off, its collector voltage practically equals the supply voltage, V_{CC}, and capacitor C_1 charges rapidly to V_{CC} through the low-resistance path from emitter to base of the conducting transitor Q_1. When the circuit action turns Q_1 fully on, its collector potential drops to approximately 0 V and since the charge on C_2 cannot change instantaneously, the base of Q_2 is at $-V_{CC}$ potential, driving Q_2 deep into cutoff.

The switching action now begins. C_2 begins to discharge exponentially through R_2. When the charge on C_2 reaches 0 V, C_2 attempts to charge up to the value of $+V_{BB}$. the base supply voltage. But this action immediately places a forward bias on Q_2 and this transistor starts to conduct. As soon as Q_2 starts conducting, its collector current causes a decrease in the collector voltage e_{c2}. This negative-going change is coupled to the base of Q_1 which starts to conduct less; i.e., it comes out of saturation. This cumulative action repeats until Q_1 finally cuts off and Q_2 conducts heavily. At this instant, the collector voltage of Q_1 reaches its maximum value of V_{CC}. Capacitor C_2 charges to the full value of V_{CC}, and a full cycle of operation has been completed.

The waveforms appearing at the base and the collector of each transistor are the result of a *symmetrical* or *balanced* operation: the time constants, R_1C_1 and R_2C_2, the transistors themselves, and the supply voltages are all identical. The conducting and nonconducting periods are therefore of almost the same duration. The waveforms for each of the two transistors are given in the waveform diagram of Fig. 11-8(b).

Assume that at time ($t = 1$) transistor Q_1 is fully turned on and transistor Q_2 is cut off. This makes the collector voltage e_{c1} of Q_1 a minimum (practically

0 V) and the collector voltage e_{c2} of Q_2 a maximum (V_{CC}). Capacitor C_1 is charging through the emitter-to-base resistance of Q_1 toward the supply voltage V_{CC} and reaches its full charge rapidly (low emitter-to-base resistance). Since e_{c1} is 0 V, capacitor C_2 begins to charge exponentially through R_2 toward the base supply voltage V_{BB} with a time constant equal to R_2C_2. Since the early part of the exponential charging curve is almost linear, the increase in Q_2 base voltage (e_{b2}) is indicated by a linear slope on the graph of Fig. 11-8(b).

At time ($t = 2$) e_{b2} reaches a value of approximately 0 V, placing a forward bias on the base of Q_2 which then starts to conduct. Within a very short time, the collector current of Q_2 reaches its maximum and collector voltage e_{c2} drops to 0 V. When Q_2 starts to draw current, the base of Q_1 becomes negative and Q_1 is quickly driven into cutoff. Its collector voltage, e_{c1} reaches the V_{CC} value and the collector current i_{c1} becomes zero. Within a very small fraction of the total Q_2 conducting time, capacitor C_2 is fully charged to V_{CC} through the low-resistance emitter-to-base path of Q_2.

Between times ($t = 2$) and ($t = 3$) transistor Q_1 is cut off, and its collector current and voltage remain constant. Similarly, the collector voltage and current for Q_2 remain constant. Only capacitor C_1 is charging and the base voltage e_{b1} of Q_1 is rising exponentially toward V_{BB}. At time ($t = 3$), the base voltage of Q_1 exceeds the cutoff value (approximately 0 V) and Q_1 starts conducting again. Obviously, one complete cycle of operation, from time ($t = 1$) to time ($t = 3$), depends on the time required for the base voltage of the cutoff transistor to reach the forward bias value. This time depends on two things: the magnitude of the reversed bias ($-V_{CC}$) and the time constant of the capacitor charging circuit involved, namely, R_1C_1 or R_2C_2.

The *analytical* evaluation of circuit operation proceeds as follows: During its nonconducting period, the collector voltage of Q_1 equals

$$e_{c1} = V_{CC}(1 - e^{-t/\tau_3}) \tag{11-13}$$

where $\tau_3 = R_3C_2$.

When Q_1 switches on, its collector voltage is at ground potential and the base voltage of Q_2 becomes $-V_{CC}$ with respect to ground The subsequent rise in Q_2 base voltage, through the R_2C_2 charging circuit, is described by

$$e_{b2} = (V_{BB} + V_{CC})(1 - e^{-t/\tau_2}) - V_{CC} \tag{11-14}$$

where $\tau_2 = R_2C_2$.

Q_2 remains cut off until e_{b2} reaches the value of 0 V (by good approximation) and the off time-interval T_2 of Q_2 can be determined by setting e_{b2} in Eq. (11-14) to zero and solving for t, so that

$$0 = (V_{BB} + V_{CC})(1 - e^{-t/\tau_2}) - V_{CC} \tag{11-15}$$

and
$$T_2 \qquad \ln\left(\frac{V_{BB}}{V_{BB} + V_{CC}}\right) \tag{11-16}$$

Similarly, when Q_2 is off and Q_1 is bottomed, the collector voltage of Q_2 can be described by

$$e_{c2} = V_{CC}(1 - e^{-t/\tau_4}) \tag{11-17}$$

where $\tau_4 = R_4C_1$.

When now Q_2 is switched on, its collector voltage drops to 0 V and the base voltage of Q_1 is given by

$$e_{b1} = (V_{BB} + V_{CC})(1 - e^{-t/\tau_1}) \tag{11-18}$$

where $\tau_1 = R_1C_1$.

Solving for the off time-interval T_1 of Q_1 by setting e_{b1} in Eq. (11-18) to zero,

$$0 = (V_{BB} + V_{CC})(1 - e^{-t/\tau_1}) - V_{CC} \tag{11-19}$$

and
$$T_1 = \tau_1 \ln\left(\frac{V_{BB}}{V_{BB} + V_{CC}}\right) \tag{11-20}$$

The total period of oscillation is given by

$$T = T_1 + T_2 \tag{11-21}$$

In the case of *symmetrical* operation, when the time constants R_1C_1 and R_2C_2 are equal, the waveform is a symmetrical square wave. By making the time constant R_1C_1 larger than the time constant R_2C_2, the output waveform becomes a *pulse train* because the off time of Q_1 will be larger than the off time of Q_2.

The *blocking oscillator* is another form of relaxation oscillator often used as a pulse generator. Although the actual circuit configuration may vary, the blocking oscillator always consists of an amplifier stage with positive or regenerative feedback from the output to the input through a pulse transformer. The pulse transformer differs somewhat from the conventional transformer in its capacity to pass a broad band of frequencies with minimum attenuation and phase shift.

One form of blocking oscillator is given in Fig. 11-9(a), where the vacuum tube is used as an amplifier. The transformer windings are so phased that increasing plate current induces a positive potential on the grid and decreasing plate current induces a negative potential. When the circuit is first energized, the plate current increases from zero to some value. The increasing plate current induces a positive voltage on the grid which causes the plate current to increase even faster. As the grid is driven more positive, the process repeats, and the plate current rises rapidly. The plate voltage drops rapidly to some low value. As the tube approaches saturation, the plate current increases at a lower rate, and the induced voltage in the grid winding of the transformer becomes smaller.

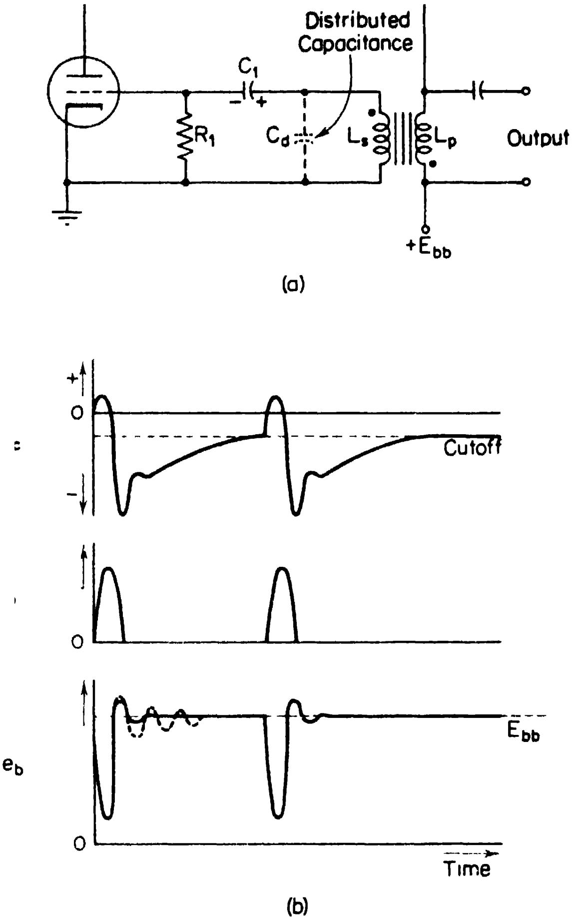


Figure 11-9
The blocking oscillator. (a) Circuit. (b) Waveforms.

In the meantime, grid current is drawn since the grid has been driven positive by the induced voltage in the transformer winding. Capacitor C_1 charges rapidly through the low-resistance cathode-to-grid path of the tube with the polarity shown and the rise in induced voltage is opposed by the rising capacitor voltage, thereby stabilizing the grid bias. When the plate current no longer increases, the induced voltage in the grid winding becomes zero, and capacitor C_1 begins to discharge through the grid resistor R_1. This makes the grid negative, which in turn causes the plate current to decrease from its saturated value. The decreasing plate current now induces a negative voltage in the grid winding which aids in the discharge of capacitor C_1. The discharge takes place at a faster rate and the tube is quickly driven to cutoff. Capacitor C_1 continues its discharge. The discharge current decreases exponentially with a time constant R_1C_1. The negative grid bias decreases until the tube finally comes out of cutoff and begins to conduct again. The nonconducting period clearly depends on the time constant R_1C_1 and is in general very much longer than the conducting period. The waveforms for the circuit are given in Fig. 11-9(b).

The *overshoot* in the plate voltage waveform and the variation in the grid voltage waveform, indicated by the dashed portions of the curves, are caused by the action of the inductance and the distributed capacitance of the transformer windings. At some (high) frequency, this parallel combination of L and C is resonant. When the current in the winding is suddenly interrupted, the LC circuit is excited into oscillation at its resonant frequency. As the voltage across the tank circuit varies sinusoidally, e_c and e_b follow the dotted waveforms. In practice, the oscillations are damped out by deliberately using a transformer with low Q windings, by the loading of the grid resistor R_1, and by the load connected to the output terminals. One important aspect of the self-resonant frequency of the transformer winding is determination of the pulsewidth by the frequency of oscillation. The transformer design is the determining factor for pulsewidth and if very narrow pulses are desired, transformer inductance and distributed capacitance must be kept to a minimum.

(*d*) *A Laboratory-Type Square-Wave and Pulse Generator.** The following description applies to a compact, general-purpose instrument providing negative pulses of variable frequency, duty cycle, and amplitude. The frequency range of the instrument is covered in seven decade steps from 1 Hz to 10 MHz, with a linearly calibrated dial for continuous adjustment on all ranges. The duty cycle can be varied from 25 per cent to 75 per cent. Two independent outputs are available: a 50-Ω source which supplies pulses with rise and fall times of 5 ns at 5-V peak amplitude, and a 600-Ω source which supplies pulses with rise and fall times of 70 ns at 30-V peak amplitude. The instrument can be operated as a free-running generator or it can be synchronized with external signals. Trigger output pulses for synchronization of external circuits or instruments are also available.

The block diagram of this generator is given in Fig. 11-10(a). The basic generating loop, which is redrawn for greater clarity in Fig. 11-10(b), consists of two current sources, the ramp capacitor, the Schmitt trigger circuit, and the current-switching circuit (here indicated by a simple switch). The two current sources provide a constant current for charging and discharging the ramp capacitor. The ratio of these two currents is determined by the setting of the *symmetry* control which then determines the duty cycle of the output waveform. The *frequency* dial controls the sum of the two currents from the current sources by applying appropriate control voltages to the bases of the current control

*Hewlett-Packard Model 211 B Square-Wave Generator.

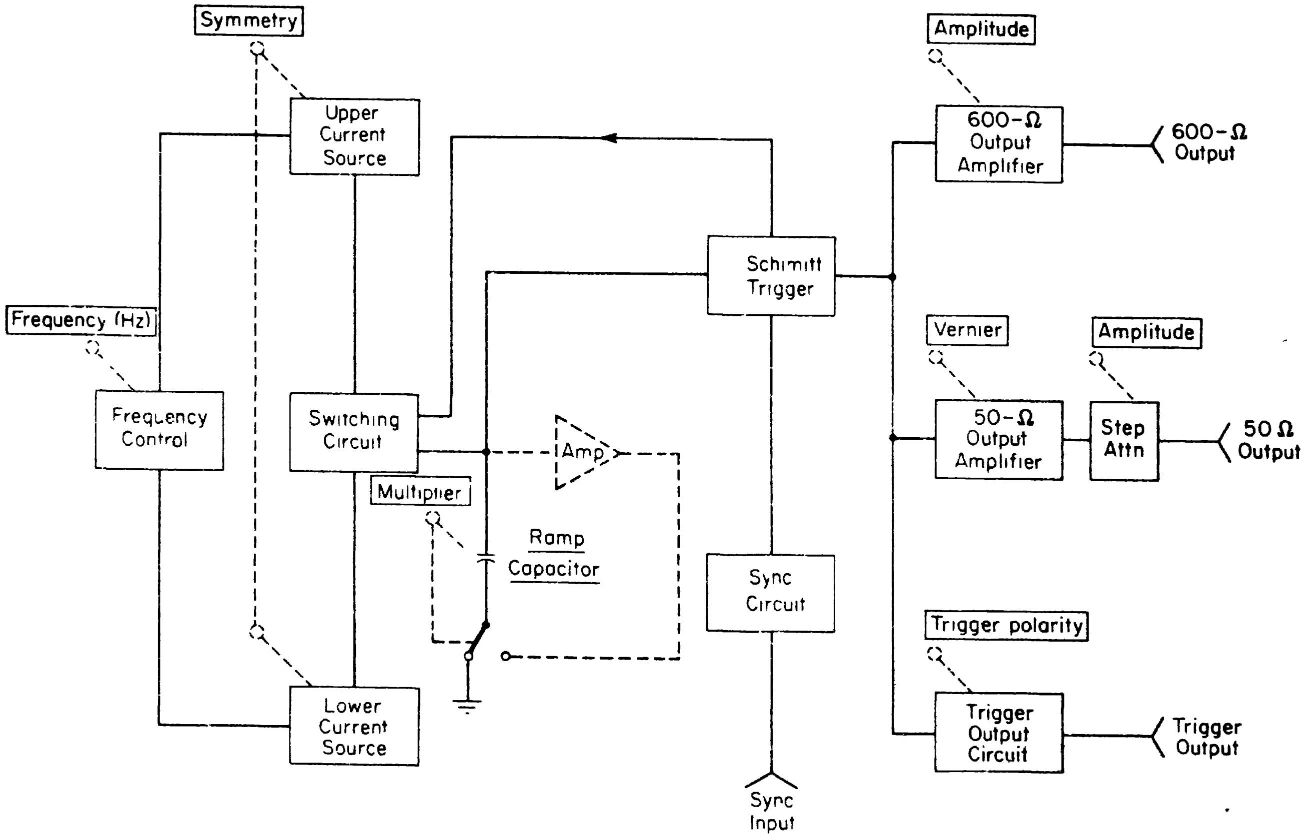


Figure 11-10 (a). Basic block diagram of a pulse generator. (*Courtesy Hewlett-Packard Co.*)

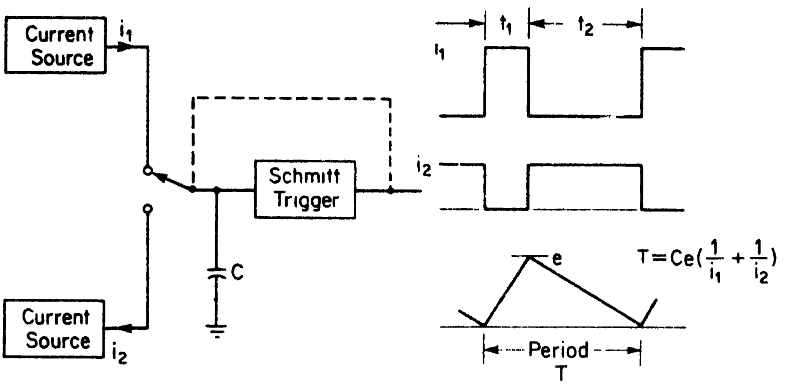


Figure 11-10 (b)
Simplified current source operation. (*Courtesy Hewlett-Packard Co.*)

transistors in the current generators. The size of the ramp capacitor is selected by the *multiplier* switch. These last two controls provide decade switching and vernier control of the frequency of the output.

The *upper* current source, supplying a constant current to the ramp capacitor, charges this capacitor at a constant rate and the ramp voltage increases linearly. When the positive slope of the ramp voltage reaches the upper limit, set by internal circuit components, the Schmitt trigger (a bistable multivibrator) changes state. The trigger circuit output goes negative, reversing the condition of the current control switch, and the capacitor starts discharging. The discharge rate is linear, controlled by the *lower* current source. When the negative ramp reaches a predetermined lower level, the Schmitt trigger switches back to its original state. This now provides a positive trigger circuit output which reverses the condition of the current switch again, cutting off the lower current source and switching on the upper current source. One cycle of operation has now been completed. The entire process, of course, is repetitive and the Schmitt trigger circuit provides negative pulses at a continuous rate.

The output of the Schmitt circuit is passed to the trigger output circuit and to the 50-Ω and 600-Ω amplifiers. The trigger output circuit differentiates the square-wave output from the Schmitt trigger, inverts the resulting pulse and provides a positive triggering pulse. The 50-Ω amplifier is provided with an output attenuator to allow a vernier control of the signal output voltage. In addition to its free-running mode of operation, the generator can be synchronized or locked in to an external signal. This is accomplished by triggering the Schmitt circuit by an external synchronization pulse.

The unit is powered by its own internal supply, which provides regulated voltages for all stages of the instrument.

11-3 Signal Generators

(*a*) *The Standard Signal Generator.** A standard signal generator is generally used as a power source in the measurement of gain, bandwidth, signal-to-noise (S/N) ratio, standing-wave ratio (SWR), and other circuit properties. It is extensively used in the testing of radio receivers and transmitters.

The standard signal generator is a source of ac energy of accurately known characteristics. The instrument is capable of *modulating* a carrier (or center) frequency which is indicated by a dial setting. The modulation is indicated by a meter reading set by appropriate control knobs. Common types of modulating signals are sine wave, square wave, and pulse; the output signal may be either amplitude modulated (AM) or frequency modulated (FM). AM is a common feature of the standard signal generator. When the FM system produces a considerable excursion in frequency at a relatively low cyclical rate, the instrument is known as a sweep-frequency generator (Sec. 11-3b). The output voltage is read by an output meter and an output attenuator setting.

The elements of a conventional standard signal generator are given in Fig. 11-11. The carrier frequency is generated by a very stable *LC* oscillator, deliv-

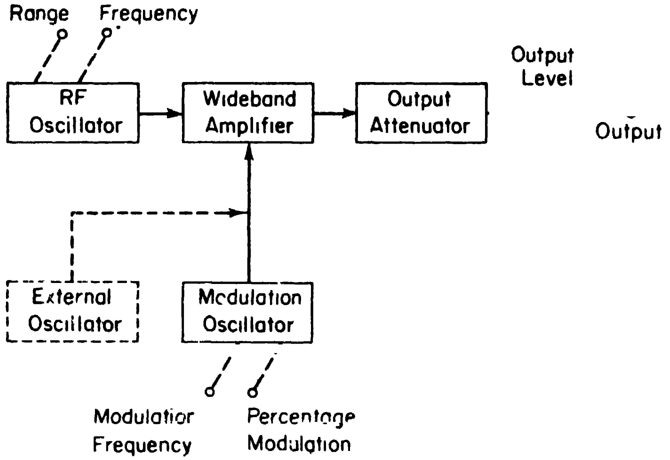


Figure 11-11
The basic elements of a standard signal generator.

ering a good sinusoidal waveform and having no appreciable hum or noise modulation. The frequency of oscillation is selected by a frequency range control and a vernier dial setting. The *LC* circuit is designed to give a reasonably constant output over any one frequency range. AM is provided from an internal, fixed-frequency sine-wave generator or from an external source. Modulation takes place in the output amplifier circuit, which delivers the modulated carrier to the output attenuator.

The frequency stability of these basic instruments is limited by the *LC* circuit design of the master oscillator. Since range switching is usually accomplished by selecting appropriate capacitive elements in the oscillator circuit, any change in frequency range upsets the circuit to some extent, and the user must wait until the circuit has stabilized at its new resonant frequency.

Some modern laboratory-type signal generators use a different approacn to frequency generation in order to improve the frequency stability. One such instrument is shown in the photograph of Fig. 11-12(b); the elementary diagram is given in Fig. 11-12(a). A single master oscillator is optimally designed for the highest frequency range, and frequency dividers are switched in to produce the lower ranges. In this manner, the stability of the top range is imparted to all other ranges. The master oscillator is made insensitive to temperature variations and also to the influence of the following stages by careful circuit design.

The RF oscillator output, after passing through an untuned buffer amplifier (B_1), enters the power amplifier unit. On the highest frequency range (34 MHz–80 MHz), the RF signal passes through an additional buffer (B_2) to the main amplifier (A). For all the lower frequency ranges, the oscillator signal is applied to a series of frequency dividers and from there through another buffer (B_3) to the power amplifier. The nine 2/1 dividers give a maximum divisor of 512. Therefore, the lowest frequency range produced by the cascaded divider chain equals the highest range divided by 512, or 67 kHz–156 kHz. Use of the buffer amplifiers provides a very high degree of isolation between the master oscillator and the power amplifier and almost eliminates all frequency-pulling effects from changes in operating and loading conditions at the output stage. 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 on the front panel is pushed which sends the indicator gliding along along the slide-rule scale of the main frequency dial at approximately 7 per cent frequency change per second. When the correct location on the main dial is reached, the oscillator can be fine-tuned by means of a large rotary control, with each division corresponding to 0.01 per cent of the main dial setting. A second front panel control ($\Delta F/F$) permits incremental tuning over a limited range and allows very great tuning resolution. All the aforementioned controls are clearly shown on the photograph of Fig. 11-12(b).

The availability of a motor-driven frequency control presents obvious opportunities for both local and remote automatic tuning, and these are

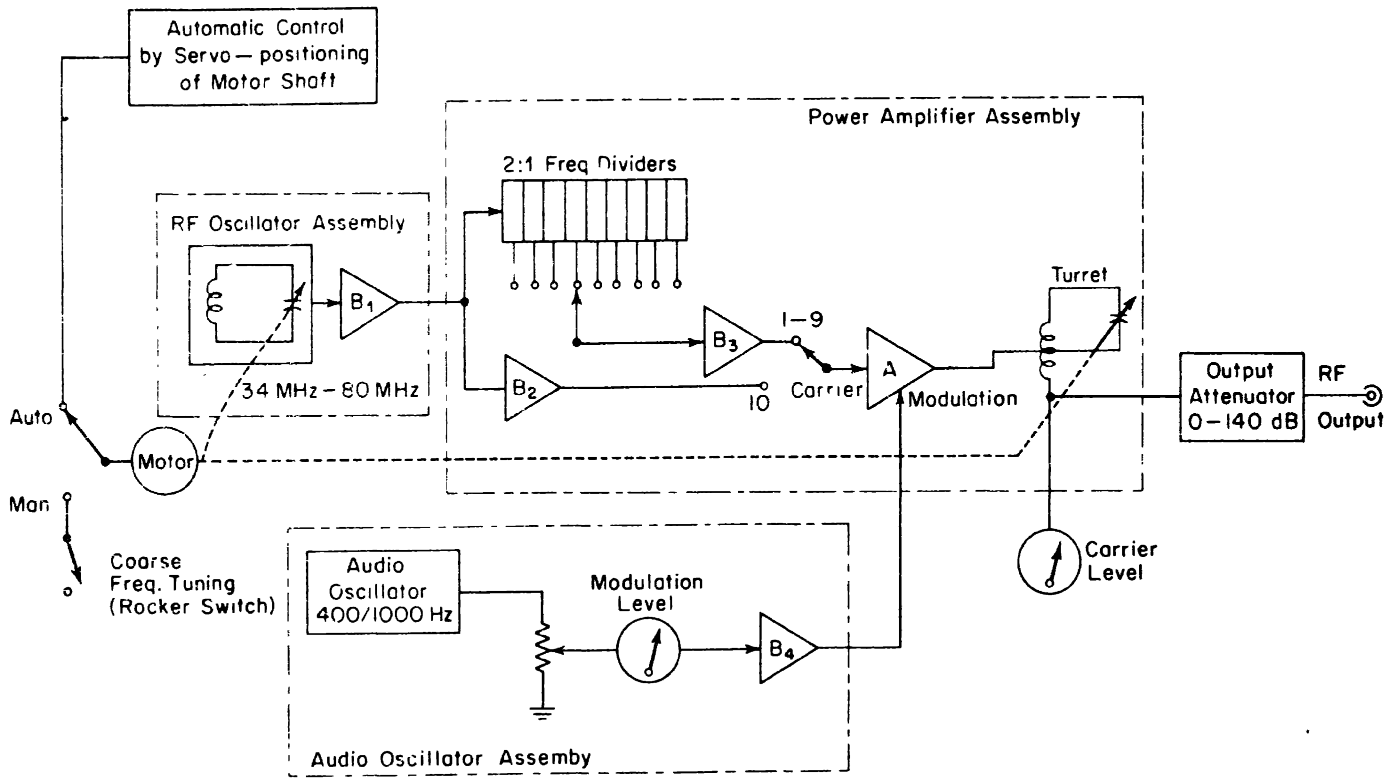


Figure 11-12 (a). Functional block diagram. (*Courtesy of General Radio Co.*)

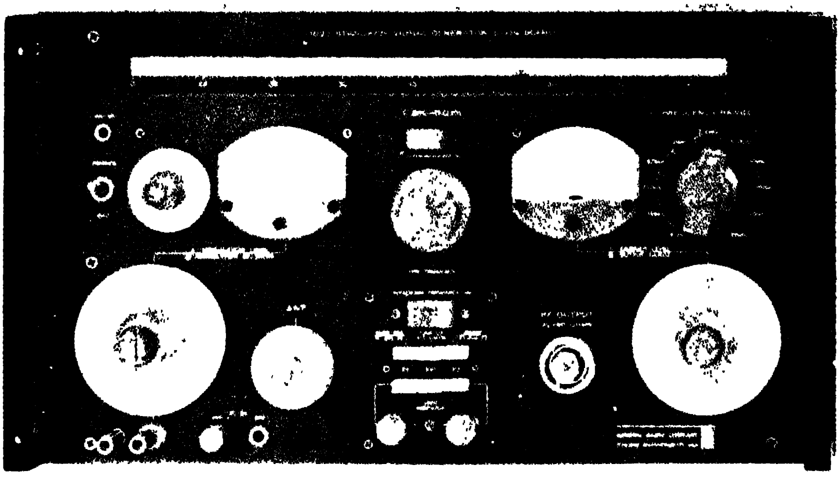


Figure 11-12 (b)
Standard signal generator. (*Courtesy General Radio Co.*)

exploited by a programmable automatic frequency-control device. With this instrument, one can sweep between adjustable frequency limits and automatically tune to preset frequencies.

The basic modulating function is performed in the power amplifier stage by variation of the base voltage of the power transistor. Two highly stable modulating signals (400 Hz and 1 kHz) are internally generated. The amplitude of the modulating signal can be adjusted by the modulation level control for up to 95 per cent modulation. Any given modulation setting is kept constant over the range of the carrier level control. Provision for application of external modulating signals is also made.

Internal calibration is provided by a 1-MHz crystal oscillator. This reference source is mixed with the master oscillator RF signal and produces a zero beat when the two signals are equal. An external reference signal can be applied to the mixer input and by using a portion of the crystal oscillator calibration circuitry, the instrument functions as a heterodyne frequency meter.

Additional refinements include the autocontrol unit which permits a number of automatic tuning operations. Power supply requirements are rather critical, but the small power consumption of the instrument makes it relatively easy to obtain excellent regulation and stability with very low ripple. The supply voltage for the master oscillator, which is especially sensitive to power variations, is regulated by a temperature-compensated reference circuit.

(*b*) *The Sweep-Frequency Generator.* The sweep-frequency generator is a logical extension of the standard signal generator. It provides a sinusoidal output voltage, usually in the RF range, whose frequency is smoothly and continuously varied over an entire frequency band, usually at a low audio rate. The process of frequency modulation (FM) may be accomplished electronically or mechanically. The mechanical method of varying the frequency of the master oscillator follows rather obviously from the discussion of the standard signal generator of Sec. 11-1(a) and may be realized by using a motor-driven variable capacitor in the *LC* oscillator circuit. This method is used to advantage in some laboratory instruments and brings the precision and stability of the conventional manually tuned signal generator to the field of swept-frequency measurements. In microwave measurements, a motor-driven, mechanically tuned klystron may be used to produce a swept RF signal, although more recent developments have led to the construction of electronically tuned oscillators. One such development is the *backward-wave oscillator tube*,* which overcomes the disadvantages of long sweep times and mechanical wear involved with motorized tuning devices.

The elements of a standard sweep-frequency generator are given in Fig. 11-13. The heart of the instrument is the master oscillator circuit, whose frequency range can be selected by the range switch. This provides a number of frequency bands, where each band usually covers several octaves. For simplicity, the frequency sweeper is assumed to be a mechanical device, which rotates the tuning capacitor in the *LC* master oscillator, causing repetitive sweeps over the complete frequency range. A representative sweep rate would be in the order of 20 sweeps per second. A manual frequency control allows independent manual adjustment of the master oscillator resonant frequency. The frequency sweeper also provides a synchronously varying sweep voltage, which can be used to drive the horizontal deflection plates of a CRO or the X axis of an X-Y recorder. Thus, the amplitude response of a device, fed by the swept-frequency output of the generator, can be displayed automatically on an oscilloscope or X-Y recorder.

To identify frequencies and bands of special interest, a *marker generator* provides half-sinusoidal waveforms at any frequency within the sweep range. The marker voltage can be added to the base line of the CRO trace during alternate cycles of the sweep voltage and appears as a superimposed identification mark on the response curve of the device under test.

The automatic level control circuit is basically a closed-loop feedback system which monitors the RF level at some point in the measurement system. This circuit holds the *forward power* (power delivered to the load) constant with frequency and load impedance variations. A constant power level ideally pre-

*Hewlett-Packard, Palo Alto, Calif. ("How a Helix Backward-Wave Tube Works"), *Application Note* No. AN 12.

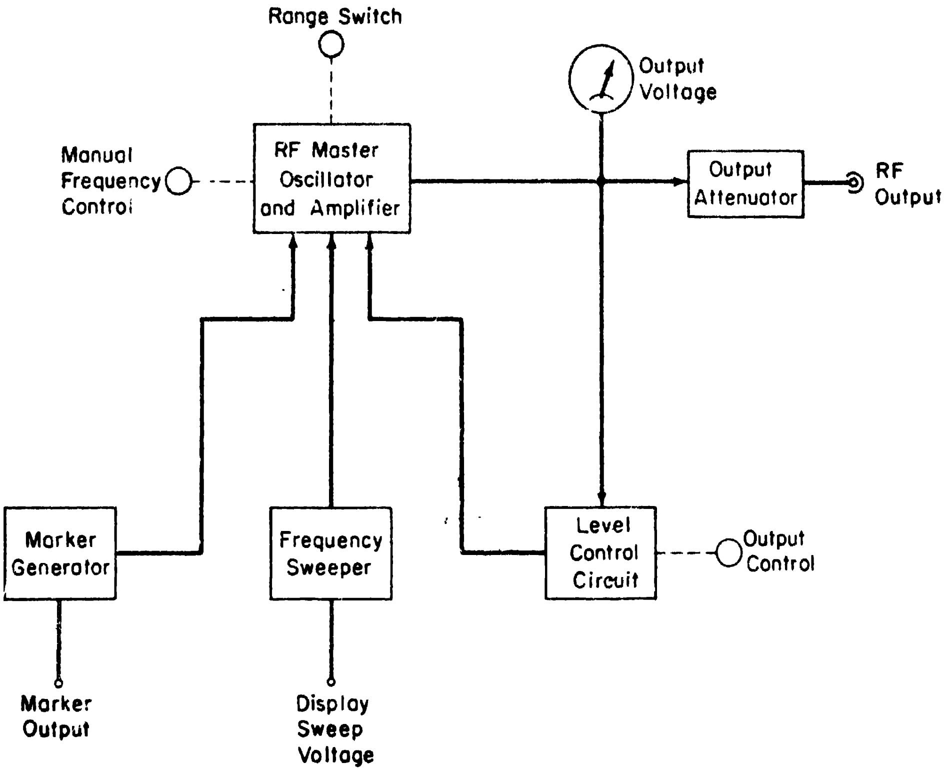


Figure 11-13
Elements of a standard sweep-frequency generator.

vents any source mismatch and also provides a constant read-out calibration with frequency.

(*c*) *The Random Noise Generator.* The random noise generator is a device which delivers a signal whose instantaneous amplitude is determined at random and is therefore unpredictable. True random noise contains no *periodic* frequency components and has a *continuous* spectrum. This instrument offers the possibility of using a single measurement as an indicator of performance over a wide frequency band.

Random noise measurements are applied in many measurement fields. In accoustical measurements, random noise is used to smooth response curves which might otherwise be difficult to interpret. Random noise in psychoaccoustical measurements has greatly increased the knowledge of the process of hearing. Random noise also best simulates the vibrations that aircraft and rockets are subjected to in flight, and it is commonly used in vibration and fatigue testing of aerospace components and assemblies.

In electrical measurements, noise can be used as the test signal itself. Intermodulation (IM) distortion and crosstalk measurements in communication systems, tests on servo amplifiers, and studies made with analog computers are but a few of the many applications. Noise of known amplitude and known

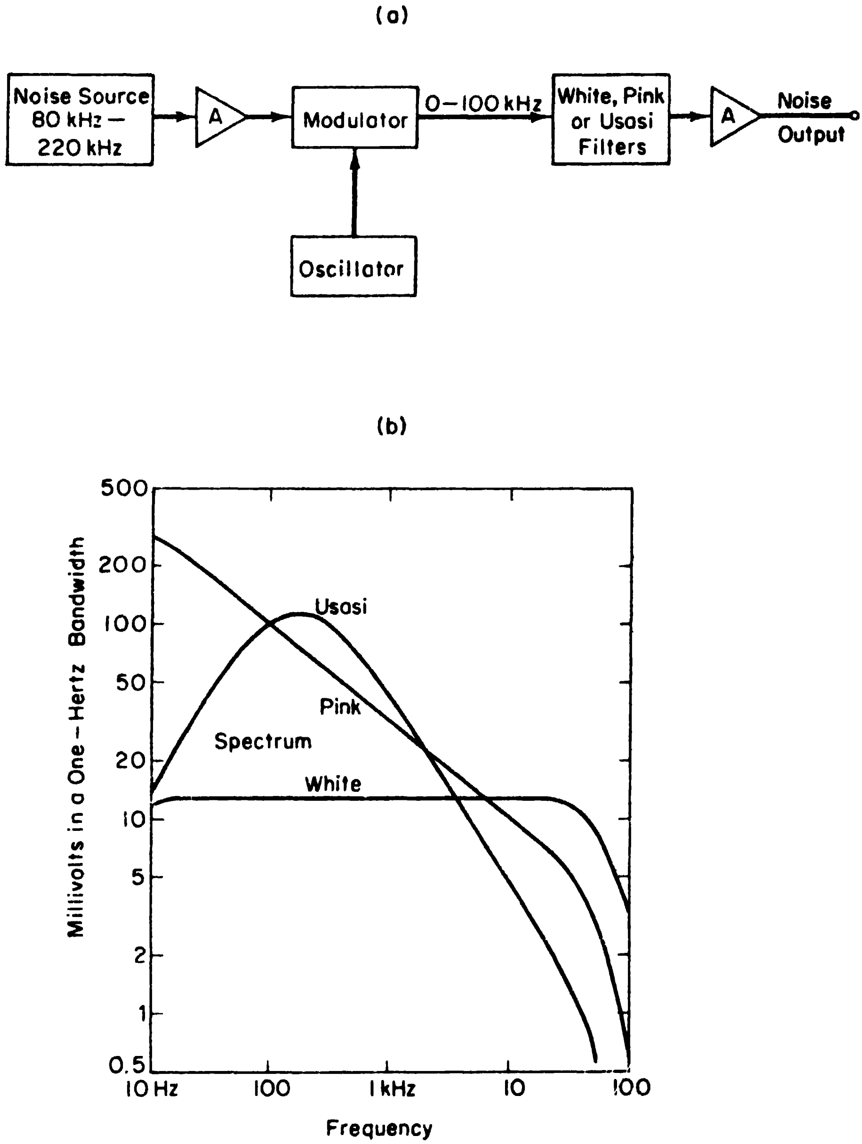


Figure 11-14

(a) Block diagram of a random noise generator. (b) Three different output spectra. (*Courtesy General Radio Co.*)

spectral characteristics is most effective for testing various methods of signal detection and recovery in the presence of noise, as in radio, telemetry, radar, and sonar systems.

Random noise generators* are available to cover frequencies from near dc to microwave. The method of generating random noise is usually by a semiconductor noise diode, which delivers frequencies in a band roughly extending from 80 kHz to 220 kHz. The simplified block diagram of a random noise generator for use in the audiofrequency range is shown in Fig. 11-14. The output from the noise diode is amplified and heterodyned down to the audiofrequency band in a balanced, symmetrical modulator delivering a symmetrical amplitude distribution.

The output amplifier, as the final stage in the generator, includes a transformer and supplies a floating, single-ended, or balanced output. The filter arrangement which follows the modulator further reduces and controls the bandwidth and supplies an output signal in three spectrum choices: *white* noise, *pink* noise, and *USASI* noise. This is illustrated in Fig. 11-14(b).

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 dB per octave.

Pink noise is so called because of its emphasis (greater amplitude) on the lower frequencies, as in reddish light. Pink noise has a voltage spectrum which is inversely proportional to the square root of the frequency and is used in bandwidth analysis.

USASI noise roughly simulates the energy distribution of speech and music frequencies and is used for testing audio amplifiers and loudspeakers.

11-4 Function Generators

A function generator is a versatile instrument which delivers a choice of different waveforms whose frequencies are adjustable over a wide range. The most common output waveforms are the sine, triangular, square, and sawtooth waves. The frequencies of these waveforms may be adjusted from a fraction of a hertz (0.00005 Hz is listed in the specification sheet of a representative function generator as the lower limit) to several hundred kilohertz.

The various outputs of the generator may be available at the same time. For instance, by providing a square wave for linearity measurements in an audio system, a simultaneous sawtooth output may be used to drive the horizontal deflection amplifier of a CRO, providing a visual display of the measurement results. The capability of the function generator to *phase lock* to an external signal source is another useful feature. One function generator may be used to phase lock a second function generator and the two output signals can be displaced in phase by an adjustable amount. In addition, one generator may be phase locked to a harmonic of the sine wave of another generator. By adjusting the phase and the amplitude of the harmonics, almost any waveform may be generated by the summation of the fundamental frequency generated by the one function generator and the harmonic generated by the other function generator.

The function generator can also be phase locked to a frequency standard, and all its output waveforms are then generated with the frequency accuracy and stability of the standard source. The function generator finds extensive application in many areas. Besides its use in the audio measurements field, it is also used in medical electronics for nerve stimulation and electroanesthesia.

The function generator can supply output waveforms at very low frequencies. Since the low frequency of a simple RC oscillator is limited, a different approach is used in the function generator of Fig. 11-15. This instrument delivers sine, triangular, and square waves with a frequency range of 0.01 Hz to 100 kHz. The frequency control network is governed by the frequency dial on the front panel of the instrument or by an externally applied control voltage. The frequency control voltage regulates two current sources.

The *upper* current source supplies a constant current to the triangle integrator whose output voltage increases linearly with time. The output voltage is given by the well-known relationship

$$e_{\text{out}} = -\frac{1}{C}\int i\,dt \tag{11-22}$$

An increase or a decrease in the current supplied by the upper current source increases or decreases the slope of the output voltage.

The voltage comparator multivibrator changes state at a predetermined level on the positive slope of the integrator's output voltage. This change of state cuts off the upper current supply to the integrator 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 voltage reaches a predetermined level on the negative slope of the output waveform, the voltage comparator again switches and cuts off the lower current source while at the same time switching on the upper source again.

The voltage at the output of the integrator has a triangular waveform whose frequency is determined by the magnitude of the current supplied by the constant current sources. The comparator delivers a square-wave output voltage of the same frequency. The third output waveform is derived from the triangular waveform, which is synthesized into a sine wave by a diode resist-

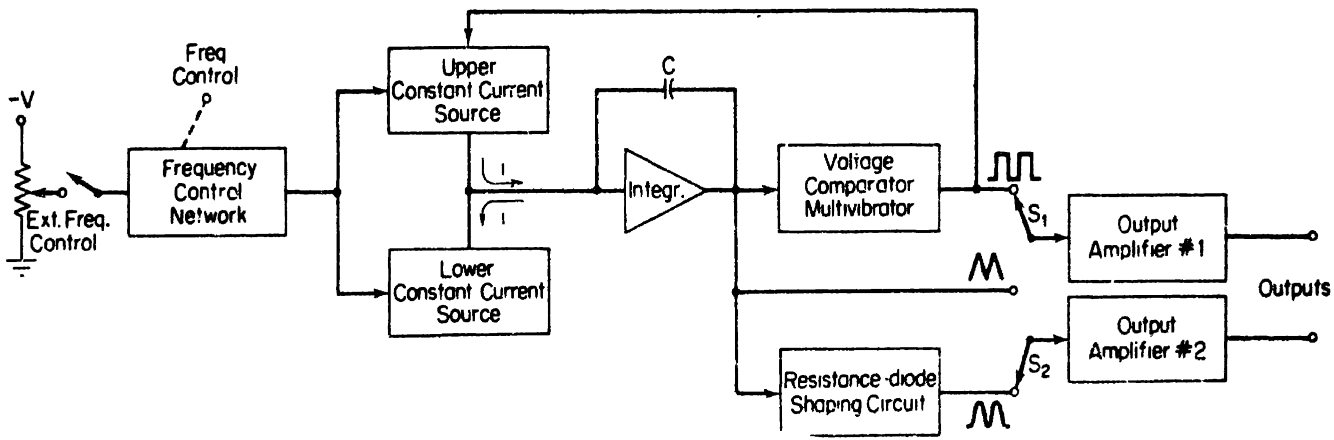


Figure 11-15. The basic elements of a function generator

ance network. In this circuit, the slope of the triangular wave is altered as its amplitude changes, resulting in a sine wave with less than 1 per cent distortion.

The output circuitry of the function generator consists of two output amplifiers which provide two simultaneous, individually selected outputs of any of the waveform functions.

11-5 Wave Analyzers

A *wave analyzer* is an instrument designed to measure the relative amplitudes of single-frequency components in a complex or distorted waveform. Basically, the instrument acts as a *frequency selective voltmeter,* which is tuned to the frequency of one signal component while rejecting all the other signal components. Two basic circuit configurations are generally used. For measurements in the audiofrequency range (from 20 Hz to 20 kHz), the analyzer has a filter section with a very *narrow passband* which can be tuned to the frequency component of interest. An instrument of this type is given in the form of a functional block diagram in Fig. 11-16(a); Fig. 11-16(b) shows a typical attenuation curve of the filter section used.

The waveform to be analyzed in terms of its separate frequency components is applied to an input attenuator which is set by the *meter range* switch on the front panel. A driver amplifier feeds the attenuated waveform to a high-Q, active filter. The filter consists of a cascaded arrangement of RC resonant sections and filter amplifiers. The passband of the total filter section is covered in decade steps over the entire audio range by switching capacitors in the RC sections. Close-tolerance polystyrene capacitors are generally used for selecting the frequency ranges. Precision potentiometers are used to tune the filter to any desired frequency within the selected passband.

A final amplifier stage supplies the selected signal to the meter circuit and to an untuned buffer amplifier. The buffer amplifier can be used to drive a recorder or an electronic counter. The meter is driven by an average-type detector and usually has several voltage ranges as well as a decibel scale.

The bandwidth of the instrument is very narrow, typically about 1 per cent of the selected frequency. Figure 11-16(b) shows a typical attenuation curve of a selected wave analyzer (General Radio Company, Type 1568-A Wave Analyzer). The initial attenuation rate is approximately 600 dB per octave; the attenuation at one-half and twice the selected frequency is about 75 dB. The filter characteristic also shows that the attenuation is still increasing far from the center frequency, well into the noise level of the instrument itself. The analyzer must have extremely low input distortion; so low, in fact, that it cannot be detected by the analyzer itself.

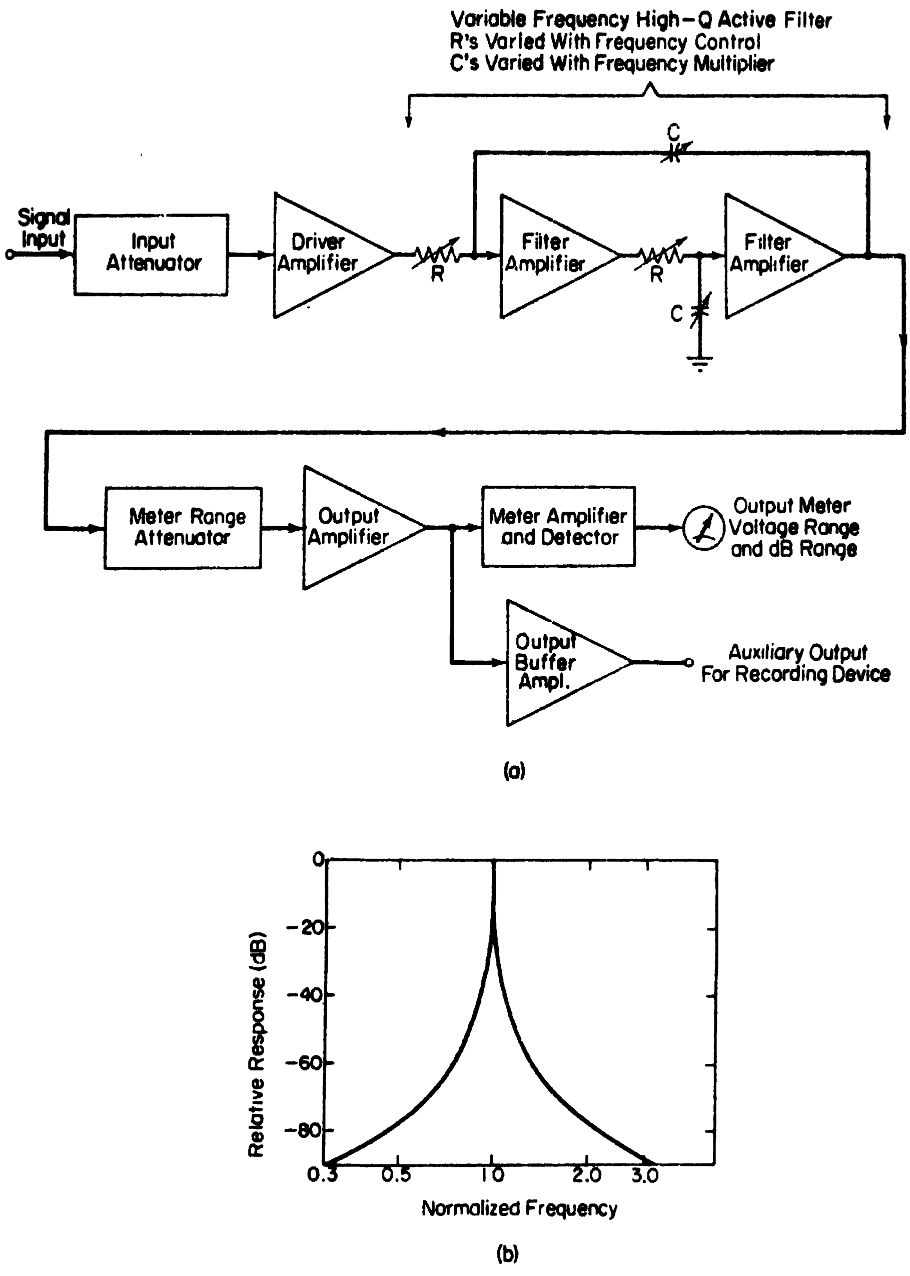


Figure 11-18

(a) Functional block diagram of an audio-range wave analyzer (*adapted from GR Type 1568-A*). (b) Attenuation characteristics of the active filter of the wave analyzer, showing the extremely sharp attenuation at the selected frequency. (*Courtesy General Radio Co.*)

Measurements in the megahertz range are usually done with another wave analyzer, particularly suited to the higher frequencies. The input signal to be analyzed is *heterodyned* to a higher intermediate frequency (IF) by an internal local oscillator. Tuning the local oscillator shifts the various signal frequency components into the passband of the IF amplifier. The output of the IF amplifier is rectified and applied to the metering circuit. An instrument which uses the heterodyning principle is often called a ***heterodyning tuned voltmeter.***

A typical wave analyzer using the heterodyning principle is shown in block diagram form in Fig. 11-17. The operating frequency range of this instrument

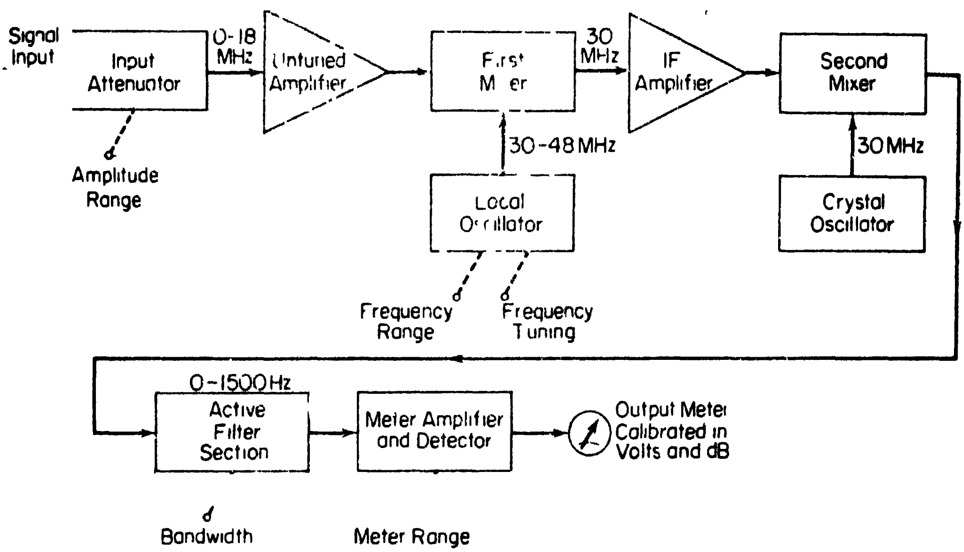


Figure 11-17
Functional block diagram of the heterodyning wave analyzer. (*Adapted from HP Model 312-A, Courtesy Hewlett-Packard Co.*)

is from 10 kHz to 18 MHz in eighteen overlapping bands, which are selected by the frequency range control of the local oscillator. The bandwidth is controlled by an active filter section and can be selected at 200, 1,000, and 3,000 Hz.

The input signal enters the instrument through a probe connector which contains a unity gain isolation amplifier. After appropriate attenuation, the input signal is heterodyned in the mixer stage with the signal from a local oscillator. The output of the mixer forms an intermediate frequency, which is uniformly amplified by the 30-MHz IF amplifier. This amplified IF signal is then mixed again with a 30-MHz crystal oscillator signal, which results in information centered on a zero frequency. An active filter with controlled bandwidth and symmetrical slopes of 72 dB per octave then passes the selected component to the meter amplifier and detector circuit. The output from the meter detector can be read off a decibel-calibrated scale or may be applied to a recording device.

Applications of the wave analyzer are found in the fields of electrical measurements, and sound and vibration analysis. For example, ***harmonic distortion*** of an amplifier can readily be measured, and the contribution of each harmonic to the total distortion figure can be determined. When the passband of the analyzer of Fig. 11-16(a) is tuned to the second harmonic, the fundamental frequency is sufficiently attenuated to reduce its level to substantially less than that of the harmonic. The curve of Fig. 11-16(b) shows that half-frequency attenuation is at least 75 dB. When the third harmonic is selected, the fundamental is attenuated by more than 85 dB. A complete harmonic analysis is then carried out by resolving the individual components of a periodic signal and measuring or displaying these components. It is not unusual to be able to separate and display about fifty harmonics.

The wave analyzer is applied industrially in the field of reduction of sound and vibration generated by machines and appliances. The source of the noise or vibration, generated by a machine must first be identified before it can be reduced or eliminated. A fine-spectrum analysis with the wave analyzer will show various discrete frequencies and resonances, which can be related to motions within the machine.

11-6 Harmonic Distortion Analyzers

In the ideal case, application of a sinusoidal input signal to an electronic device, such as an amplifier, should result in the generation of a sinusoidal output waveform. Generally, however, the output waveform is not an exact replica of the input waveform because of the various types of distortion that may arise. This distortion may be a result of the inherent nonlinear characteristics of the transistors or vacuum tubes in the circuit or of the circuit components themselves. Nonlinear behavior of circuit elements introduces harmonics of the fundamental frequency in the output waveform, and the resultant distortion is often referred to as *harmonic distortion* (HD).*

A measure of the distortion represented by a particular harmonic is simply the ratio of the harmonic to that of the fundamental frequency, expressed as a percentage. Harmonic distortion (HD) is then represented by

$$D_2 = \frac{B_2}{B_1}, \qquad D_3 = \frac{B_3}{B_1}, \qquad D_4 = \frac{B_4}{B_1} \tag{11-23}$$

*Millman (*op. cit*, pp. 405–408) treats harmonic generation in a vacuum tube.

where $D_n (n = 2, 3, 4, \ldots)$ represents the distortion of the nth harmonic and B_n represents the amplitude of the nth harmonic. (B_1 is the amplitude of the fundamental.)

The *total harmonic distortion* or *distortion factor* is defined as

$$D = \sqrt{D_2^2 + D_3^2 + D_4^2 + \cdots} \tag{11-24}$$

Several methods have been devised to measure the harmonic distortion, caused either by a single harmonic or by the sum of all the harmonics. Some of the better-known methods are described in the following subsections.

(*a*) *The Tuned-Circuit Harmonic Analyzer.* One of the oldest methods of determining the harmonic content of a waveform uses a tuned circuit (Fig. 11-18). A series-resonant circuit, consisting of inductor L and capacitor C,

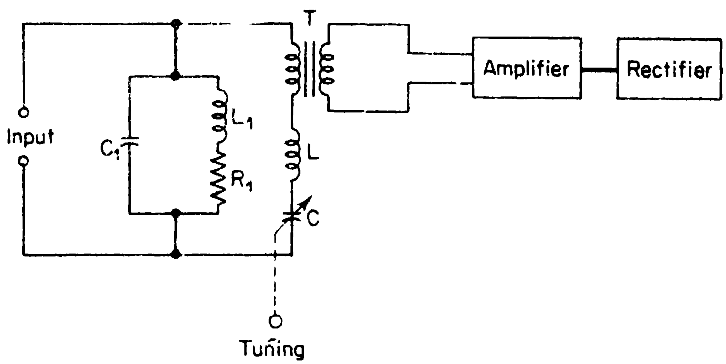


Figure 11-18
Functional block diagram of the tuned circuit harmonic analyzer.

is tuned to a specific harmonic frequency. This harmonic component is transformer coupled to the input of an amplifier. The output of the amplifier is rectified and applied to a meter circuit. After a reading is obtained on the meter, the resonant circuit is retuned to another harmonic frequency and the next reading is taken, and so on. The parallel-resonant circuit consisting of L_1, R_1, and C_1 provides compensation for the variation in the ac resistance of the series-resonant circuit and also for the variations in the amplifier gain over the frequency range of the instrument.

Numerous modifications of this basic circuit have been developed, but all tuned-circuit analyzers have *two major drawbacks*: (1) At low frequencies, very large values for L and C are required and their physical size becomes rather impractical: (2) Harmonics of the signal frequency are often very close in frequency so that it becomes extremely difficult to distinguish between them. Some circuit refinements can lessen this problem and the analyzer does find useful application where it is important to measure each harmonic component individually rather than take a single reading for the total harmonic distortion.

(*b*) *The Heterodyne Harmonic Analyzer or Wavemeter.* The difficulties of the tuned circuit are overcome in the heterodyne analyzer, by using a highly selective, fixed-frequency filter. The simplified block diagram of Fig. 11-19 shows the basic functional sections of the heterodyne harmonic analyzer.

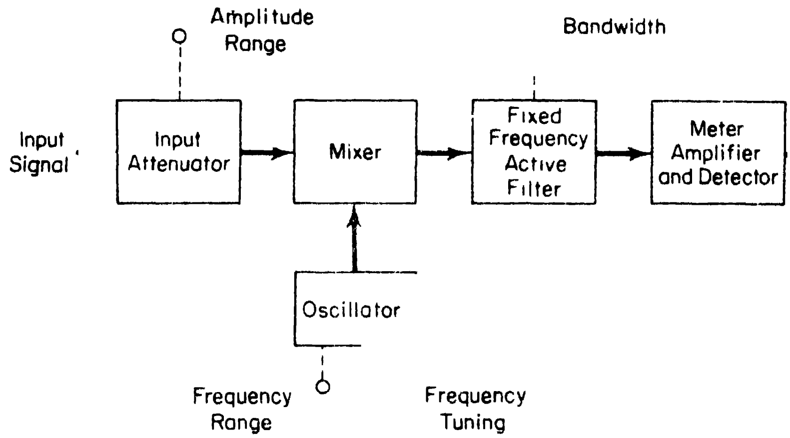


Figure 11-19
Block diagram of a heterodyne-type harmonic analyzer or wavemeter.

The output of a variable-frequency oscillator is *mixed* (heterodyned) successively with each harmonic of the input signal and either the sum or the difference frequency is made equal to the frequency of the filter. Since now each harmonic frequency is converted to a constant frequency, it is possible to use highly selective filters of the quartz-crystal type. With this technique, only the constant-frequency signal, corresponding to the particular harmonic being measured, is passed and delivered to a metering circuit. This metering circuit is usually a balanced-bridge VTVM. The mixer usually consists of a balanced modulator since it offers a simple means of eliminating the original frequency of the harmonic. The low harmonic distortion generated by the balanced modulator is another advantage over different types of mixers. Excellent selectivity is obtained by using quartz-crystal filters or inverse feedback filters.

On some heterodyne analyzers the meter reading is calibrated directly in terms of voltage; other analyzers compare the harmonics of the impressed signal with a reference voltage, usually by making the reference voltage equal to the amplitude of the fundamental. Direct reading instruments of the heterodyne type are sometimes known as *frequency-selective voltmeters.* In these instruments, the frequency of the input signal is read off a calibrated dial. A low-pass filter in the input circuit excludes the sum of the mixed frequencies and passes only the difference frequency. This voltage is compared to the input signal and read off on a calibrated voltmeter in dBm and volts. The level range for most of these meters is from -90 dBm to $+32$ dBm.

(*c*) *The Fundamental Suppression Harmonic Distortion Analyzer.* The fundamental suppression method of measuring distortion is used when it is important to measure total harmonic distortion (THD) rather than the distortion caused by each component. In this method, the input waveform is applied to a net-

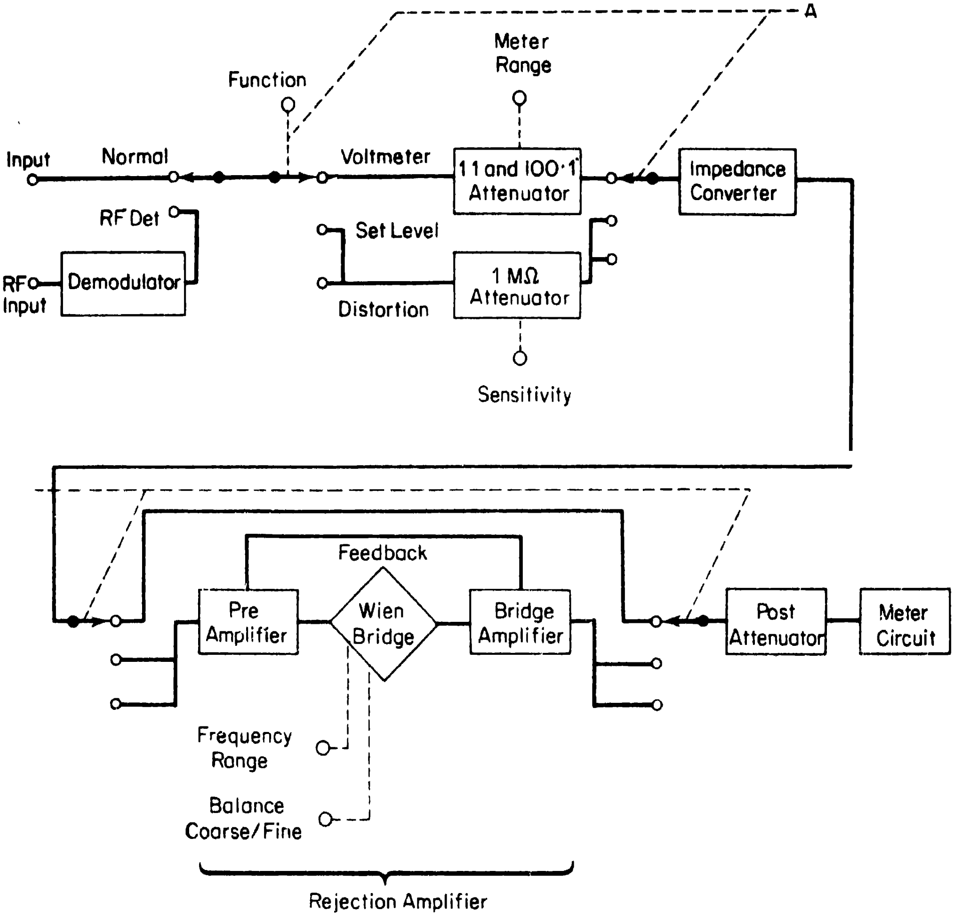


Figure 11-20
Block diagram of a fundamental suppression distortion analyzer. (*Courtesy Hewlett-Packard Co.*)

work which *suppresses* or *rejects* the *fundamental* frequency component but passes all the harmonic-frequency components for subsequent measurement. These instruments are simpler and less expensive than instruments of the heterodyne type and can be used to advantage in combination with other types of analyzers. The fundamental suppression type of instrument has two advantages: (1) The harmonic distortion generated within the instrument itself is very small and can be neglected. (2) The selectivity requirements are less because only the fundamental must be suppressed; all other frequencies may be passed.

The block diagram of the fundamental suppression HD analyzer is shown in Fig. 11-20. The instrument consists of four major sections: (1) the impedance converter circuit, (2) the rejection amplifier, (3) the metering circuit, (4) the power supply section. The impedance converter provides a low-noise, high-impedance input circuit, independent of the signal source impedance placed at the input terminals to the instrument. The rejection amplifier rejects the fundamental frequency of the input signal and passes the remaining frequency components on to the metering circuit where the HD is measured. The metering circuit provides a visual indication of total HD in terms of a percentage of total input voltage.

Two modes of operation are possible: When the function switch is in the *voltmeter* position, the instrument operates as a conventional ac voltmeter, a very convenient feature. With the function switch in the *distortion* postion, the rejection amplifier becomes part of the circuit and distortion measurements are made.

In the *distortion mode* of operation, the input signal is applied to a 1-MΩ input attenuator which provides 50-dB attenuation in 10-dB steps, controlled by a front panel switch marked *sensitivity*. When the desired attenuation is selected, the signal is fed to the impedance converter, which is a low-distortion, high-input impedance amplifier circuit with a gain independent of the source impedance placed at the input terminals. The over-all negative feedback in this amplifier results in unity gain and low distortion. Signals having a high source impedance can be measured accurately and the sensitivity selector can be used in the high-impedance positions without distorting the input signal.

The rejection amplifier circuit consists of a preamplifier, a Wien bridge circuit, and a bridge amplifier. The preamplifier receives the signal from the impedance converter and provides additional amplification at extremely low distortion levels. The Wien bridge circuit is used as a rejection filter for the fundamental frequency of the input signal. With the function switch in the "distortion" position, the Wien bridge is connected as an interstage coupling element between the preamplifier and the bridge amplifier. The bridge is first tuned to the fundamental frequency of the input signal by setting the *frequency range* selector (R_1 and R_2) and varying the tuning capacitors (C_1 and C_2). After tuning, the bridge is brought into balance by adjustment of the coarse and fine balance controls (R_3 and R_4). When the bridge is tuned and balanced, the voltage and phase of the fundamental, which appears at the junction of the series reactance and the shunt reactance, are the same as the voltage and phase at the midpoint of the resistive branch. When these two voltages are equal and in phase, no output signal will appear.

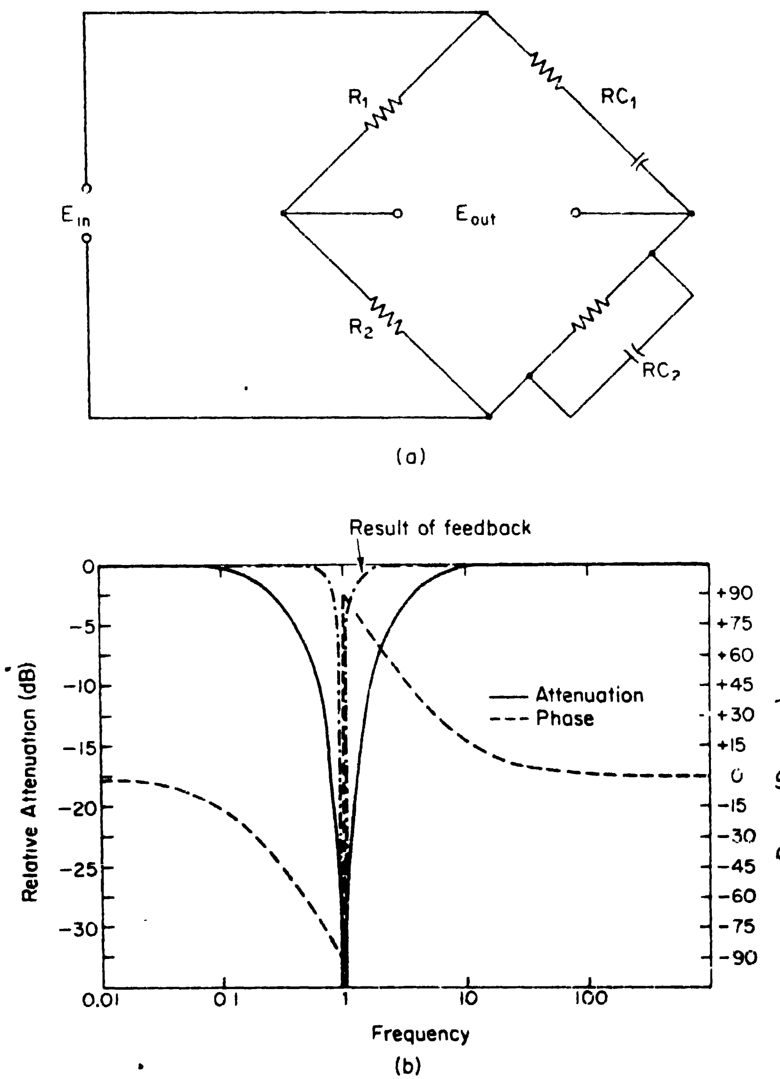


Figure 11-21
(a) Typical Wien bridge, used as a suppression circuit. (b) Rejection characteristics of the Wien bridge.

For frequencies other than the fundamental, the reactive branch of the Wien bridge offers varying degrees of phase shift and attenuation. The difference voltage between the reactive and resistive branches is amplified by the bridge-amplifier circuit. The output of the bridge amplifier is connected through a post-attenuator to the metering circuit.

Negative feedback is applied from the bridge amplifier output to the preamplifier input to narrow the frequency-rejection characteristic. Figure 11-21 shows that rejection of harmonic voltages is not constant. The second harmonic is attenuated several decibels more than the third harmonic, the third harmonic more than the fourth, and so on. Fortunately, however, the higher harmonics are far smaller in amplitude than the lower harmonics. The result of the negative feedback is illustrated by the rejection characteristic shown in dashed lines on the attenuation and phase curves of Fig. 11-21.

The meter circuit consists of a post-attenuator, an amplifier, and a rectifier circuit. The attenuator limits the signal level to the meter amplifier to 1 mV for full-scale deflection on all ranges. The meter amplifier itself is a multistage amplifier circuit, designed for low drift and low noise, and flat response characteristics. The meter is connected in a bridge-type rectifier and reads the average value of the signal impressed on the circuit. The meter scale is calibrated to the rms value of a sine wave.

The AM detector circuit allows measurement of envelope distortion in AM carriers. The input signal is applied to the demodulator, where the modulating signal is recovered from the RF carrier. The signal is then applied to the impedance converter through the 1-MΩ attenuator and is processed in the same manner as the normal distortion measurements.

In the voltmeter mode of operation, the input signal is applied to the impedance converter circuit through the 1/1 and 100/1 attenuator, which selects the appropriate meter range. The output of the impedance converter then bypasses the rejection amplifier and the signal is applied directly to the metering circuit. The voltmeter section can be used separately for general-purpose voltage and gain measurements.

11-7 Spectrum Analysis

(*a*) *Introduction. Spectrum analysis** is defined as the study of the ***energy distribution*** across the frequency spectrum of a given electrical signal. From this study comes valuable information about bandwidth, effects of various types of modulation, spurious signal generation, and so on—all of it useful in the design and testing of RF and pulse circuitry.

*Hewlett-Packard, Palo Alto, Calif., *Application Notes* AN 63, AN 63A. General principles and measurement applications are treated in a practical manner.

Because of instrumentation capabilities and limitations, spectrum analysis is usually divided into two major categories: (1) audio spectrum analysis, (2) RF spectrum analysis. The RF spectrum analysis, covering the frequency range of 10 MHz to 40 GHz, is the more important, since it includes the vast majority of the communications, industrial instrumentation, navigation, and radar bands.

Originally designed to view the spectrum of a burst of RF energy in radar applications, the spectrum analyzer has been developed into a sophisticated instrument capable of graphically presenting amplitude as a function of frequency in a portion of the RF spectrum. The instrument finds application as a tool for measuring attenuation, FM deviation, for frequency measurements and in pulse studies.

(*b*) *The Basic Spectrum Analyzer.* The spectrum analyzer is designed to represent, graphically, a plot of amplitude versus frequency of a selected portion of the frequency spectrum under investigation. One of the earliest instruments used during the development of radar pulse techniques, involved little more than a simple RF indicator, which lacked calibrated controls or broad spectrum coverage but sufficed for the relatively simple task at hand. The modern spectrum analyzer basically consists of a narrow-band *superheterodyne receiver* and CRO. The receiver is electronically tuned by varying the frequency of the local oscillator. The simplified block diagram of Fig. 11-22 shows the elements of a spectrum analyzer using the swept-frequency design.

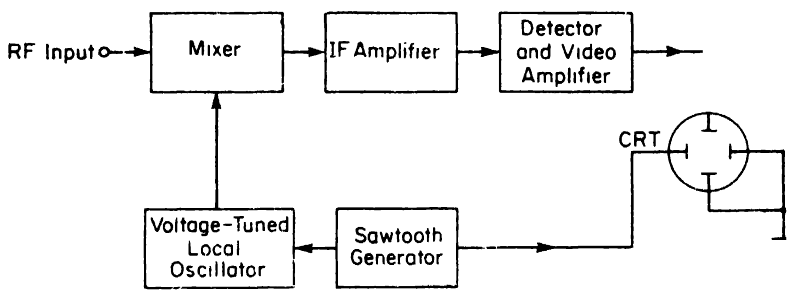


Figure 11-22
The basic elements of a spectrum analyzer of the swept-receiver design.

A sawtooth generator supplies a sawtooth voltage to the frequency control element of the voltage-tuned local oscillator, which then sweeps through its frequency band at a recurring linear rate. The same sawtooth voltage is simultaneously applied to the horizontal deflection plates of the CRT in the CRO. The RF signal under investigation is applied to the input of the mixer stage. As the local oscillator is swept through its frequency band by the sawtooth generator, it will beat with the input signal to produce the required intermediate frequency (IF). An IF component is produced only when the corresponding component is present in the RF input signal. The resulting IF signals are amplified and detected and then applied to the vertical deflection plates of the CRT, producing a display of amplitude versus frequency.

(*c*) *Spectral Displays.* To better understand the application of spectrum analysis, it is important to have a clear understanding of the term *spectral display.*

A CRO is generally used to portray electrical signals with respect to time. For example, pulse rise time, pulsewidth, and repetition rate are read directly on the calibrated X axis of a CRT and the observed pattern is a plot of signal amplitude versus time. Such measurements are said to be in the *time domain.* In the spectrum analyzer, signals are broken down into their individual frequency components and displayed along the X axis of a CRT which is calibrated in frequency, for a presentation of signal amplitude versus frequency. These measurements are then said to be in the *frequency domain.*

Figure 11-23 gives a three-dimensional representation of a fundamental frequency (f_1) and its second harmonic ($2f_1$) and illustrates the time domain and the frequency domain characteristics. In Fig. 11-23(b) the two signals are portrayed as dashed lines; the solid line is the algebraic sum of the instantaneous values of the two signals, as seen on the CRT screen. In Fig. 11-23(c) the two signals are shown in the amplitude-frequency plane and are portrayed on the CRO as two components of the composite signal, as the *window* of the spectrum analyzer sweeps across the frequency range of the signal.

It is illustrative to consider the spectra of some common signals and the CRT display which results when these signals are applied to the spectrum analyzer.

(*1*) *Continuous-wave (CW) signals:* If the analyzer's local oscillator sweeps through a CW input signal slowly, the resulting response on the screen is simply a plot of the IF amplifier passband. A pure CW signal, by definition, has energy only at one frequency and should therefore appear as a single spike on the CRT screen. This will occur provided that the total RF sweepwidth or *spectrum width* is wide compared to the IF bandwidth in the analyzer.

(*2*) *Amplitude modulation:* When a CW signal of frequency f_c is amplitude modulated by a singel tone f_a, sidebands are generated at $f_c + f_a$ and $f_c - f_a$. The analyzer will then display the carrier frequency f_c, flanked by the two sideband frequencies whose

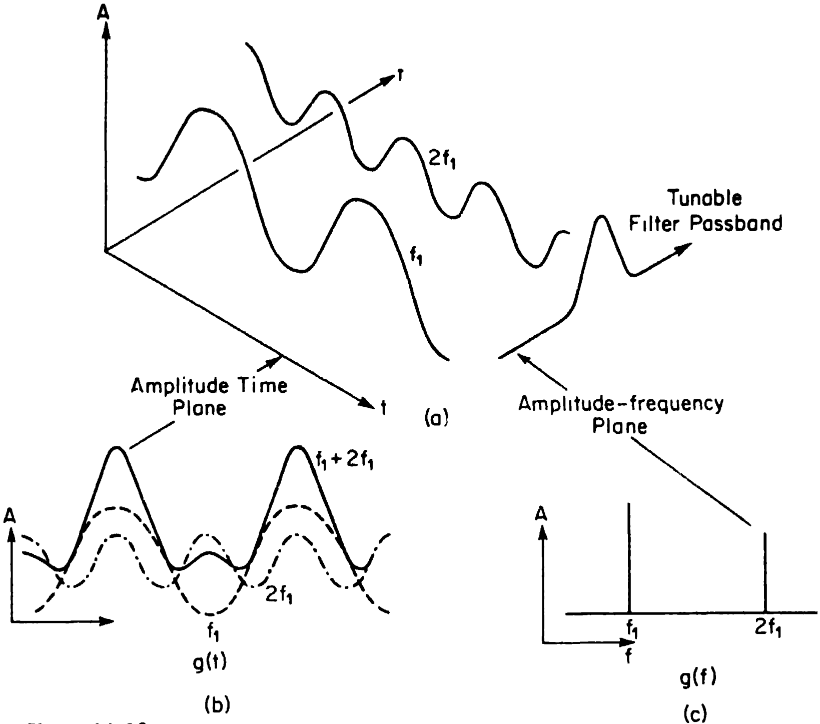


Figure 11-23
A three-dimensional presentation of amplitude, frequency, and time. (a) The addition of a fundamental and its second harmonic. (b) View seen in the t-A plane. On an oscilloscope, only the composite $f_1 + 2f_1$ would be seen. (c) View seen in the f-A plane. The components of the composite signal are clearly seen. (*Courtesy Hewlett-Packard Co.*)

amplitude relative to the carrier depends on the percentage of modulation, as shown in Fig. 11-24.

Provided that the frequency, the spectrum width, and the vertical response of the analyzer are calibrated, the display on the CRT screen provides the following (numerical) information: (a) carrier frequency, (b) modulation frequency, (c) modulation percentage, (d) nonlinear modulation, (e) incidental frequency modulation (as evidenced by jitter of the spectral lines), (f) spurious signal location and strength.

(*3*) *Frequency modulation:* If a CW signal f_c is frequency modulated at a rate f_r, it will produce an infinite number of sidebands. These are located at intervals of $f_c + nf_r$, where $n = 1, 2, 3, \ldots$. In practice, only the sidebands containing significant power are usually considered. An FM display is shown in Fig. 11-25.

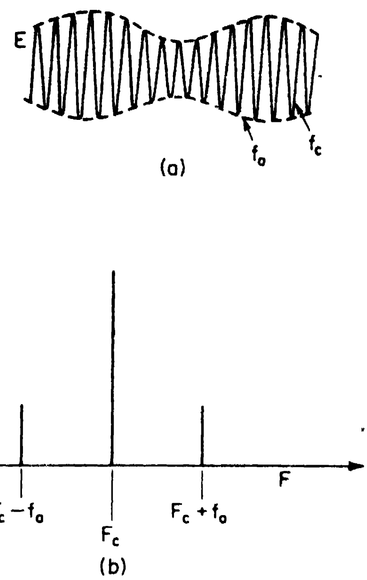


Figure 11-24
single-tone amplitude modulation. (a) Time-amplitude plot.) Frequency-amplitude plot.

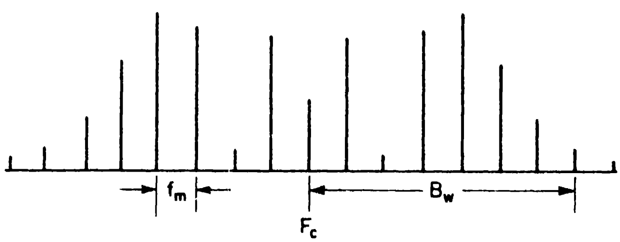


Figure 11-25
Amplitude spectrum of single-tone frequency modulation.

(4) *Pulse modulation:* An idealized rectangular waveform, with zero rise time and no overshoot or other aberrations, is given in Fig. 11-26(a). This pulse is shown in the time domain, but when its frequency spectrum is to be analyzed, it must be broken down into its individual frequency components. This is shown in Fig. 11-26(b), where a constant voltage, a fundamental frequency, and its third harmonic are added algebraically to form a wave which eventually becomes a square wave as more odd harmonics are added in phase with the fundamental. If an infinite number of odd harmonics were added, the resulting pulse would be perfectly rectangular. A spectral plot, in the frequency domain, would have the form given in Fig. 11-26(c), where the amplitudes and phases of an infinite number of harmonics are plotted, resulting in a smooth envelope as shown.

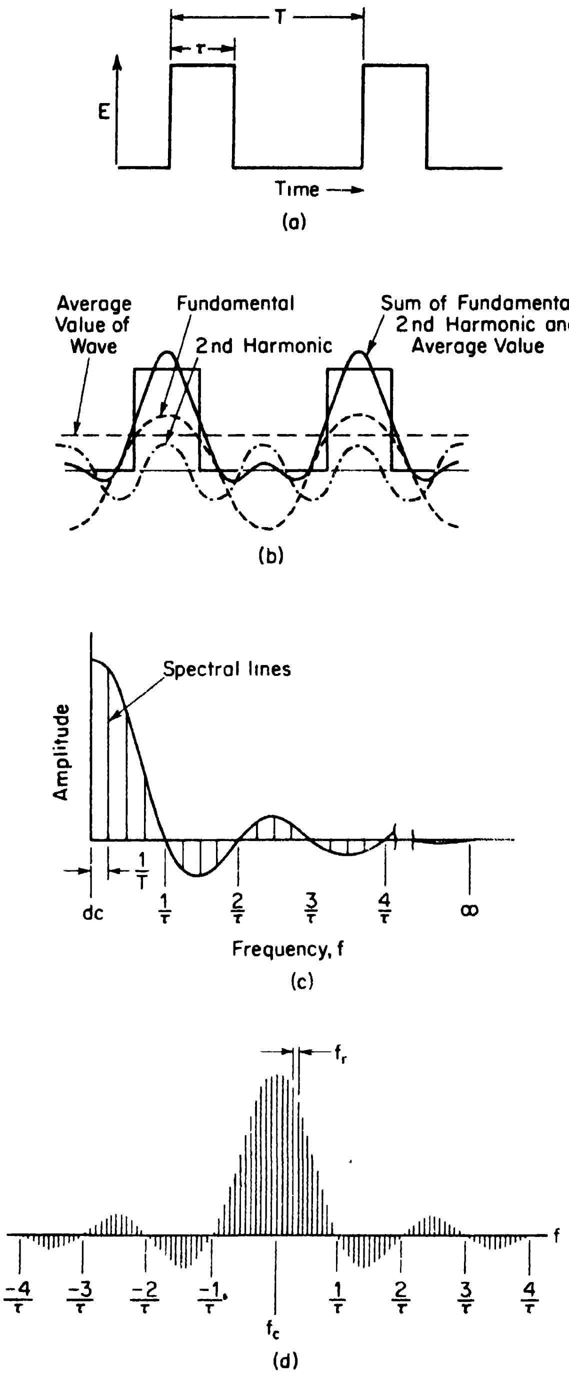


Figure 11-26

Pulse modulation. (a) Periodic rectangular pulse train. (b) Addition of a fundamental cosine wave and its harmonics to form rectangular pulses. (c) Spectrum of a perfectly rectangular pulse. Amplitudes and phases of an infinite number of harmonics are plotted resulting in smooth envelope as shown. (d) Resultant spectrum of a carrier amplitude, modulated with a rectangular pulse. (*Courtesy Hewlett-Packard Co.*)

When this pulse is used to amplitude modulate a carrier, the sums and differences of the carrier and *all* the harmonic components contained in the pulse are produced. The harmonic frequencies therefore produce multiple sidebands, in exactly the same manner that the modulating signal in amplitude modulation does. These multiple sidebands are generally referred to as *spectral lines* on the analyzer display. There will be twice as many sidebands or spectral lines as there are harmonic frequencies contained in the modulating pulse. Figure 11-26(d) shows the spectral plot resulting from rectangular pulse modulation of a carrier. The individual lines represent the modulation product of the carrier and the modulating pulse frequency with its harmonics. Therefore, the lines will be spaced in frequency by an amount equal to the pulse repetition rate of the original pulse waveform. The *main lobe* in the center and the *side lobes* are shown as groups of spectral lines extending above and below the base line. For a perfectly rectangular pulse, the number of side lobes is infinite. The main lobe contains the carrier frequency, represented by the longest line in the center.

The number of application possibilities is as great as the imagination of the user of this instrument. A few representative examples are cited:

(1) Pulse-width and repetition rate measurements.
(2) Tuning a parametric amplifier.
(3) FM deviation measurement.
(4) RF interference testing.
(5) Antenna pattern measurements.

References

1. Doyle, John N., *Pulse Fundamentals*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

2. Millman, Jacob, and Herbert Taub, *Pulse, Digital and Switching Waveforms*. New York,: McGraw-Hill Book Company, Inc., 1965.

3. Ryder, John D., *Electronic Fundamentals and Applications*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

Problems

1. With reference to Fig. 11-7, define the following terms: (a) pulsewidth, (b) rise time, (c) overshoot, (d) ringing, (e) duty cycle, (f) PRR.

2. Consider the Wien bridge oscillator of Fig. 11-2 and qualitatively explain why the frequency of oscillation depends on the following: (a) coupling capacitance, C_1, (b) Q_2 base resistor, R_4; (c) load resistance, R_6; (d) supply voltage, V_{CC}.

3. The phase shift oscillator of Fig. 11-3 uses three RC elements in cascade as the phase-shifting network between output and input. Show that a two-element RC network *cannot* work and explain why a four-element RC network is *not necessary*.

4. Which factors affect the stability and the accuracy of RC oscillators, such as the phase shift or Wien bridge?

5. What is the advantage of RC oscillators over LC oscillators? Which factors limit the high-frequency oscillation of the RC oscillator?

6. Consider the blocking oscillator of Fig. 11-9. (a) What factor(s) limit the pulse duration? (b) What effect does the ratio of pulsewidth to period have on the transistor or vacuum tube in the circuit?

7. Define the following terms: (a) harmonic distortion, (b) total harmonic distortion, (c) intermodulation distortion.

8. The Wien bridge circuit of Fig. 11-2 is slightly altered by adding an inductor L in series with the series RC combination and replacing the parallel RC combination by a resistor R_p. Calculate (a) the frequency of oscillation of this circuit; (b) the minimum gain of the two-stage amplifier for a finite value of R_p.

9. Verify Eq. (11-9) for the feedback factor of the phase-shift oscillator of Fig. 11-3 by applying conventional network analysis to the RC feedback circuit. Prove that the phase shift is 180° for $\alpha = 6$ and that at this frequency $\beta = \frac{1}{29}$.

10. Design a phase-shift oscillator to operate at a frequency of 10 kHz. Select a suitable transistor and find the minimum value for the load resistance (R_1 in Fig. 11-3) for which the circuit will oscillate. Calculate the RC product required for 10-kHz oscillation and calculate the value for C after choosing a reasonable value of R.

11. For the astable multivibrator of Fig. 11-8 the following component values are specified:

$$
\begin{aligned}
R_1 &= R_2 = 50\ k\Omega \\
C_1 &= C_2 = 0.02\ \mu\text{F} \\
R_3 &= R_4 = 1\ \text{k}\Omega \\
V_{BB} &= -10\ \text{V} \\
V_{CC} &= 10\ \text{V}
\end{aligned}
$$

Calculate (a) frequency of oscillation; (b) amplitude of the output pulse, appearing at the collector of Q_2; (c) duty cycle of the output waveform.

12. It is desired to modify the multivibrator of Problem 11 so that the duty cycle is reduced to 20 per cent while maintaining the original frequency. Indicate which circuit components must be changed and calculate the values of these components.

13. Using a single set of components, draw a circuit arrangement whereby the frequency of an astable multivibrator can be varied over an appreciable range.

14. Draw a circuit arrangement illustrating how an astable multivibrator may be used to display two traces on the screen of a single-beam CRO.

15. The blocking oscillator of Fig. 11-9 has the following component values: $R_1 = 100\ \mathrm{k\Omega}$; $C_1 = 0.02\ \mu\mathrm{F}$; $L_s = 1.2$ H; $C_d = 120$ pF. Calculate the approximate pulsewidth and the PRR.

TWELVE

NUMBER SYSTEMS

12-1 Introduction

The *decimal* number system, which requires ten digits, has been universally accepted as the system in which to perform calculations or to express quantities. It is, however, by no means the only number system nor is it the simplest to use in many applications. The *binary* number system, for example, is an *on-off* system, using only two digits (0 and 1) and is particularly suited for use in switching and logic circuits.

Many electronic devices characteristically operate in the binary number system. The bistable multivibrator or flip-flop is a good example of an electronic device having on-off characteristics: the flip-flop is either in one stable state or the other stable state; there are only these two possibilities. Because of the simplicity and reliability of binary electronic circuits, the binary number system has proved to be very efficient for use in digital computers. The result of an operation performed by a digital computer, however, may become available in a display register as a readout in the *octal* number system, because this may happen to be a more efficient way of making the information available than any other number system.

A recording device, such as an X-Y plotter, may receive its input information from a digital computer in terms of the binary number system. The output of the X-Y plotter, written by the pen on the recording paper, may represent a plot of voltage versus time. Obviously, the binary input signal has been translated into an analog voltage output.

Two types of *conversion* generally occur: (1) digital information (numbers representing some quantity) is transposed from one number system into the other number system; (2) information in digital form is converted into its corresponding analog equivalent, or conversely, analog information is converted into its digital equivalent. Since so many instruments in the field of gathering and handling data operate in the digital mode but accept input data or produce output data in analog form, it has become increasingly important to have some knowledge of the basic number systems and the basic conversion techniques.

This chapter deals with some of the common number systems and with conversion from one system into the other. Chapter 15 discusses some techniques used for converting analog information to digital information or digital to analog information.

12-2 Counting in Different Number Systems

The decimal number system is the common *radix-10* number system and uses ten symbols representing the quantities 0 through 9. Numbers larger than nine are constructed by assigning different values or *weights* to the position of the symbol with respect to the *decimal point.* It is necessary to learn only the ten basic numerals and the *positional notation* in order to count to any desired number.

Each position in a decimal number has a weight assigned to it which is ten times the value of the next position to the right. For example, in the decimal number 23, the positional value of the digit 2 is ten times the positional value of the digit 3. Every positional weight is a multiple of ten and can be expressed as 10 raised to some power. Starting at the decimal point, the positional weight of the *ones* position is 10^0 (= 1). The next position to the left, which is the *tens* position, carries a positional weight of 10^1 (= 10). The next position to the left, the *hundreds* position, carries a positional weight of 10^2 (= 100). This progression of increasing exponents can be carried as far to the left of the decimal point as desired. A similar progression can be extended to the *right* of the decimal point, but here the exponents are *negative.* The first position to the right of the decimal point is the *tenths* position with a positional weight of 10^{-1} ($= \frac{1}{10}$). The next position to the right is the *hundredths* position with a positional weight of 10^{-2} ($= \frac{1}{100}$).

Any decimal number can be represented by the diagram of Fig. 12-1,

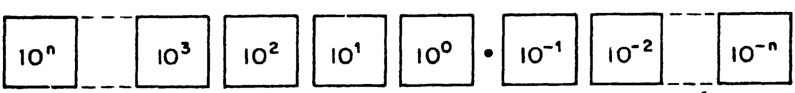


Figure 12-1
Skeleton for any decimal number, showing the positional weights of each digit with respect to the decimal point.

where the digit, placed in each "box," indicates how many multiples of the indicated power of ten are part of the total quantity represented by that number. The meaning of the positional value is illustrated by Example 12-1.

Example 12-1: Consider the decimal number 461.25 and write the value of each digit as represented by its relative position.

SOLUTION: $461.25 = 4 \times 10^2 + 6 \times 10 + 1 + 2 \times 10^{-1} + 5 \times 10^{-2}$
and, combining,

$$400 + 60 + 1 + 0.2 + 0.05 = 461.25$$

The number ten as the radix of a system (the number of symbols used in the system) is not a particularly magical choice. In fact, any number can be used as the radix of a number system, as long as the positional weights are adjusted accordingly. In the decimal system, only one value can be ascribed to a particular number, and in order to keep this feature in a number system with a different number of symbols, it is necessary to change the weights of the different positions. The values which must be assigned, turn out to be powers of the number of symbols available, that is, the *radix*.

The skeleton of a general number system with radix R is then given by Fig. 12-2.

$$\boxed{R^3}\ \boxed{R^2}\ \boxed{R^1}\ \boxed{R^0} \bullet \boxed{R^{-1}}\ \boxed{R^{-2}}\ \boxed{R^{-3}}$$

Figure 12-2
Skeleton number for a general number system using radix R.

The *binary* number system (with a radix of 2) uses only two symbols: 0 and 1. The positional value of the first digit to the left of the *binary point* is 2^0, following the skeleton layout of Fig. 12-2. The next position has a value of 2^1 $(= 2)$; the next position has a value of 2^2 $(= 4)$, and so on. Similarly, the positional values of the digits to the right of the binary point are given by Fig. 12-2 and the first digit has a value of 2^{-1} $(= \frac{1}{2})$. The next digit to the right has a positional value of 2^{-2} $(= \frac{1}{4})$, and so on.

A similar procedure may be carried out for number systems with a different radix. Examples of counting in a different number system are given in Table 12-1. The *octal* number system has a radix of eight; the *duodecimal* number system has a radix of twelve. In the last system, symbols A and B are used here to represent the quantities 10 and 11 (decimal equivalents). The positional weights are given in decimal values at the top of each column.

TABLE 12-1
SOME COMMON NUMBER SYSTEMS

Decimal		Binary					Octal		Duodecimal	
TENS	ONES	SIXTEENS	EIGHTS	FOURS	TWOS	ONES	EIGHTS	ONES	TWELVES	ONES
	0					0		0		0
	1					1		1		1
	2				1	0		2		2
	3				1	1		3		3
	4			1	0	0		4		4
	5			1	0	1		5		5
	6			1	1	0		6		6
	7			1	1	1		7		7
	8		1	0	0	0	1	0		8
	9		1	0	0	1	1	1		9
1	0		1	0	1	0	1	2		A
1	1		1	0	1	1	1	3		B
1	2		1	1	0	0	1	4	1	0
1	3		1	1	0	1	1	5	1	1
1	4		1	1	1	0	1	6	1	2
1	5		1	1	1	1	1	7	1	3
1	6	1	0	0	0	0	2	0	1	4
1	7	1	0	0	0	1	2	1	1	5
1	8	1	0	0	1	0	2	2	1	6
1	9	1	0	0	1	1	2	3	1	7
2	0	1	0	1	0	0	2	4	1	8
2	1	1	0	1	0	1	2	5	1	9
2	2	1	0	1	1	0	2	6	1	A
2	3	1	0	1	1	1	2	7	1	B
2	4	1	1	0	0	0	3	0	2	0

12-3 Binary-To-Decimal Conversion

Binary quantities are often translated into their decimal equivalents. This can be done electronically, and the techniques used are therefore important. The most obvious manner in which a binary number can be transposed into a decimal number is the *direct method,* a technique which follows directly from the skeleton number given in Fig. 12-2, where the positional weights are indicated by the powers of the radix. The following example illustrates this technique:

Example 12-2: Convert the binary number 101011 into its decimal equivalent by the direct method.

SOLUTION: The binary number is rearranged with the corresponding powers of 2 for each bit in the following way:

$$\begin{aligned} 101011 &= 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &= 32 + 0 + 8 + 0 + 2 + 1 = 43 \end{aligned}$$

The solution shows the decimal equivalent for each bit, determined by its position in the original number. Therefore,

$$\text{binary } 101011 = \text{decimal } 43$$

Binary fractions are converted in exactly the same manner. The corresponding positional values are assigned to each bit and the number is converted as shown in the following examples:

Example 12-3: Convert the binary number 110.110 into its decimal equivalent by using the direct method of conversion.

SOLUTION: The binary number is rearranged with the corresponding powers of 2 for each bit, according to the general skeleton given in Fig. 12-2:

$$\begin{aligned} 110.110 &= 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 + 1 \times 2^{-1} \\ &\quad + 1 \times 2^{-2} + 0 \times 2^{-3} \\ &= 4 + 2 + 0 + 0.5 + 0.25 + 0 \\ &= 6.75 \end{aligned}$$

Therefore,

$$\text{binary } 110.110 = \text{decimal } 6.75$$

Example 12-4: Convert the binary fraction 0.1111 into its decimal equivalent.

SOLUTION: The corresponding powers of 2 are assigned to each bit position and the number is written as follows:

$$\begin{aligned} 0.1111 &= 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3} + 1 \times 2^{-4} \\ &= 0.5 + 0.25 + 0.125 + 0.0625 \\ &= 0.9375 \end{aligned}$$

Therefore,

binary 0.111 = decimal 0.9375

The double-dabble method converts a binary number into its decimal equivalent in a different way. This method is preferred by some people, although at first glance it appears to be more difficult than the direct method. It requires the following steps:

(1) Multiply the high-order bit by two. Add this product to the next highest-order bit.

(2) Multiply the total obtained in step (1) by two. Add this new product to the next highest-order bit.

(3) Repeat this process until all the bits have been processed.

The final total is the decimal equivalent of the original binary quantity.

Example 12-5: Convert the binary number 101011 (see Example 12-2) into its decimal equivalent by the double-dabble method.

SOLUTION: The procedure just outlined is followed and the conversion proceeds as shown:

Step 1: $(1 \times 2) + 0 = 2$
Step 2: $(2 \times 2) + 1 = 5$
Step 3: $(5 \times 2) + 0 = 10$
Step 4: $(10 \times 2) + 1 = 21$
Step 5: $(21 \times 2) + 1 = 43$

Therefore,

binary 101011 = decimal 43

This answer is, of course, identical to the answer obtained in Example 12-2, where the same binary number was converted.

12-4 Decimal-to-Binary Conversion

The process of converting binary numbers into their decimal equivalents is usually done at the output terminals of a binary device (such as a digital computer). Assuming that the computer operates in the binary number system, the original input signals into the computer must be converted into their binary equivalents. Several techniques may be used to perform this conversion.

The *direct method* examines the magnitude of the decimal number and subtracts the largest possible power of two from it. This is repeated with the remainder until there is no remainder left. The power of two which was subtracted from the number at various stages in the process determines the position of the binary bits in the final answer. The method is illustrated in Example 12-6.

Example 12-6: Convert the decimal number 51 into its binary equivalent by using the direct method of conversion.

SOLUTION:

Step 1: The largest possible power of 2 in 51 is $2^5(= 32)$. Subtracting 32 from 51 gives a remainder of 19.

Step 2: The largest possible power of 2 in 19 is $2^4(= 16)$. Subtracting 16 from 19 gives 3.

Step 3: The largest possible power of 2 in 3 is $2^1(= 2)$. Subtracting 2 from 3 gives 1.

Step 4: The largest possible power of 2 in 1 is $2^0(= 1)$. Subtracting 1 from 1 gives 0.

Combining the powers of 2 in their appropriate places:

$$51 = 1 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0.$$

Therefore,

$$\text{decimal } 51 = \text{binary } 110011$$

The *double-dabble method* of converting a decimal number into its binary equivalent is the reverse of the double-dabble method for converting binary into decimal. The method requires the following steps:

(1) Divide the decimal number by 2. Record the remainder.

(2) Divide the first quotient by 2. Record the remainder.

(3) Divide the next quotient by 2. Record the remainder.

(4) Repeat this process until the final quotient is 0.

(5) The remainders, starting with the last one obtained, make up the equivalent binary number.

Example 12-7: Convert the decimal number 37 into its binary equivalent by the double-dabble method.

SOLUTION: The conversion proceeds smoothly when columns for the quotients and remainders are set up.

Number		Quotient	Remainder
$37 \div 2$	=	18	1
$18 \div 2$	=	9	0
$9 \div 2$	=	4	1
$4 \div 2$	=	2	0
$2 \div 2$	=	1	0
$1 \div 2$	=	0	1

Reading the remainder column from bottom to top yields the answer:

decimal 37 = binary 100101

When a *decimal fraction* is to be converted into its binary equivalent, the following rules are applied:

(1) Multiply the original decimal fraction by 2. Remove the integer part from the product and record the integer.

(2) Multiply the remaining part of the product by 2. Remove the new integer and record it. If there is no integer, indicate this by recording a 0.

(3) Repeat this process until the remaining fractional portion of the product is 0, or until the required number of binary digits are obtained.

(4) The binary equivalent is equal to the integers which were removed from the intermediate products, starting with the binary point to the left of the first integer, obtained in step (1).

Example 12-8: Convert the decimal fraction 0.6372 into its binary equivalent.

SOLUTION: The conversion again proceeds best when columns are set up for the products and the integers.

Number		Product	Integer
0.6372×2	=	1.2744	1
0.2744×2	=	0.5488	0
0.5488×2	=	1.0976	1
0.0976×2	=	0.1952	0
0.1952×2	=	0.3904	0
0.3904×2	=	0.7808	0
0.7808×2	=	1.5616	1
0.5616×2	=	1.1232	1

and so on.

When the last product is other than 0, the fraction is not terminated and the result shows this with a + sign at the right end. Therefore,

$$\text{decimal } 0.6372 = \text{binarv } 0.10100011+$$

12-5 Basic Arithmetic Operations in the Binary Number System

The arithmetic processes of addition, subtraction, multiplication, and division in the binary number system are identical to those used in the decimal number system.

(a) *Binary Addition.* Since the binary number system has only two digits, a 0 and a 1, *a carry of 1* to the next higher digit position is required when the limit of a position is exceeded. The four possibilities for addition in the binary system are

$$\begin{aligned} 0 + 0 &= 0 \\ 0 + 1 &= 1 \\ 1 + 0 &= 1 \\ 1 + 1 &= 0 \quad \text{with a carry of 1.} \end{aligned}$$

The process of addition is illustrated in Example 12-9.

Example 12-9: Find the sum of the binary numbers 1101 and 1001.

SOLUTION: Arranging the numbers in columns and allowing some room to record the carries,

$$\begin{array}{cccccc l} & 1 & & 1 & & & \text{carry} \\ & & 1 & 1 & 0 & 1 & \text{first number} \\ + & & 1 & 0 & 0 & 1 & \text{second number} \\ \hline & 1 & 0 & 1 & 1 & 0 & \text{sum} \end{array}$$

The answer can be checked by converting the three binary numbers to their decimal equivalents:

$$\begin{aligned} \text{Binary } 1101 &= 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 8 + 4 + 0 + 1 = 13 \text{ (decimal)} \end{aligned}$$

$$\begin{aligned} \text{Binary } 1001 &= 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 8 + 0 + 0 + 1 = 9 \text{ (decimal)} \end{aligned}$$

$$\begin{aligned} \text{Binary } 10110 &= 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 \\ &= 16 + 0 + 4 + 2 + 0 = 22 \quad \text{(decimal)} \end{aligned}$$

Adding in decimal notation,

$$13 + 9 = 22$$

verifying that the binary addition was performed correctly.

Binary fractions are added in exactly the same manner as the integer binary numbers. The procedure is illustrated in Example 12-10.

Example 12-10: Find the sum of 101.101 and 110.100.

SOLUTION: Arranging the numbers in columns and allowing some space to record the carries,

$$\begin{array}{rl} 1\,1\,1\,1 & \text{carry} \\ 1\,0\,1.1\,0\,1 & \\ +\;\; 1\,1\,0.1\,0\,0 & \\ \hline 1\,1\,0\,0.0\,0\,1 & \end{array}$$

(b) *Binary Subtraction.* The simplest type of subtraction in the binary system is *direct subtraction.* (This method of subtraction is not used in digital computers, where it would require a rather large amount of additional logic circuits. It is usually much simpler to use the existing "add" circuitry in the *complemented* method of subtraction.)

Direct subtraction is performed according to these simple rules:

$$\begin{array}{l} 0 - 0 = 0 \\ 1 - 1 = 0 \\ 1 - 0 = 1 \\ 0 - 1 = 1 \quad \text{with a } \textit{borrow of 1.} \end{array}$$

The process of direct subtraction is illustrated by Example 12-11.

Example 12-11: Subtract 1010 from 1111.

SOLUTION:

$$\begin{array}{rl} 1\,1\,1\,1 & \text{minuend} \\ -\,1\,0\,1\,0 & \text{subtrahend} \\ \hline 1\,0\,1 & \text{remainder} \end{array}$$

No borrows were necessary and the process is self-explanatory.

Consider Example 12-12 in which borrowing is required.

Example 12-12. Subtract 110 from 11010.

SOLUTION: Arranging the numbers in columns and allowing some space above the columns to record the borrows,

$$\begin{array}{rl} 0\;\,10 & \text{changes in minuend after borrowing} \\ 1\;\,1\;\,0\;\,1\;\,0 & \text{original minuend} \\ -\qquad 1\;\,1\;\,0 & \text{subtrahend} \\ \hline 1\;\,0\;\,1\;\,0\;\,0 & \text{remainder} \end{array}$$

The operation can be checked by converting the minuend, subtrahend, and remainder into their decimal equivalents.

$$\begin{aligned}
\text{binary } 11010 &= 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 \\
&= 16 + 8 + 0 + 2 + 0 = 26 \quad \text{(decimal)} \\
\text{binary } 110 &= 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 \\
&= 4 + 2 + 0 = 6 \quad \text{(decimal)} \\
\text{binary } 10100 &= 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 \\
&= 16 + 0 + 4 + 0 + 0 = 20 \quad \text{(decimal)}
\end{aligned}$$

Subtracting in decimal notation,

$$26 - 6 = 20$$

which checks with the binary subtraction.

(c) *Binary Multiplication.* Multiplication with binary numbers follows essentially the same rules as multiplication with decimal numbers. Since both the multiplicand and the multiplier contain only the digits 0 and 1, the multiplication becomes exceedingly simple: A binary number which is multiplied by (binary) 1, equals that binary number; a binary number multiplied by (binary) 0, equals 0. When the partial products are added, the established rules for binary addition must be observed.

The procedure of binary multiplication is illustrated in Example 12-13.

Example 12-13: Compute the product of the two binary numbers 101 and 10.

Solution: The binary numbers are set up for multiplication in the usual manner:

$$\begin{array}{rl}
1\ 0\ 1 & \text{multiplicand} \\
1\ 0 & \text{multiplier} \\
\\
0\ 0\ 0 & \text{first partial product} \\
1\ 0\ 1 & \text{second partial product} \\
\hline
1\ 0\ 1\ 0 & \text{final product (sum of the partial products)}
\end{array}$$

The multiplication can be checked by converting the binary numbers into their decimal equivalents:

$$\begin{aligned}
\text{binary } 101 &= 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 4 + 0 + 1 = 5 \text{ (decimal)} \\
\text{binary } 10 &= 1 \times 2^1 + 0 \times 2^0 = 2 + 0 = 2 \quad \text{(decimal)} \\
\text{binary } 1010 &= 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 \\
&= 8 + 0 + 2 + 0 = 10 \quad \text{(decimal)}
\end{aligned}$$

Multiplying in decimal notation,

$$5 \times 2 = 10$$

which checks with the binary multiplication.

A more challenging problem is given in Example 12-14.

Example 12-14: Compute the product of 11011 and 1011; check the answer by converting the binary numbers into decimal equivalents.

SOLUTION: The binary numbers are set up for multiplication.

$$
\begin{array}{r}
1\;1\;0\;1\;1 \\
1\;0\;1\;1 \\
\hline
1\;1\;0\;1\;1 \\
1\;1\;0\;1\;1 \\
0\;0\;0\;0\;0 \\
1\;1\;0\;1\;1 \\
\hline
1\;0\;0\;1\;0\;1\;0\;0\;1
\end{array}
$$

The multiplicand, multiplier, and product are converted into their decimal equivalents:

$$
\begin{aligned}
\text{multiplicand: binary } 11011 &= 1\times 2^4+1\times 2^3+0\times 2^2+1\times 2^1+1\times 2^0 \\
&= 16+8+0+2+1 \\
&= 27 \quad \text{(decimal)}
\end{aligned}
$$

$$
\begin{aligned}
\text{multiplier: binary } 1011 &= 1\times 2^3+0\times 2^2+1\times 2^1+1\times 2^0 \\
&= 8+0+2+1 \\
&= 11 \quad \text{(decimal)}
\end{aligned}
$$

$$
\begin{aligned}
\text{product: binary } 100101001 &= 1\times 2^8+1\times 2^5+1\times 2^3+1\times 2^0 \\
&= 256+32+8+1 \\
&= 297 \quad \text{(decimal)}
\end{aligned}
$$

The product of the decimal numbers is

$$27\times 11 = 297$$

which checks with the product of the binary numbers.

(d) *Binary Division.* The general rules for decimal division also apply in binary division. Be careful also to obey the rules for binary multiplication and binary subtraction discussed in the preceding paragraphs. The division process is illustrated in Example 12-15.

Example 12-15: Divide the binary number 1010 by the binary number 10.

SOLUTION: The divisor and the dividend are placed in the conventional positions, leaving some space to record the quotient above the dividend.

$$\begin{array}{r l}
 & \;1\;0\;1 \quad \text{quotient} \\
\text{divisor}\;\; 1\;0 & \overline{)\,1\;0\;1\;0} \quad \text{dividend} \\
 & \,1\;0 \\
 & \,\overline{0\;1\;0} \\
 & \;1\;0 \\
 & \;\overline{0\;0} \quad \text{remainder}
\end{array}$$

In most cases the remainder will not be 0. The *binary fraction* must then be included in the final answer, as shown in Example 12-16.

Example 12-16: Find the quotient of 10011 divided by 101.

SOLUTION:

$$\begin{array}{r r}
 & 1\;1 \\
1\;0\;1 & |\,1\;0\;0\;1\;1 \\
 & 1\;0\;1 \\
 & 1\;0\;0\;1 \\
 & 1\;0\;1 \\
 & 1\;0\;0
\end{array}$$

The quotient equals 11 and the remainder equals 100. The answer is then correctly written as $11\frac{100}{101}$. The problem can be reworked in decimal notation and the answer checked.

binary divisor 101 = decimal 5
binary dividend 10011 = decimal 19
binary quotient 11 = decimal 3
binary remainder $\frac{100}{101}$ = decimal $\frac{4}{5}$

The decimal division yields

$$19 \div 5 = 3\tfrac{4}{5}$$

The binary division therefore was correct.

Several problems for practicing simple binary arithmetic are given at the end of this chapter.

12-6 Conversions in the Octal Number System

The octal number system has a radix of eight and uses the symbols 0, 1, 2, 3, 4, 5, 6, and 7. The positional value of each digit in an octal number follows the arrangement given in the skeleton number of Fig. 12-2. The powers of eight, associated with each position, increase to the left of the octal point and decrease to the right of the octal point.

(a) *Octal-to-Decimal Conversion.* Octal numbers are easily converted to their decimal equivalents by using the technique of positional value of the digits The procedure is identical to that given for converting binary numbers to their decimal equivalents. The method is illustrated in Example 12-17.

Example 12-17: Convert octal 264 to decimal notation.

SOLUTION: The octal number is rearranged with the corresponding powers of eight for each bit:

$$\begin{aligned} \text{octal } 264 &= 2 \times 8^2 + 6 \times 8^1 + 4 \times 8^0 \\ &= 128 + 48 + 4 \\ &= 180 \quad \text{(decimal)} \end{aligned}$$

The solution shows the decimal equivalent for each bit, determined by its position with respect to the octal point.

(b) *Decimal-to-Octal Conversion.* The best method for converting a decimal number into its octal equivalent is the *double-dabble* method, which was first introduced for the decimal-to-binary conversion (Sec. 12-3). The steps to be followed are quite similar and allowance need be made only for the change in radix from 2 in the binary system to 8 in the octal system.

The following steps are required:

(1) Divide the decimal integer by eight. Record the remainder in a separate column.

(2) Divide the quotient obtained in step (1) by eight. Record the remainder again.

(3) Repeat the process until the quotient is 0.

(4) The octal equivalent is given by the column of remainders written from the bottom to the top.

The procedure is illustrated in Example 12-18.

Example 12-18: Convert the decimal number 428 into its octal equivalent.

SOLUTION: A simple table is set up and the method just outlined is followed.

Number		Quotient	Remainder
$428 \div 8$	=	53	4
$53 \div 8$	=	6	5
$6 \div 8$	=	0	6

Reading the remainder column from bottom to top yields

$$\text{decimal } 428 = \text{octal } 654$$

The answer may be checked by converting the octal number 654 back into its decimal equivalent.

$$\begin{aligned}\text{octal } 654 &= 6 \times 8^2 + 5 \times 8^1 + 4 \times 8^0 \\ &= 384 + 40 + 4 \\ &= \text{decimal } 428\end{aligned}$$

(c) *Conversion from Binary to Octal.* Since the octal system finds application in digital computers, where binary numbers are often recorded in octal notation, conversions from the binary to the octal system and from the octal to the binary system should be understood. Also, a large decimal number is most easily translated into its binary equivalent by first converting it into octal notation.

To convert a binary number into an octal number the following rules apply:

(1) Starting at the binary point, group the binary number into three-digit clusters, both to left and right of the binary point.

(2) In order to complete the groups of three digits when forming the clusters, zeros may be placed at the extreme right and left ends of the original binary number.

(3) Read out each group of three binary numbers. The combined readout is the octal equivalent of the original binary number.

The procedure is given in Example 12-19.

Example 12-19: Convert the binary number 11010110.1001 into its octal equivalent.

Solution: The original binary number is separated into group-of-3 clusters, adding zeros where necessary.

$$011\ 010\ 110\ .\ 100\ 100$$

Reading the binary value of each cluster gives:

$$011\ 010\ 110\ .\ 100\ 100 = 3\ 2\ 6\ .\ 4\ 4$$

Therefore,

$$\text{binary } 11010110.1001 = \text{octal } 326.44$$

An octal number may be converted into its binary equivalent with equal ease. The procedure is essentially the reverse of the binary-to-octal conversion and is given in Example 12-20:

Example 12-20: Convert the octal number 372.25 into its binary equivalent.

SOLUTION: Each digit in the original octal number is stated in a binary cluster-of-3:

$$372.25 = 011\ 111\ 010\ .\ 010\ 101$$

Combining the cluster,

$$\text{octal } 372.25 = \text{binary } 11111010.010101$$

12-7 Binary-Coded Decimal Numbers

The binary-coded-decimal (BCD) notation is a special way of expressing decimal quantities in binary form. A variety of special codes may be used, but they are all referred to as *BCD notation.* In BCD notation, each digit of the original decimal number is expressed directly as a four-digit binary group. The number of binary groups is then equal to the number of digits in the original decimal number.

One often used BCD notation is the so-called 8-4-2-1 code. This is often referred to simply as *the BCD code,* since the weights of the positions are the same as in the binary number system, as illustrated by Example 12-21.

Example 12-21: Express the decimal number 2,963 in the 8-4-2 1 code.

SOLUTION: Each digit of the decimal number is expressed in binary form as follows:

$$2 = 0010$$
$$9 = 1001$$
$$6 = 0110$$
$$3 = 0011$$

The binary groups are moved together to form the BCD notation and

$$\text{decimal } 2963 = 0010\ 1001\ 0110\ 0011 \quad \text{(BCD)}$$

Notice that the binary groups are separated by one space to indicate that the number is not a pure binary number but is expressed in a hybrid number system.

A typical example of the 8-4-2-1 code is the perforated tape used with digital computers. Figure 12-3 illustrates how the binary bits are represented

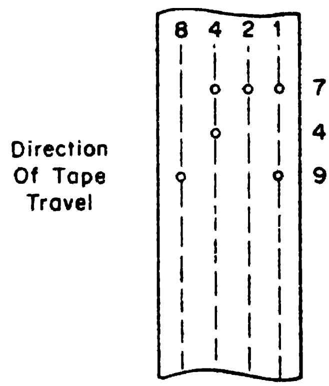


Figure 12-3
The decimal number 749 is punched on a paper tape in binary-coded-decimal (BCD) notation.

on the tape by small punched holes. A hole indicates a binary 1; the absence of a hole indicates a binary 0. A total of four poistions are available across the width of the tape to accommodate the four-digit code. The holes, punched in the tape illustrated in Fig. 12-3, represent the decimal number 749, if the tape is read from the top of the figure to the bottom.

BCD numbers do not always follow the pure binary number system. Special-purpose systems, such as the Gray code, the excess-3 code, and the biquinary code are often used in computer systems.

12-8 Notation

In working with different number systems, it is extremely important to be certain which system is being used. Where there may be some doubt as to the number system being used, it is good practice to indicate this by writing the radix of the system, in decimal notation, as a subscript to the number.

For instance, the number 625_8 indicates that this is the number 625 written in the octal number system. This same number, written in decimal notation, would be 405_{10}.

Problems

1. Convert the following binary quantities into their decimal equivalents:

(a) 101	(i) 0.0101
(b) 11010	(j) 0.111001
(c) 11111	(k) 10.10
(d) 1000001	(l) 110.011
(e) 0.1	(m) 111.111
(f) 0.101	(n) 1110.001
(g) 0.1111	(o) 1000.001
(h) 0.001	(p) 111000.101

2. Convert the following decimal quantities into their binary equivalents:

(a) 13	(e) 0.25	(i) 2.68
(b) 32	(f) 0.326	(j) 12.76
(c) 41	(g) 0.572	(k) 37.82
(d) 93	(h) 0.6841	(l) 476.663

3. Find the sum of the following binary quantities.

(a) 101 and 110
(b) 111 and 111
(c) 1000 and 1011
(d) 110011 and 1111
(e) 101010 and 10101
(f) 10.111 and 1.11
(g) 1010.10 and 110.11
(h) 101.10001 and 1010.111
(i) 1.111101 and 111.001
(j) 1110011 and 1001100011

4. Subtract the following binary quantities:

(a) 100 from 110
(b) 110 from 1110
(c) 100 from 100001
(d) 10001 from 11101
(e) 1001100 from 11001101
(f) 10111 from 101000
(g) 1011001 from 10110010
(h) 1001001001 from 1101101100
(i) 1000111 from 110000000
(j) 111111 from 1000000

5. Compute the products of the following binary quantities:

(a) 10 and 01
(b) 11 and 101
(c) 111 and 100
(d) 110 and 111
(e) 101 and 101
(f) 1101 and 1010
(g) 1001 and 1101
(h) 10101 and 010
(i) 1001 and 1111
(j) 1101 and 11011

6. Perform the following binary divisions:

(a) $100 \div 10$
(b) $111 \div 10$
(c) $111 \div 11$
(d) $100 \div 11$
(e) $1101 \div 101$
(f) $11001 \div 11$
(g) $10101 \div 101$
(h) $1001101 \div 10$
(i) $11001100 \div 111$
(j) $11111 \div 1001$

7. Convert the following octal numbers into their decimal equivalents:

(a) 25
(b) 36
(c) 77
(d) 152
(e) 256
(f) 351
(g) 334
(h) 2662
(i) 45321

8. Convert the following decimal numbers into their octal equivalents:

(a) 29
(b) 362
(c) 112
(d) 588
(e) 642
(f) 361
(g) 667
(h) 811
(i) 328

9. Convert the following octal quantities into their binary equivalents:

(a) 25
(b) 11
(c) 254
(d) 543
(e) 2667
(f) 2315
(g) 452
(h) 777
(i) 1026

10. Convert the following binary quantities into their octal equivalents:

(a) 1010
(b) 1111
(c) 10000
(d) 111011
(e) 100100
(f) 10101
(g) 1000001
(h) 10100
(i) 11001100
(j) 101110111

11. Convert the following decimal quantities into the 8–4–2–1 BCD notation:

(a) 130	(d) 3765	(g) 337
(b) 249	(e) 1001	(h) 2999
(c) 485	(f) 8163	(i) 1915

12. Express the following binary-coded decimal numbers (8–4–2–1 code) in their equivalent decimal numbers:

(a) 1001 0001	(g) 1000 0001 1001
(b) 0100 0101	(h) 0101 1001 0101
(c) 1000 0101	(i) 1001 0101 0001
(d) 0111 0111	(j) 0011 0000 1001 0011
(e) 0101 1001	(k) 1001 1000 0111 0100
(f) 0001 0111	(l) 1001 0111 1000 0101

THIRTEEN

FREQUENCY- AND TIME-MEASURING INSTRUMENTATION

13-1 Introduction

Electronic counters have proved to be the most accurate, flexible, and convenient instruments for making both *frequency* and *time-interval* measurements. Electronic counters and associated equipment can measure frequencies from dc to the gigahertz range and time intervals from nanoseconds to days. Although many vacuum tube models of electronic counters are still available, they have generally been superseded by solid-state instruments. This discussion deals only with the circuits of these newer instruments.

An *electronic counter* is an instrument designed to measure an unknown frequency or time interval by *comparing* it to a known frequency or time interval. The counter's logic is designed to present the result of this comparison measurement in an easy-to-read numerical display. The *accuracy* of the measurement depends primarily upon the stability of the known frequency which is derived from the counter's internal oscillator.

The electronic counter has several functional sections, which can be interconnected in different ways to make different types of measurement. The most important subsystems of the counter are

(1) *Decade-counting assemblies* (DCAs) with their numerical display, to totalize and display the count.

(2) *Signal gate,* which controls the start and stop times of the actual count.

(3) *Time-base,* which supplies precise increments of time to control the gate for a frequency or a time-interval measurement.

Other sections of a general-purpose counter include circuitry for signal shaping, logic control, and binary-coded-decimal output circuitry.

The first half of this chapter deals with some of the basic building blocks of the electronic counter; the second half is concerned with the application of counters in some typical measurement situations.

13-2 The Decade-Counting Assembly (DCA)

A *decade-counting assembly* is a circuit which produces one single output pulse for every ten input pulses applied to the circuit. Since the base (the radix) of our numbering system is *ten*, the *decade* counter is by far the most widely used of all counting assemblies. A counting assembly uses bistable multivibrators, more commonly known as *flip-flops.* A flip-flop is a circuit with two stable states which is capable of remaining in either state indefinitely until triggered by an external signal. This indicates that the flip-flop has a *memory;* it will remember or store the last signal it received until a new signal is applied to the circuit.

One very common form of the bistable multivibrator is the emitter-coupled circuit of Fig. 13-1. This circuit forms an entirely symmetrical arrangement of two transistors, Q_1 and Q_2, together with their associated resistive and capacitive components. In order to explain the operation of the circuit, the assumption is made that initially transistor Q_1 is conducting heavily. Notice that we have two PNP transistors using a negative voltage collector supply.

When Q_1 is conducting, its collector voltage is only very slightly negative and is assumed to be zero for the sake of clarity. The base of Q_2, which is connected to the junction of R_3 and R_6, is then also zero volts. The relatively heavy current through Q_1 causes a voltage drop across R_7 which produces sufficient negative voltage at the emitter of Q_2 to hold it at cutoff. With Q_2 cut off, voltage divider R_2-R_4-R_5 delivers a negative voltage to the base of Q_1 and keeps it conducting The flip-flop is therefore in a stable state with Q_1 conducting and Q_2 cut off.

This stable state can be reversed by applying a positive voltage to the *trigger input* terminal. Resistors R_8 and R_9 connect the bases of both transistors to this input terminal through capacitor C_3. Now consider the waveforms shown in the diagram of Fig. 13-1. At time $t = t_1$, a positive input voltage is applied

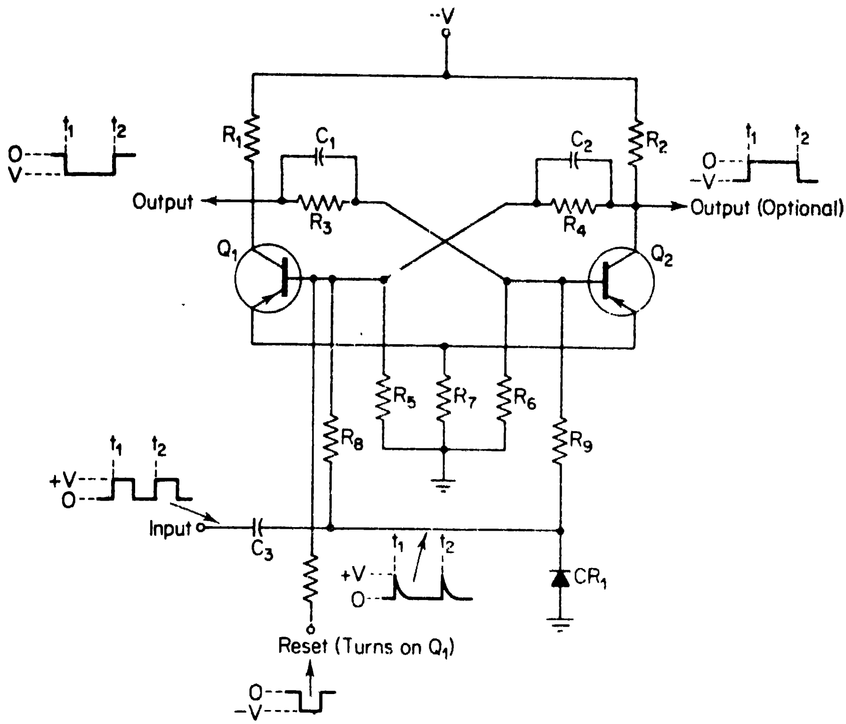


Figure 13-1
Circuit diagram of an emitter-coupled bistable multivibrator or flip-flop. Relevant waveforms are indicated.

to the trigger input terminal, cutting Q_1 off. The Q_1 collector potential goes negative toward the supply voltage. This drives Q_2 into conduction since the base of Q_2 is connected to the R_3-R_6 junction. When Q_2 conducts, its collector voltage becomes considerably less and so does the base of Q_1, allowing Q_1 to remain cut off. The R_1-R_3-R_6 divider delivers a sufficiently large negative voltage to the base of Q_2 to drive it into heavy conduction. The application of this positive input pulse has therefore caused the flip-flop to go to its other stable state (where Q_2 is conducting and Q_1 is cut off). The flip-flop remains in this state until another positive trigger pulse comes along, causing a new transition of states. This next positive pulse, occurring at time $t = t_2$, cuts off Q_2 and starts a sequence of events which ends with Q_1 conducting and Q_2 cut off.

Notice that a positive trigger pulse has no effect on the transistor which is already cut off; it affects only the *conducting* transistor. The waveforms which accompany the diagram of Fig. 13-1 indicate that one *output* pulse is present at the collector of either Q_1 or Q_2 for every two *input* pulses triggering the flip-flop. This means that the circuit divides by two.

Diode CR_1 in the circuit of Fig. 13-1 removes the negative pulse from the differentiated input waveform. Without this diode, the negative pulse would drive the cut-off transistor *on* and the stage would switch from one state to the other but it would not divide by two. The ac coupling through capacitors C_1 and C_2 assists the circuit in making fast transitions between states; they are called speed-up or *commutating* capacitors. The dc coupling through R_3 and R_4 insures the bistable characteristics of the circuit. A flip-flop which completes its cycle and produces one output pulse after receiving two input pulses, is called a *binary circuit,* since it is counting in the binary system, by twos. Another feature of the circuit of Fig. 13-1 is the *reset* terminal. A negative pulse, applied to this terminal, reaches the base of Q_1 and forces it to conduct. The flip-flop then returns or resets to its initial condition with Q_1 conducting and Q_2 cut off.

Now consider Fig. 13-2 where four binaries, represented by their logic

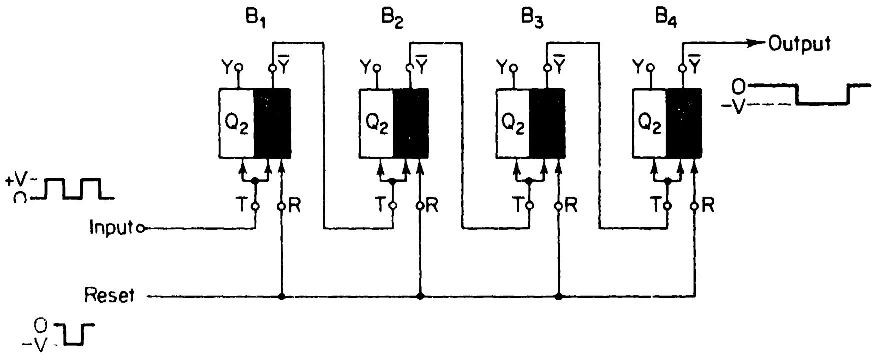


Figure 13-2
Four flip-flops in cascade, forming a binary counter. The zero state of each binary is represented by the shaded half of the logic symbol, where transistor Q_1 is conducting.

symbols, are connected in cascade. For convenience and ease of reference, assume that the flip-flops used in this cascaded arrangement, are identical to the circuit of Fig. 13-1. In the logic symbols, point T represents the input or *trigger* terminal. The bases of both transistors are connected to this point so that each successive positive trigger pulse will reverse the state of the binary. The points marked Y and $\bar{Y}$ (*read as Y-bar*) represent the collectors of the transistors. Comparing the logic symbol with the actual circuit diagram, one sees that the collector of Q_1 has been designated the $\bar{Y}$ output terminal and the collector of Q_2, the Y output terminal. The $\bar{Y}$ output terminal of each binary in the chain is connected to the trigger input of the next binary. The positive-going transition at the collector of Q_1 (while switching states from Q_1-conducting to Q_1-conducting) provides the positive input pulse to trigger the next binary. Arbitrarily state that a binary is in state 0 when transistor Q_1 conducts. The binary is in state 1 when this transistor does *not* conduct. When a binary is in state 1, its output voltage (at the $\bar{Y}$ terminal) is highly negative; when the binary is in state 0, its output voltage is approximately zero volts. A transition from the 1 state to the 0 state then produces a voltage step in the positive-going direction; a transition from the 0 state to the 1 state produces a negative-going output pulse. In order to start our discussion, a reset pulse, consisting of a negative voltage, is applied to the bases of all the Q_1 transistors, forcing them to conduct. By our earlier convention, all the binaries are now in state 0.

Starting from this reference position we can observe the waveforms, shown in Fig. 13-3, which appear at the output of each binary as the result of

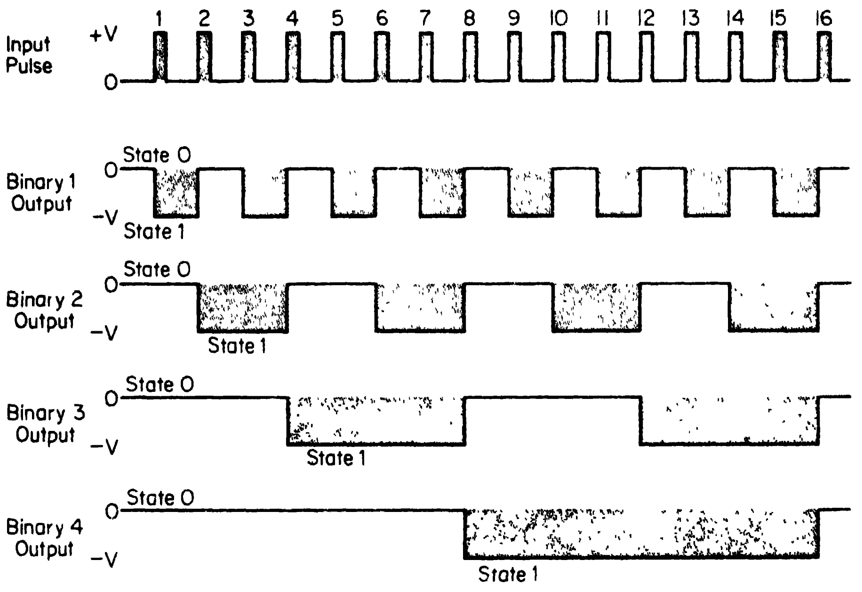


Figure 13-3
Waveform chart of a four-binary counter.

sixteen successive positive trigger pulses applied to the input of the first binary. The first input pulse causes a transition of binary B_1 from state 0 to state 1. As a result of this transition a negative-going voltage appears at the output of B_1. This negative pulse does *not* cause a transition in the next stage because of the action of diode CR_1.

The second input pulse causes the first binary B_1 to return to its original 0 state and a positive-going voltage appears at the output terminal, $\bar{Y}$. This output pulse triggers the following binary (B_2) and it makes a transition from state 0 to state 1. Its output voltage is again a negative-going pulse which does not affect the next binary (B_3). After two input pulses have been applied, therefore, flip-flop B_1 is in state 0 and flip-flop B_2 is in state 1.

The third input pulse causes B_1 to go to its 1 state, causing a negative pulse at its output. A negative pulse into binary B_2 does not affect this flip-flop and it remains in its 1 state. This process can be carried through all of its sixteen steps and the waveform chart of Fig. 13-3 will be found to be correct.

Summarizing the entire operation of the cascaded flip-flops:

(1) Flip-flop B_1 makes a transition at each applied trigger pulse.
(2) Each of the other flip-flops makes a transition only when the preceding flip-flop switches from state 1 to state 0.

Table 13-1 lists the states of the four binaries in the chain as a function of the number of input pulses. This table corresponds exactly to the waveform chart of Fig. 13-3. Notice that the order in which the binaries appear in Table 13-1 is *opposite* to the order in which they are drawn in Fig. 13-2. The ordered array of states 0 and 1 in any row of Table 13-1 is the binary representation of

TABLE 13-1
STATES OF THE BINARIES

Number of Input Pulses	State of Binary B_4	B_3	B_2	B_1
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	0	0
5	0	1	0	1
6	0	1	1	0
7	0	1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1
16	0	0	0	0

the number of input pulses. We can therefore say that the chain of four flip-flops of Fig. 13-2 counts in the binary system of numbers. The chain will reset itself to the 0 state with the application of the sixteenth pulse and the counting cycle starts again with the seventeenth input trigger.

A chain of 4 binaries will count up to the number $2^4 = 16$ before it resets itself. It is clear that the counting range may be extended by adding more binaries: a chain of 5 flip-flops counts up to the number $2^5 = 32$ and n binaries, therefore, count to 2^n.

The decimal system, with which we are more familiar, counts to the base of ten. To construct a counter which provides one output pulse for every ten input triggers we start with a cascaded arrangement of four flip-flops. Feedback is introduced from the later stages to the earlier stages so that the count is advanced by six at some time during the first ten counts. This advance of six counts may be made in one step or in several stages. In fact, many different circuit arrangements are possible.

The circuit of Fig. 13-4 shows a *scale-of-16 counter* (four flip-flops in cascade) which is modified by feedback into a *scale-of-10* counter. This circuit is commonly used in commercially available decade counters and deserves some attention at this point. The feedback connections are shown in the logic diagram of Fig. 13-4(a) by the heavy lines, and the waveform chart is given in Fig. 13-4(b). Inspection of the waveform chart shows that the count proceeds normally for the first three trigger pulses. After the third trigger has been received, flip-flops B_1 and B_2 are in the 1 state, which is quite normal. When the fourth input pulse arrives, a series of events takes place: Binary B_1 and binary B_2 go from the 1 state to the 0 state and binary B_3 goes from the 0 state to the 1 state.

At this point, the Y-output terminal of binary B_3 gives a positive transition which is coupled to the "set" terminal of binary B_2. This connection is indicated by the heavy line in Fig. 13-4. The set terminal of the flip-flop is connected to the base of the Q_1 transistor. The positive pulse provided by the feedback line from B_3 forces transistor Q_1 of binary B_2 to cutoff. When Q_1 is cut off, Q_2 conducts, and the binary is in the 1 state. Therefore, the feedback from the Y terminal of B_3 to the S terminal of B_2 causes a transition of B_2 from state 0 to state 1. At this point, the binary chain reads (counting from B_4 to B_1) $0110 = 6$, so that the counter has been advanced by two counts through feedback. The fifth input trigger puts B_1 back to state 1, but provides no further output pulses for the following binaries who stay in their respective states. The sixth trigger pulse starts a new sequence of events. Firstly, binary B_1 returns to the 0 state. This transition generates an output pulse which returns binary B_2 to the 0 state. This in turn generates an output pulse which moves binary B_3 to the 0 state.

Incidentally, the feedback loop from B_3 to B_2 receives a negative pulse at this instant which is of no consequence. The transition of binary B_3 to the 0 state produces an output pulse which moves the last binary B_4 to the 1 state. Here again a feedback loop is connected to the Y terminal of B_4, indicated by a heavy line in the diagram. The Y terminal of B_4 is connected to the "set" terminal of binary B_3. The transition of B_4 produces a positive pulse at the Y

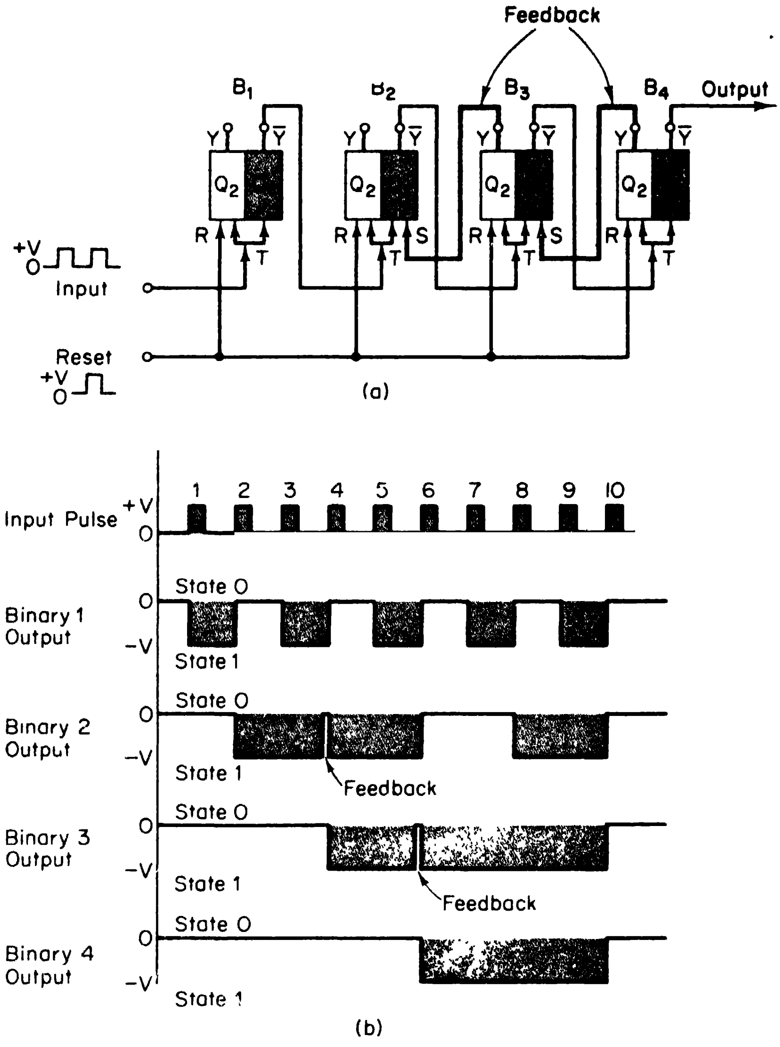


Figure 13-4
(a) A scale-of-16 counter, modified by feedback into a scale-of-10 counter. (b) Waveform chart of the feedback counter.

output which immediately returns binary B_3 to the 1 state. After receiving the sixth input pulse, the situation is then as follows: Binaries B_1 and B_2 are in the 0 state and binaries B_3 and B_4 are in the 1 state. Again reading from B_4 to B_1, we find that this corresponds to the binary code 1100 = 12 decimal. The count of the binary chain has therefore been advanced another four steps for a total advance of six counts. Inspection of the waveform chart in Fig. 13-4(b) shows that the next four pulses simply bring the count up to a total of ten, at which time the binaries all reset to their initial 0 state and the counting process can start all over again.

The net result of this exercise has been that 1 output pulse has been generated at the $\bar{Y}$ terminal of the last binary after the application of ten input pulses. This output pulse may serve as a "carry" pulse to a following decade counter assembly (DCA). Notice the difference between a regular input pulse and a reset pulse: A regular input pulse is positive and causes a conducting transistor to cut off; a reset pulse is negative and causes a cut-off transistor to conduct.

The DCA usually incorporates a digital display system to indicate the state of each individual binary in the chain. A simple indicator which may be used for this purpose is a neon bulb in series with a resistor. This indicator is then connected in the collector circuit of the Y transistor of each binary. When a binary is in the 1 state, the Y transistor will be conducting heavily and the neon lamp lights up. The neon lamps connected to binaries B_1, B_2, B_3, and B_4 are assigned values of 1, 2, 4, and 8, respectively. To determine the count of the DCA it is necessary only to add the numbers assigned to those neon lamps that are lit.

A much more elegant arrangement involves an electrical readout of the DCA plus a *digital display*. The electrical readout in this case consists of a four-line binary-coded-decimal output voltage, where a voltage representing the state of each binary in the DCA is taken from the collector of each of the Y transistors. A binary 1 is then represented by a relatively positive voltage on each line and a binary 0 is represented by a relatively negative voltage on each line. The ten allowable combinations are summarized in Table 13-2,

TABLE 13-2

FOUR-LINE CODE TRUTH TABLE

git	Four-line Code			
	B_4	B_3	B_2	B
0	0	0	0	0
1	0	0	0	1
2	0	0	1	0
3	0	0	1	1
4	0	1	1	0
5	0	1	1	1
6	1	1	0	0
7	1	1	0	1
8	1	1	1	0
9	1	1	1	1

representing the digits 0 through 9. This binary-coded electrical readout is available for use elsewhere; for example, it may be used for recording on magnetic tape. In most cases the electrical readout is used to produce a digital display.

A typical application is given in the simplified circuit diagram of Fig.

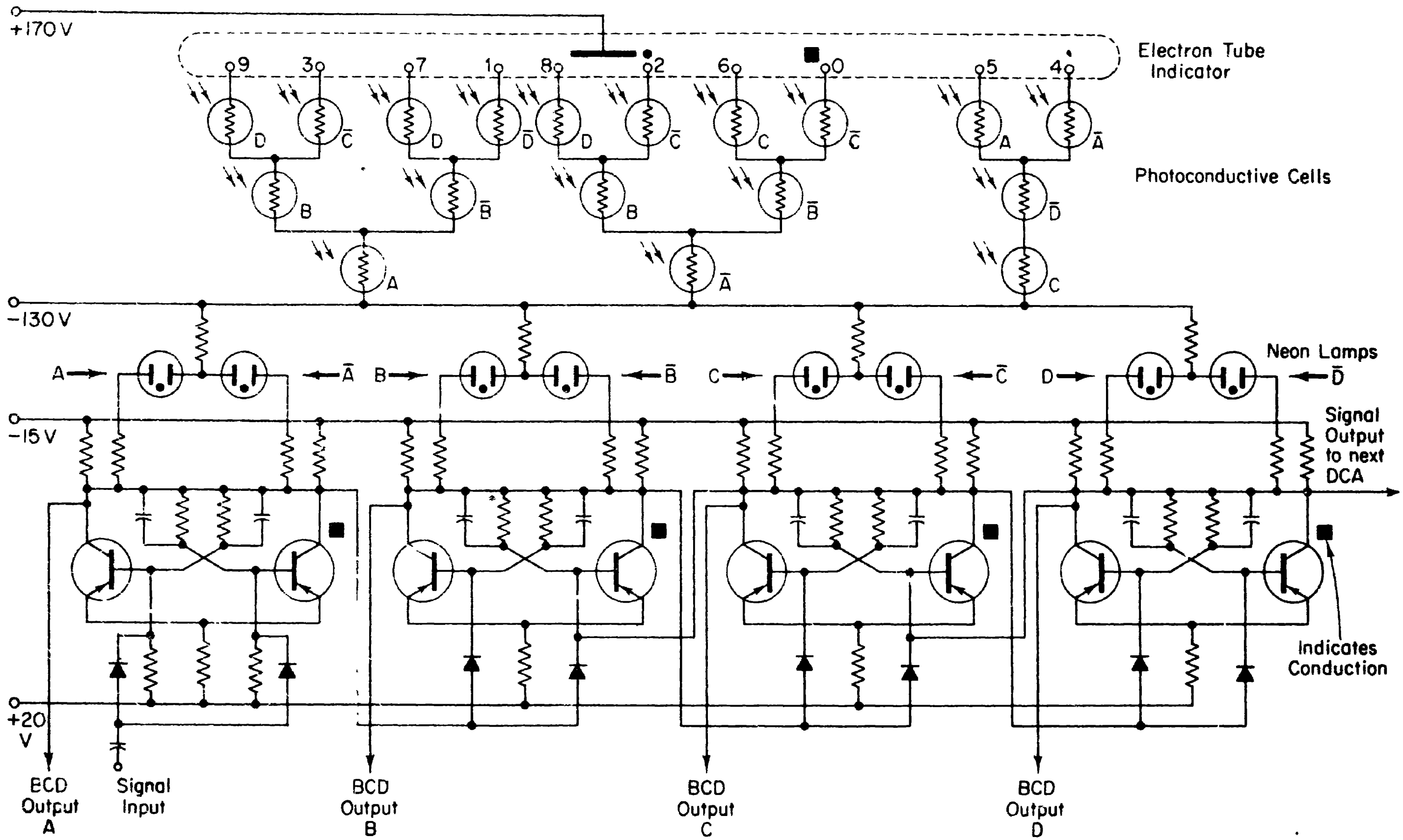


Figure 13-5. Simplified schematic diagram of a decimal counter with binary-coded-decimal outputs and electron-tube display of the digits. (*Courtesy Hewlett-Packard Co.*)

13-5, which shows the display matrix and the decimal-counter schematics in their basic form. The *display matrix,* consisting of eight neon input lamps and eighteen photoconductive elements is used to convert the binary-coded representation to a digital representation. As indicated in the schematic diagram, the circuit to each numeral in the ten-digit electron tube indicator is brought through three series-connected photocell elements. A characteristic* of the photocell element is that it has a high resistance (several MΩ) when dark and a relatively low resistance (less than 7,000 Ω) when illuminated. Thus when the three photocell elements which constitute a circuit path are illuminated, the resistance in the display path drops to about 20,000 Ω and sufficient current is available to light the display digit.

Illuminating elements for the photocells are neon lamps, which are connected in the collector circuit of each of the eight transistors in the counting circuit. The lamp lights when the transistor conducts. As explained earlier in this section, a four-binary counting circuit has ten states, ten combinations of conducting and nonconducting transistors each combination corresponding to one digit. There is therefore a pattern of lighted neon lamps for each digit. Assigning a binary weight of 1 when the plain-letter lamp (A, B, C, or D) lights and a weight of 0 when the bar-letter lamp ($\bar{A}$, $\bar{B}$, $\bar{C}$, or $\bar{D}$) lights, the lamp pattern for any digit can be determined from Table 13-2. Figure 13-5 shows the counting circuit with transistors $\bar{A}$, $\bar{B}$, $\bar{C}$, and $\bar{D}$ conducting. The lamps associated with these circuits illuminate the photocell elements in the circuit to the digit 0 display. It is clear that the decade-counting assembly serves a *dual* purpose. It *counts* and *displays* from 0 to 9 input pulses. On the arrival of the tenth input pulse, it resets and displays a zero, but it also produces an output pulse. It therefore *divides* by a factor of ten and may then be called a *decade divider*.

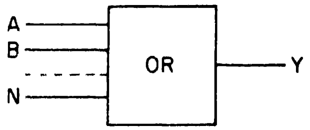
Both functions of the DCA are used in very many instrumental applications. Consider, for example, the case where several DCAs, each with its own display unit, are connected in cascade. Upon application of trigger pulses to the input of the first DCA, its display unit will portray the count of the units from 0 to 9. Its output pulse, which occurs after every tenth input pulse, fires the second DCA, which therefore displays the count of the tens. The third DCA is triggered by every tenth output pulse from the second DCA and therefore portrays the count of the hundreds. The total display is limited only by the number of DCAs available.

13-3 The Control Gate

The *control gate* is usually a part of the *function-control* assembly which switches the counter logic circuits to perform the various counting operations.

Since the counter contains, in addition to the main control gate, a number of additional gating circuits for logic control, we shall briefly review some of the principles of logic gates.

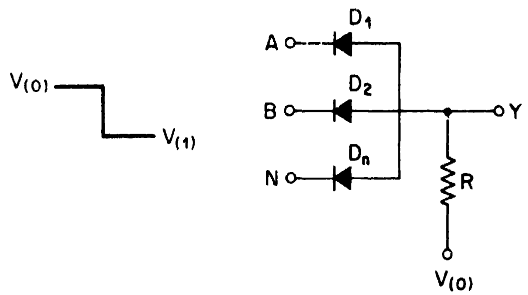
The OR gate is a multiple input circuit with a single output, as indicated by the logic symbol in Fig. 13-6(a). The circuit obeys the following definition:



(a)

Input		Output
A	B	Y
0	0	0
0	1	1
1	0	1
1	1	1

(b)



(c)

Figure 13-6
(a) The IEEE standard symbol for an OR gate. (b) The truth table for a two-input OR gate. (c) A negative-logic OR gate.

The output of an OR *assumes the* 1 *state if one or more inputs assume the* 1 *state.* The *n* inputs to a logic circuit are usually designated by A, B, . . . N and the output by Y, as shown in Fig. 13-6. The operation of an OR gate can then be described by: Y equals A *or* B *or* C . . . *or* N.

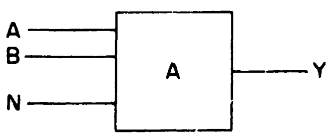
It is usual practice to write a truth table for the logic circuit, tabulating all the possible inputs and their corresponding outputs. This is done in Fig. 13-6(b) for the OR gate. The actual circuitry contains semiconductor diodes in an arrangement as shown in Fig. 13-6(c). The circuit shown here represents *negative* logic, where the voltage for the 1 state is negative with respect to the voltage for the 0 state. If all the inputs are in the 0 state, the voltage across each diode is equal to $V_{(0)} - V_{(0)}$. In order for a diode to conduct, it must be forward biased, and in our case none of the diodes will then conduct. Therefore the output voltage equals $V_{(0)}$, and Y is in the 0 state. If now input A is changed to the 1 state, diode D_1 will conduct because it now is forward biased and the output voltage becomes

$$V_{(0)} - [V_{(0)} - V_{(1)}] \approx V_{(1)} \quad \text{(by approximation)}$$

Y is therefore in the 1 state, and each diode, except D_1, is back biased. If two or more inputs are in the 1 state, then the diodes connected to these inputs conduct and all other diodes remain reverse biased. If for some reason the voltage level $V_{(1)}$ is not identical for all inputs, then the most negative value of $V_{(1)}$ will appear at the output and all diodes except that one are nonconducting.

The AND gate has two or more inputs and a single output and it operates on the following condition: *The output of an* AND *gate assumes the* 1 *state only if all the inputs assume the* 1 *state.* The logic symbol for the AND gate is given in Fig. 13-7(a), accompanied by the two-input truth table in Fig. 13-7(b). The operation of the AND gate can be described by the expression: Y equals A *and* B *and* C *and . . . and* N.

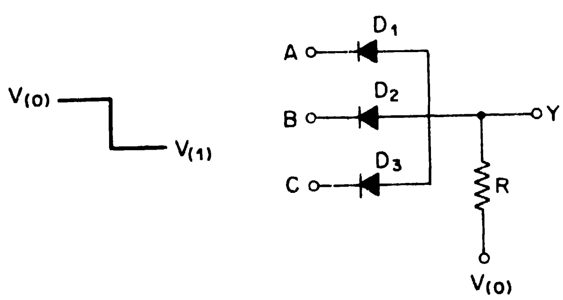
The diode logic circuit for a negative AND gate is given in Fig. 13-7c. To understand the operation of the AND circuit assume that the diodes are ideal (zero forward resistance and infinite reverse resistance) and that the inter-



(a)

Input		Output
A	B	Y
0	0	0
0	1	0
1	0	0
1	1	1

(b)



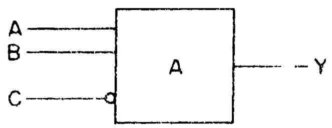
(c)

Figure 13-7

(a) The IEEE standard for an AND gate. (b) Truth table for a two-input AND gate. (c) A negative-logic AND circuit.

nal resistances of the input voltage sources are zero. If any input to the AND gate is at the 0 level $V_{(0)}$, the diode connected to that input is *forward* biased and *conducts*, clamping the output voltage at $V_{(0)}$. Therefore Y is at 0. If all the inputs are in the 1 state at the same time, all the diodes are *reverse* biased and the output voltage equals $V_{(1)}$ or Y is at 1. The AND gate is often called a *coincidence* gate.

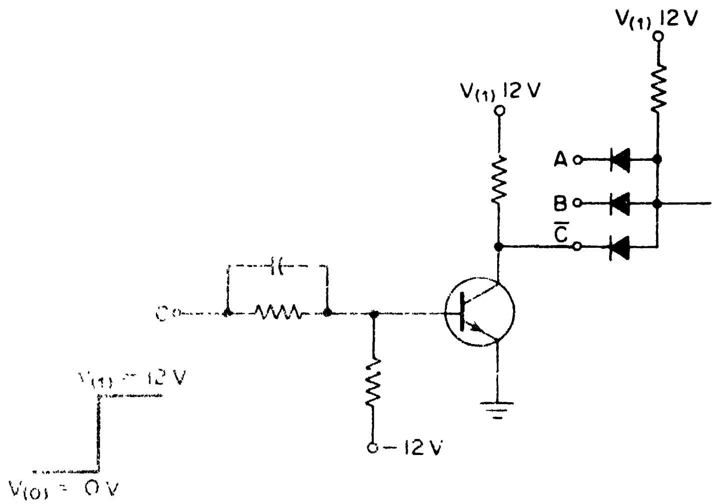
The INHIBIT gate is a modification of the AND gate. Here one input



(a)

Input			Output
A	B	C	Y
0	0	0	0
0	1	0	0
1	0	0	0
1	1	0	1
0	0	1	0
0	1	1	0
1	0	1	0
1	1	1	0

(b)



(c)

Figure 13-8

(a) The logic symbol for an INHIBIT gate. (b) The truth table for one INHIBIT and two normal terminals. (c) Circuit diagram of a positive-logic INHIBIT circuit.

terminal is preceded by an *inverter,* shown in the circuit diagram of Fig. 13-8(c). This modified AND circuit operates on the following principle: The output of an INHIBIT gate assumes the 1 state if all the inputs, except the *negation* input, assume the 1 state. The logic symbol and the truth table are also given in Fig. 13-8. The truth table is valid for a three-input AND gate with one inhibitor terminal (C).

The circuit operates as follows: If either of the inputs A or B, or both, are in the 0 state where $V_{(0)} = 0$ V, then at least one of the diodes conducts and the output is clamped to 0 V. If a *coincidence* occurs, so that *both* inputs A and B are in the 1 state [$V_{(1)} = +12$ V], then both diodes D_1 and D_2 are reverse biased and nonconducting. If at the same time input C is also at the 1 level of +12 V, transistor Q in the inverter circuit will be conducting, and the negation terminal $\bar{C}$ will be at the 0 level of 0 V. This forward biases diode D_3 which clamps the output at 0 V and output terminal Y is in the 0 state. On the other hand, if input terminal C is in the 0 state, $\bar{C}$ is at +12 V and diode D_3 is back biased. The output at the Y terminal will then be at +12 V and Y is in the 1 state. The truth table in Fig. 13-8 verifies the anti-coincidence conditions of the INHIBIT circuit. An INHIBIT gate with only two input terminals normally passes the input signal. Adding a second input signal of the same logic level prevents the signal from passing through the gate.

Now that the operation of various types of gate circuits is understood, their function in the different parts of an electronic counter can be more clearly explained.

13-4 The Time Base and Associated Circuitry

The time base of an electronic counter invariably consists of a crystal oscillator and a decade divider assembly which reduces the output frequency of the oscillator in steps of a factor of ten. The crystal which controls the frequency of the oscillator is usually kept in a temperature-controlled environment (oven) to increase the frequency stability of the instrument. A functional block diagram of a time-base generator for a general-purpose electronic counter is shown in Fig. 13-9.

Figure 13-9 shows three major blocks. The first block is the *crystal oven assembly,* which consists of a thermally insulated chamber containing a heating element, a temperature-sensing circuit, and a piezoelectric crystal.

The *oven control assembly* includes an LF oscillator whose amplitude is controlled by the temperature-sensing element in the crystal oven. This temperature-sensing element is a thermistor in a bridge configuration, whose change in resistance due to temperature variations upsets the balance of the

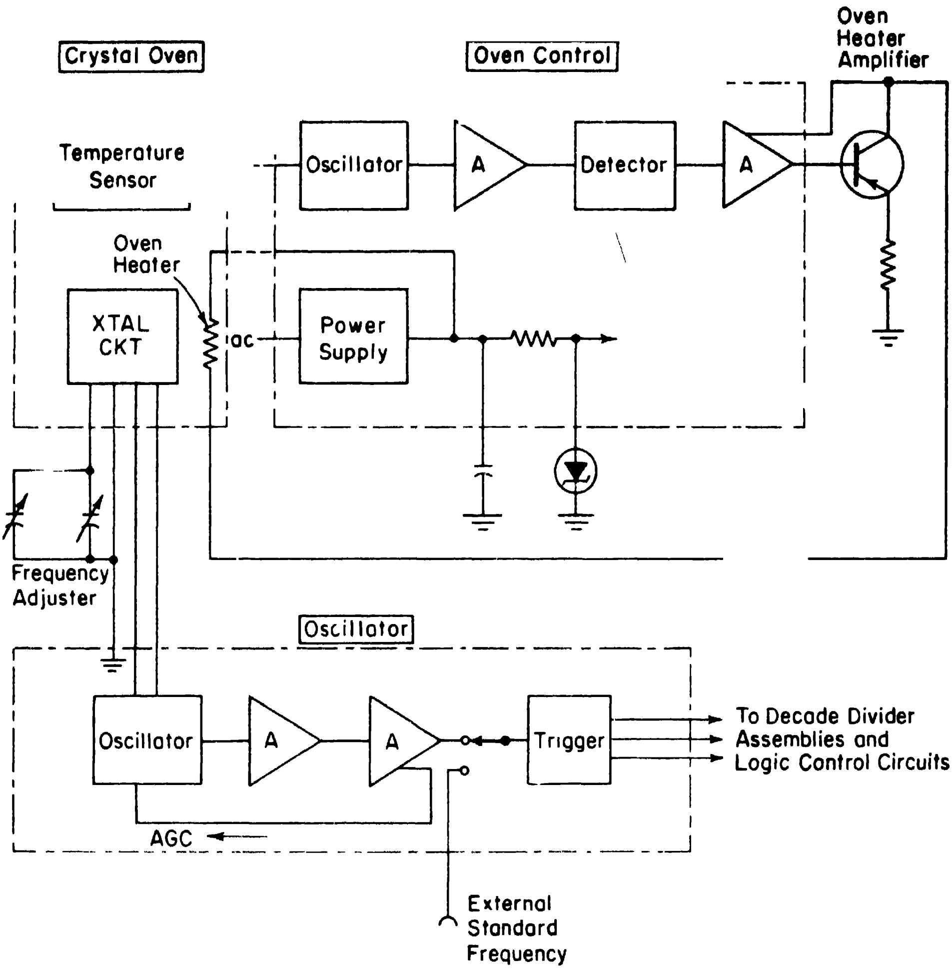


Figure 13-9
Functional block diagram of the crystal oven, the oven-control circuit, and the oscillator of a general-purpose electronic counter.

bridge circuit. The oscillator output is amplified and detected to produce a dc level whose amplitude is inversely proportional to the oven temperature. This dc level is amplified and applied to the heating element in the oven.

The *oscillator assembly* includes the oscillator circuit which is connected to the crystal in the oven. The oscillator output is amplified and made available to be supplied to the decade divider assemblies (DDAs) for reduction by factors of ten. The decade divider assembly operates identically to the decade counter assembly described in Sec. 13-2, except that there is no displayed count.

13-5 Frequency Measurements

Frequency can be accurately measured by counting the number of cycles of the unknown signal for a precisely controlled time interval. Figure 13-10 shows the logic block diagram for a counter in the *frequency* mode of operation.

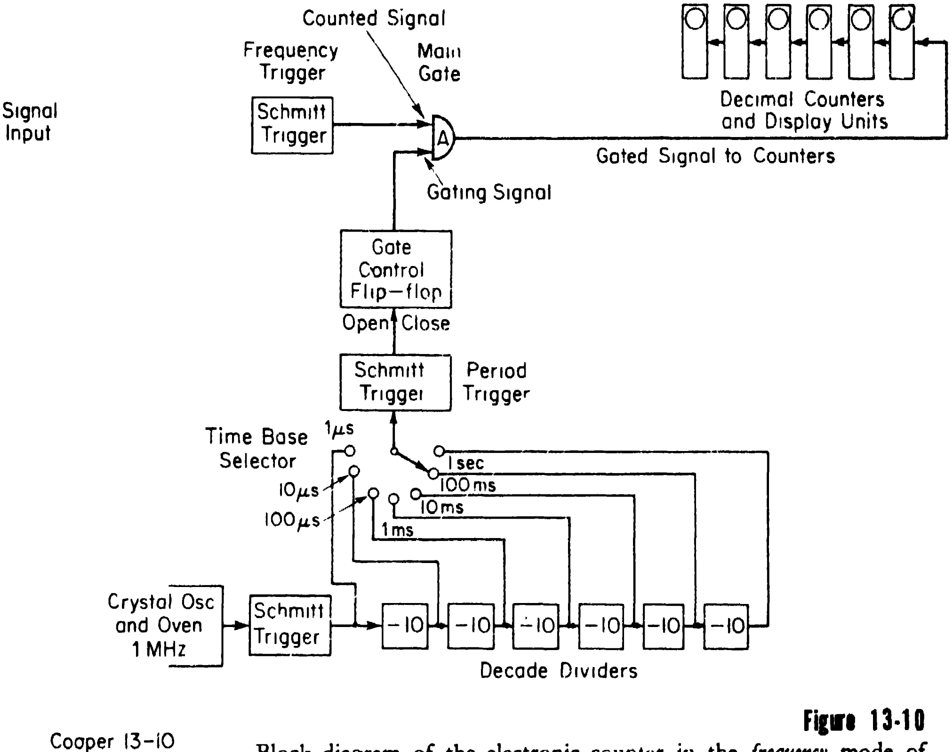


Figure 13-10
Block diagram of the electronic counter in the *frequency* mode of operation.

There are two signals that need to be traced: the *input signal* (or measured frequency) and the *gating signal,* which determines the length of time during which the DCAs are allowed to accumulate pulses. The input signal is amplified and passed to a Schmitt trigger. Here it is converted into a square wave with very fast rise and fall times, then differentiated and clipped. As a result, the signal which arrives at the input to the main gate consists of a series of pulses separated by the period of the original input signal. In the block diagram of Fig. 13-10, the oscillator frequency is 1 MHz. The time-base output is shaped by a Schmitt trigger, so that positive spikes, 1 μs apart, are applied to a number of decade dividers. In the example shown, six DDAs are used whose outputs are connected to a time-base selector switch. This allows the time interval to be selected from 1 μs to 1 s. The first output pulse from the time-base selector switch passes through a Schmitt trigger to the gate control flip-flop. The gate flip-flop assumes a state such that an *enable* signal is applied to the main gate. Since this is an AND gate, the input signal pulses are allowed to enter the DCAs and they are totalized and displayed. This continues until the second pulse from the DDAs arrives at the control flip-flop. The gate control assumes the other state which removes the enable signal from the main gate. The main gate closes and no further pulses are admitted to the DCAs. The DCA display is now in a state which corresponds to the number of input pulses received during a precise time interval which was determined by the time base.

Since frequency is defined as the number of occurrences of a particular phenomenon in some length of time, the counter display corresponds to *frequency*. Usually the time-base selector switch moves the decimal point in the display area, allowing the frequency to be read directly in Hz, kHz, or MHz.

13-6 Period Measurement

It is often desirable to measure the *period* of a signal rather than its frequency. The *same* building blocks used for the frequency measurement can be rearranged so that the counted signal and the gating signal are reversed. Figure 13-11 shows the block diagram of the period measurement using the same counter components as in the previous application. The gating signal is derived from the unknown input signal which now controls the opening and closing of the main gate. The precisely spaced pulses from the crystal oscillator are counted for one period of the unknown frequency. In the example shown in Fig. 13-11, the time base is set to 10 μs (100-kHz time-base frequency), and the number of pulses which occur during one period of the unknown signal are counted and displayed by the DCAs.

The accuracy of the period measurement may be increased greatly by using the multiple-period average mode of operation. This type of measurement is similar to the single-period measurement in that the gating signal is derived from the unknown input signal and the counted signal from the time-base oscillator. The basic difference is that the main gate is held open for more than one period of the unknown signal. This is accomplished by passing the unknown signal through one or more DDAs so that the period is extended by a factor of 10, 100, or more.

Figure 13-11 shows the multiple-period average mode of operation as a modification of the single-period measurement by the dashed portion of the block diagram. Notice that in Fig. 13-11 the 1-MHz crystal frequency is

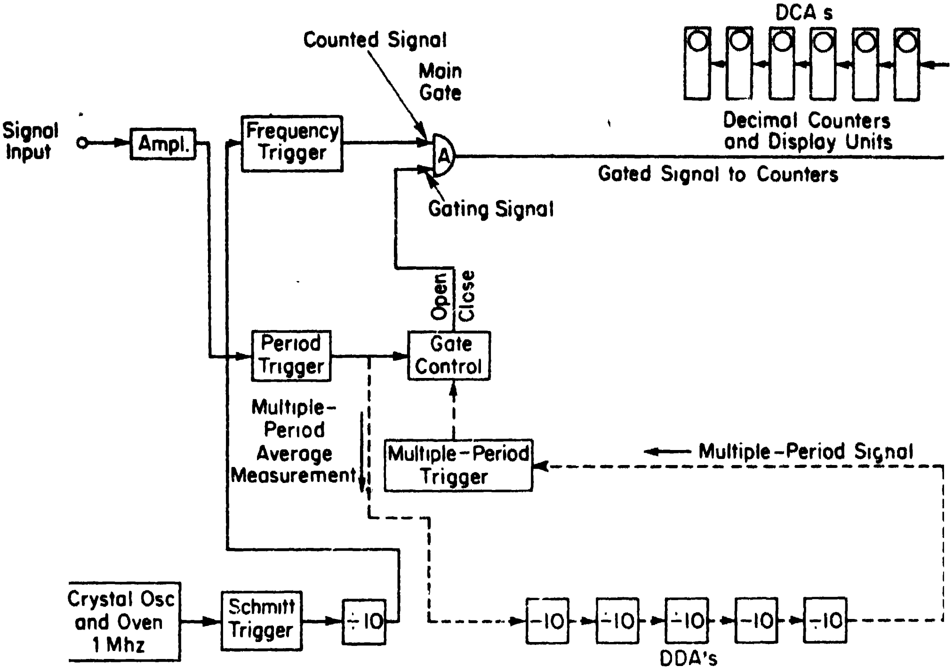


Figure 13-11
Block diagram of the *single period* and *multiple period average* measurement mode of operation.

divided by 1 DDA to a frequency of 100 kHz (10 μs period). These clock pulses are shaped by the frequency trigger and fed to the main gate to be counted. The input signal whose period is to be measured, is amplified, shaped by the period trigger, and fed to 5 DDAs in cascade counting the input frequency down by a factor of 10^5. This divided signal is now shaped by the multiple-period trigger (another Schmitt trigger circuit) and applied to the gate control flip-flop. The gate control provides the enable pulse and the stop pulse for the main gate. Obviously, the main gate will remain open for a greatly increased time interval, in fact increased by a factor of 10^5. The DCAs will, therefore, count the number of 10-μs intervals which occur during 100,000 periods of the input signal. The readout logic is so designed that the decimal point will be automatically positioned to display the proper units.

13-7 Ratio and Multiple-Ratio Measurements

A ratio measurement is, in effect, a period measurement with the *lower* of the two frequencies used as the *gating* signal and the *higher*-frequency signal

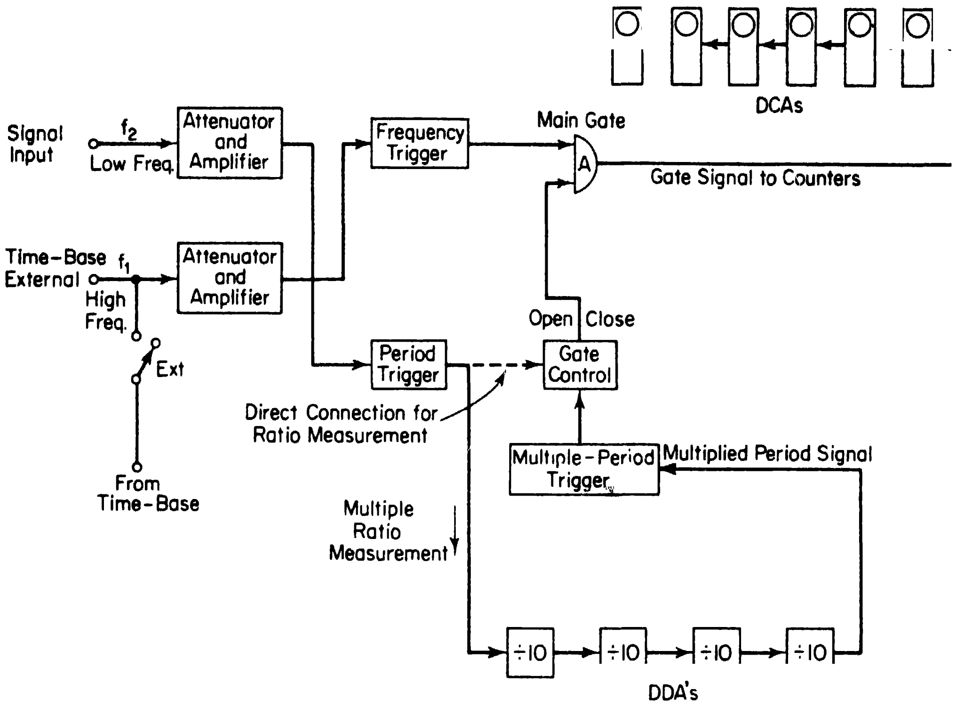


Figure 13-12
Block diagram of the *ratio* and *multiple ratio* mode of operation.

as the *counted* signal. In other words, the lower-frequency signal takes the place of the time base. The block diagram of Fig. 13-12 applies. The number of cycles of the higher-frequency signal f_1 which occur during a period of the lower-frequency signal f_2 are counted and displayed in the DCAs. A multiple-ratio measurement extends the number of periods of the lower-frequency signal by a factor of 10, 100, etc. Note that the standard selector is in the *external* position and f_1 takes the place of the internal oscillator.

13-8 Time-Interval Measurements

Time-interval measurements can be made with the same basic blocks as ratio measurements. This measurement is very useful in determining the *pulse-width* of a certain waveform.

The block diagram for this measurement is given in Fig. 13-13. This configuration shows two parallel input signal channels, where one channel supplies the *enabling* pulse for the main gate and the other channel supplies the

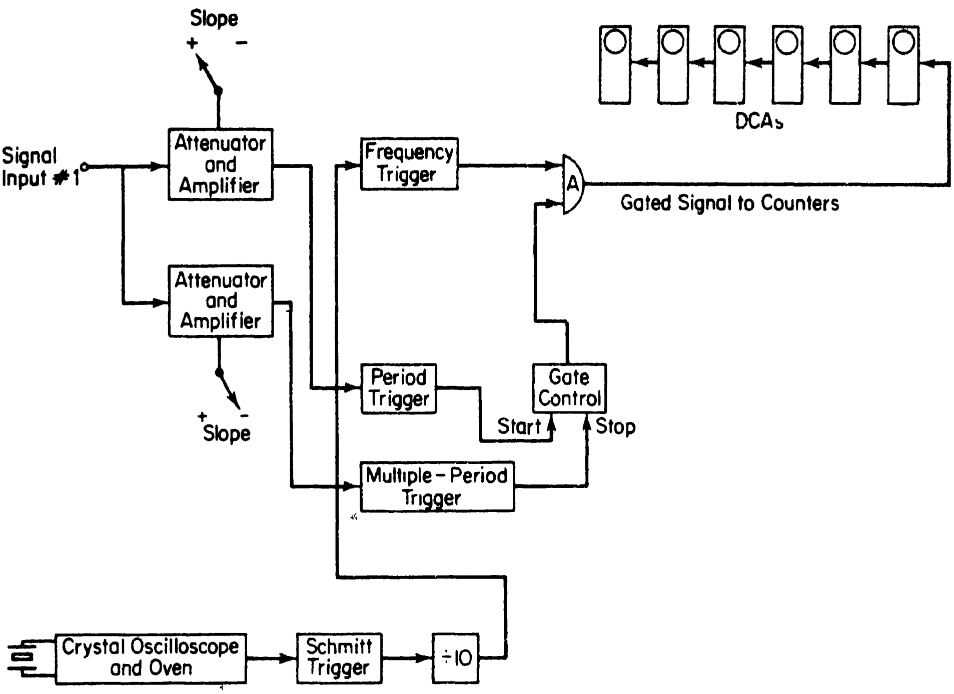


Figure 13-13
Block diagram of the *time interval* mode of operation.

disabling pulse for the same gate. The main gate is *enabled* at a point on the *leading* edge of the input signal waveform and *closed* at a point on the *trailing* edge of the same waveform. The counter must then have a *slope-selection* feature, as indicated in the block diagram. The *trigger level* control permits selection of the point on the incoming signal waveform at which the measurement begins and ends.

13-9 The Universal Counter-Timer

The measurements described briefly in the preceding sections all use the *same* fundamental building blocks, which together, form the modern *universal counter*. The type of measurement to be performed dictates how the different building blocks must be interconnected. Modern solid-state universal counters use *diode logic* to interconnect the various elements of the system. These diode logic elements consist of different types of gates (AND, OR, NOR, INHIBIT, etc.) which are selected and controlled by a single front-panel switch, called the *function* switch.

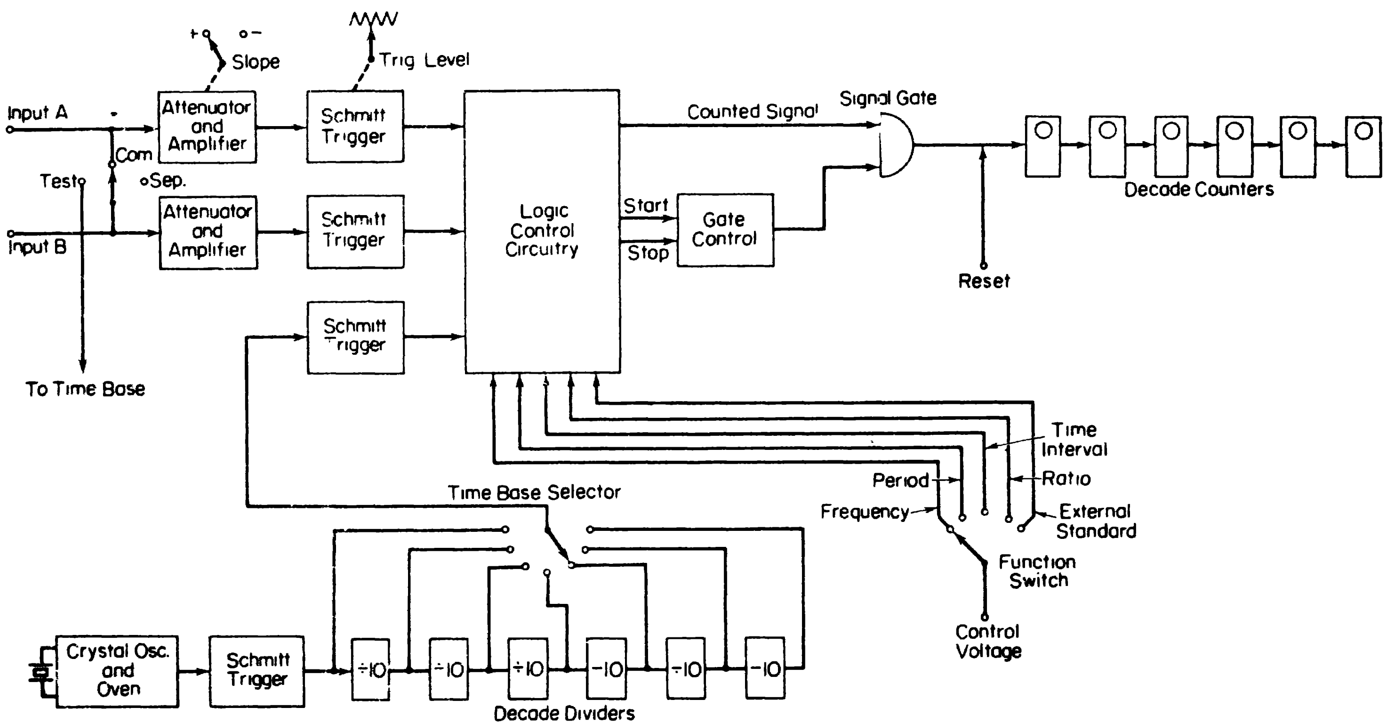


Figure 13-14 (a). Block diagram of a Universal Counter.

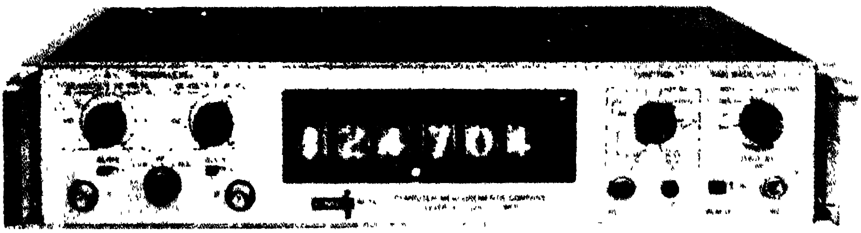


Figure 13-14 (b)
A Universal Counter-timer. (*Courtesy Computer Measurements Co.*)

The simplified block diagram of Fig. 13-14(a) shows the elements of a universal counter; Fig. 13-14(b) is a photograph of this instrument. With the function switch in the *frequency* mode, as shown on the diagram, a control voltage is applied to specific gates in the logic control circuitry. The gates in this circuit allow the *input signal* to be connected to the counted-signal channel of the main gate. The selected output from the time-base dividers is simultaneously gated to the control flip-flop, which enables and disables the main gate. Both control paths are *latched internally* to allow them to operate only in the proper sequence.

With the function switch in the *period* mode, the control voltage is connected to certain gates in the logic control circuitry, which connect the *time-base signal* to the counted-signal channel of the main gate and also the signal input to the gate control for enabling and disabling the main gate. The remaining function switch positions perform similar control functions in the logic control circuitry. Exact details of switching and control procedures vary from instrument to instrument. (See the instruction manual of the counter available.)

13-10 Measurement Errors

Frequency and time measurements made by an electronic counter are subject to several inaccuracies inherent in the instrument itself. Intelligent use of the counter for a given application requires an understanding of the *limitations* of the instrument.

One very common instrumental error resulting from counter measurement is the *gating error,* present whenever frequency and period measurements are made. Recall that, for a frequency measurement (Fig. 13-10), the main gate (also called the *signal* gate) is opened and closed by the oscillator output pulse. This allows the input signal to pass through the gate and be counted by the DCAs. The gating time is not synchronized with the input signal; they are, in fact, two totally unrelated signals.

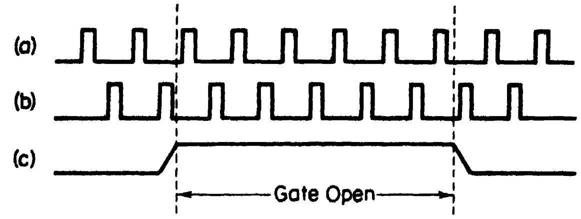


Figure 13-15 Illustrating the gating error.

In Fig. 13-15 the gating interval is indicated by waveform (c). Waveforms (a) and (b) represent the input signal in different phase relationships with respect to the gating signal. Clearly, in one case, six pulses will be counted; in the other case, only five pulses are allowed to pass through the gate. We have therefore a ± 1 count ambiguity in the measurement. When measuring *low frequencies,* the gating error may have an appreciable effect on the results. Take, for example, the case where a frequency of 10 Hz is to be measured and the gating time equals 1 s (a reasonable assumption). The decade counters would indicate a count of 10 ± 1 count, an inaccuracy of 10 per cent. *Period* measurements are therefore, to be preferred over frequency measurements at the *lower* frequencies.

The dividing line between frequency and period measurements may be determined as follows: Let

f_c = crystal (or clock) frequency of the instrument
f_x = frequency of the unknown input signal

In a *period* measurement, the number of pulses counted equals

$$N_p = \frac{f_c}{f_x} \tag{13-1}$$

In a *frequency* measurement with a 1-s gate time, the number of pulses counted is

$$N_f = f_x. \tag{13-2}$$

The *crossover frequency* (f_o) at which $N_p = N_f$ is

$$\frac{f_c}{f_o} = f_o \quad \text{or} \quad f_o = \sqrt{f_c} \tag{13-3}$$

Signals with a frequency *lower* than f_o should therefore be measured in the *period* mode, signals of frequencies *above* f_o should be measured in the *frequency* mode in order to minimize the effect of the ± 1 count gating error. The accuracy degradation at f_o caused by the ± 1 count gating error is $100/\sqrt{f_c}$ per cent.

Inaccuracies in the time base also cause errors in the measurement. In frequency measurements, the time base determines the opening and closing of the signal gate, and provides the pulses to be counted. *Time-base errors* consist of oscillator calibration errors, short-term crystal stability errors, and long-term crystal stability errors.

Several methods of *crystal calibration* are in common use. One of the simplest calibration techniques is to zero-beat the crystal oscillator against the standard frequency transmitted by a standards radio station, such as WWV (see Sec. 3-3). This method gives reliable results with a usual accuracy in the order of 1 part in 10^6, which corresponds to 1 cycle of a 1-MHz crystal oscillator. If the zero-beating is done with visual (rather than audible) means, for example, by using a CRO, the calibration accuracy can usually be improved to 1 part in 10^7.

Several very low-frequency (VLF) radio stations are presently covering the North American continent with precise signals in the 16–20-kHz range. Low-frequency receivers are available with automatic servo-controlled tuning which can be slaved to the signal of one of these stations. The error between the local crystal oscillator and the incoming signal can then be recorded on a strip-chart recorder. A simplified diagram of this procedure is given in Fig. 13-16. Improved calibration accuracy is obtainable by using VLF stations

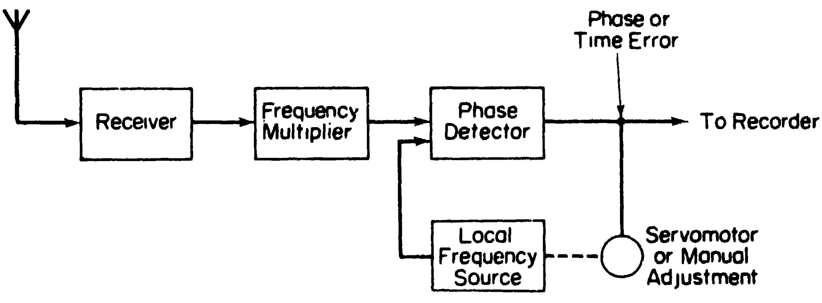


Figure 13-16
Calibration of a local frequency source.

rather than HF stations because the transmission paths for very low frequencies is shorter than for high-frequency transmissions.

Short-term crystal stability errors are caused by momentary frequency variations due to voltage transients, shock and vibration, cycling of the crystal oven, electrical interference, etc. These errors can be *minimized* by taking frequency measurements over *long* gate times (10 s to 100 s) and *multiple-period* average measurements. A reasonable figure for short-term stability of a standard crystal-oven combination is in the order of 1 or 2 parts in 10^7.

Long-term stability errors are the more subtle contributors to the inaccuracy of a frequency or time measurement. Long-term stability is a function of aging and deterioration of the crystal. As the crystal is temperature-cycled and kept in continuous oscillation, internal stresses induced during manufacture are relieved and minute particles adhering to the surface are shed reducing its thickness. Generally, these phenomena will cause an *increase* in the oscillator frequency.

A typical curve of frequency change versus time is shown in Fig. 13-17.

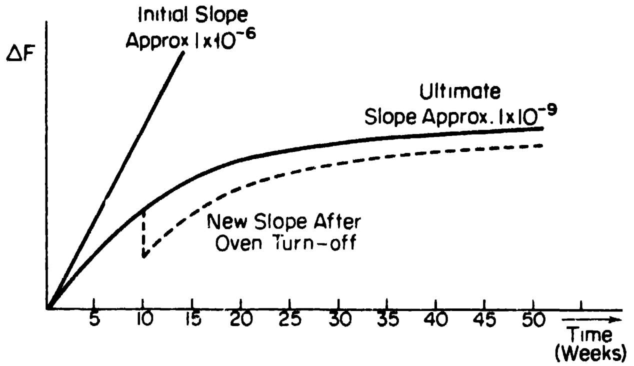


Figure 13-17
Frequency change versus time for an oven-controlled crystal.

The *initial* rate of change of crystal frequency may be in the order of 1 part in 10^6 per day. This rate will decrease, provided that the crystal is maintained at its operating temperature, normally about 50° to 60°C, with ultimate stabilities of 1 part in 10^9. If, however, the instrument containing the crystal is unplugged from the power source for a period of time sufficient to allow the crystal to cool appreciably, a new slope of aging will ensue when the instrument is put back into operation. It is possible that the actual frequency of oscillation after cool-off will vary by several cycles and that the original frequency will not again be reached unless calibration is done.

To show the effect of long-term stability on the absolute accuracy of the measurement, assume that the oscillator was calibrated to within 1 part in 10^9 and that a long-term stability of 1 part in 10^8 per day was reached. Assume further that calibration was done sixty days ago. The guaranteed accuracy at this time is then $1 \times 10^{-9} + 60 \times 10^{-8} = 6.01 \times 10^{-7}$, or 6 parts in 10^7. It can be seen therefore that maximum absolute accuracy can be achieved only if an exact calibration is performed a relatively short time *before* the measurement is taken.

In time-interval and period measurements the signal gate is opened and closed by the input signal. The accuracy with which the gate is opened and closed is a function of the *trigger-level error*. In the usual application the input signal is amplified and shaped and then applied to a Schmitt trigger circuit which supplies the gate with its control pulses. Usually the input signal contains a certain amount of unwanted components or noise, which is amplified along with the signal. The time at which triggering of the Schmitt circuit occurs, is a function of the input signal amplification and of its signal-to-noise ratio. In general, we can say that trigger-time errors are reduced with large signal amplitudes and fast rise times.

Maximum accuracy can be obtained from digital frequency and time measurements if the following suggestions are followed:

(1) The effect of the one-count gating error can be minimized by making frequency measurements above $\sqrt{f_c}$ and period measurements below $\sqrt{f_c}$, where f_c is the clock frequency of the counter.

(2) Since long-term stability has a cumulative effect, the accuracy of measurement is mostly a function of the time since the last calibration against a primary or secondary standard.

(3) The accuracy of time measurements is greatly affected by the *slope* of the incoming signal, controlling the signal gate. *Large* signal amplitude and *fast* rise time assure maximum accuracy.

13-11 Some Measurement Applications

(a) *Frequency Measurements.* The measurement capability of an electronic counter in the *frequency* mode of operation can be extended by using a *heterodyne* converter. This is shown in the block diagram of Fig. 13-18. The input signal is applied to a heterodyne converter, which consists of a reference oscillator and a mixer stage with a low-pass filter. The input signal frequency, f_s, and the *reference* oscillator frequency, f_o, are applied to the mixer stage, which produces *both* the sum and the difference of the two input frequencies. The low-pass filter allows *only* the difference frequency to be passed on to the gating circuits of the counter. The counter then counts the frequency $(f_o - f_s)$ or $(f_s - f_o)$, depending on whether the input signal frequency is above or below the reference oscillator frequency. The need to know whether to add or to subtract the counter readout to the reference frequency in order to obtain the frequency of the unknown signal, may complicate the operation somewhat, but this method does extend the range of a general-purpose counter very effectively. A counter with a time-base frequency of 1 MHz usually has an input frequency range of approximately 5 MHz. The use of the frequency converter extends this range to 500 MHz or higher.

Some of the more sophisticated counters have provision for plug-in units, which allow for frequency conversion simply by plugging the appropriate

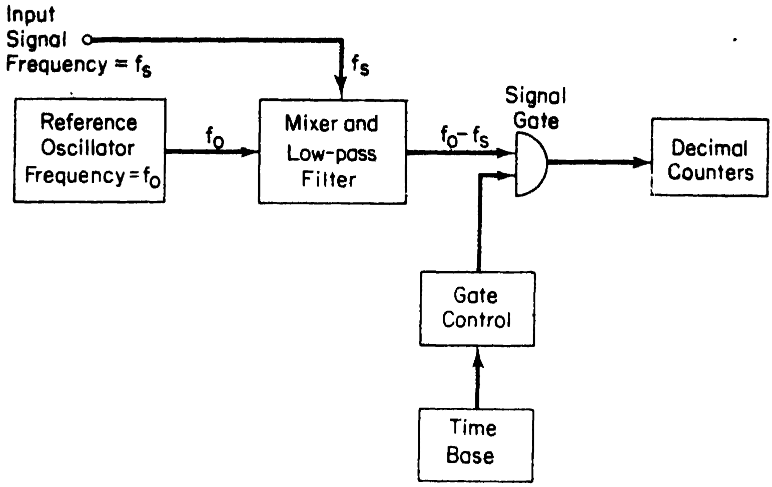


Figure 13-18
Heterodyne frequency conversion generates differences between signal and reference oscillator frequencies.

extension into the basic counter frame. The decade divider assemblies (DDAs) in the oscillator circuit of the counter usually count the time-base frequency of, say, 1 MHz down to 1 Hz, providing a *period* of 1 s. The advantage of a 1-s time base is that the digital readout of the input frequency is then in *cycles per second,* a very convenient feature. If a different time base is selected by proper positioning of the front-panel *time-base* control, the decimal point in the digital readout window usually is placed in such a position that the readout is again in cycles per second.

It is not necessary to use a time base of 1 s and, in fact, many applications require a different time base. For example, if the winch drum of Fig. 13-19 has a circumference of 10 ft, the rope speed (v) in feet per second is 10 times the drum's angular speed (R) in revolutions per second; therefore, $v = 10R$. The rope speed is read out directly in feet per second if the counter counts 10 pulses per revolution for 1 s. If the rope speed in feet per minute is desired, the counter could be arranged to count 10 pulses per revolution for 60 s, by having 10 cams on the drum. A time base commonly used is one which reads out the pulses from a turbine flow meter in units of gallons per minute.

(b) *Time-Interval Measurements.* In these measurements (see the block diagram of Fig. 13-13), the signal gate is opened and closed by the *input signal,* allowing the *time-base* frequency to be counted. In the block diagram of Fig. 13-13, the period trigger provides the opening pulse for the main gate while the multiple-period trigger supplies the closing pulse for the main gate.

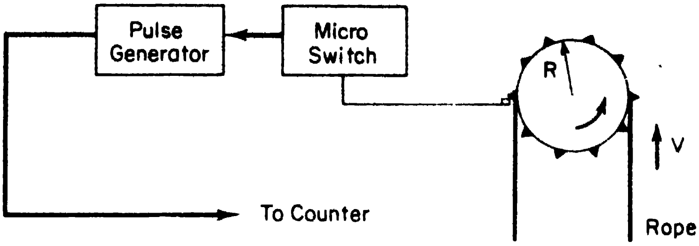


Figure 13-19
Winch drum produces 10 pulses per revolution for application to the counter.

Both pulses are derived from the same input waveform, but one Schmitt trigger responds to the *positive*-going signal while the other Schmitt trigger responds to the *negative*-going waveform. A *trigger level* control allows selection of the point on the incoming waveform, either positive or negative, at which the circuit is triggered. This control can increase noise rejection and decrease the effect of harmonic content upon the measurement. The operation of the trigger level control is shown in Fig. 13-20.

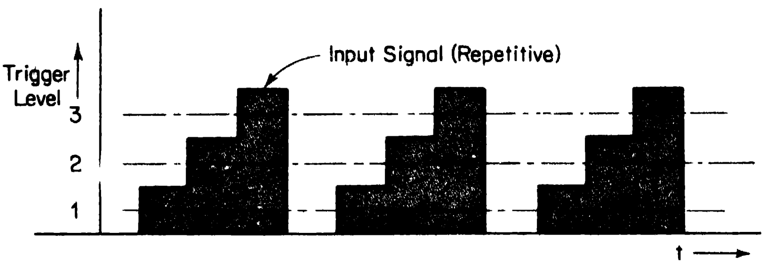


Figure 13-20
The trigger-level control.

One application of time-interval measurements involves determination of *pulsewidth* and *rise time* of an unknown waveform, using the *slope-selection* feature of the instrument (see Fig. 13-21). The signal gate is opened at a point on the *leading* edge of the input signal by the trigger level control of amplifier 1.

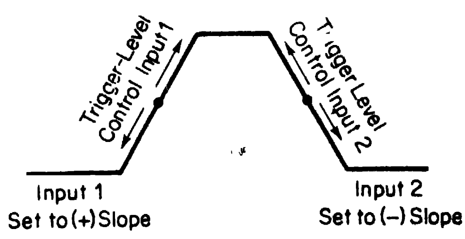


Figure 13-21
Slope-triggering.

The gate is closed at a point on the *trailing* edge of the input signal by the trigger level control of amplifier 2. The width of the pulse is registered by the digital readout and depends on the setting of the time-base selector. If the time-base selector was set at 1 μs (1-MHz frequency) the counter reads the time interval directly in μs.

A different application is shown in Fig. 13-22. Here an electronic counter

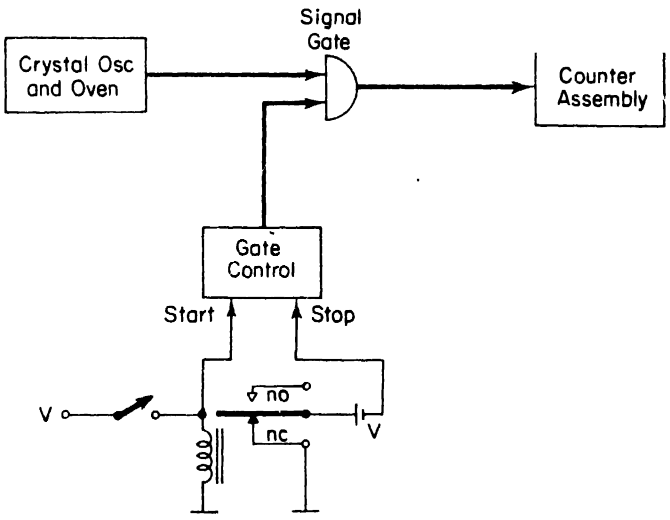


Figure 13-22
Measurement of relay delay-times.

is used to measure the delay times of a *relay*. The relay functions control the opening and closing of the signal gate and the number of cycles of the time-base generator is counted by the DCAs. The various response times are measured as follows:

Delay time: Gate is opened by the coil voltage, upon its application. Gate is closed by the normally closed contacts, when they open.
Transfer time: Gate is opened by the normally closed contacts, when they open. The gate is closed by the normally open contacts, when they close.
Pick-up time: The gate is opened by the application of coil voltage. The gate is closed by the normally open contacts, when they close.
Drop out time: The gate is opened by the removal of coil voltage. The gate is closed by the normally open contacts, when they return to their normally open position at the deenergization of the coil.

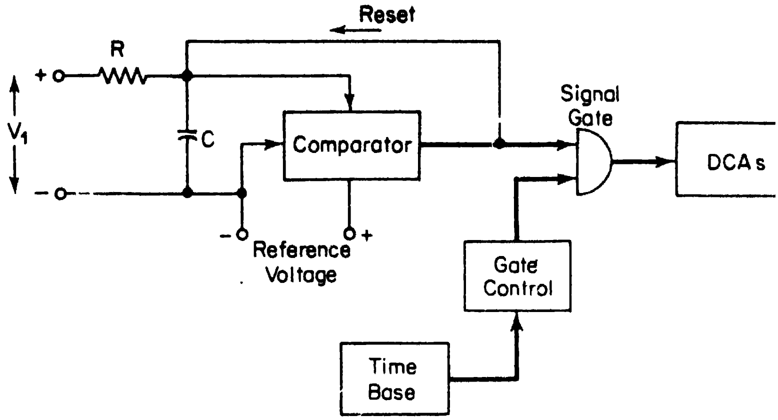


Figure 13-23
Digital voltmeter counts the number of charging cycles during the time-base controlled gate-interval. Each output pulse resets the capacitor voltage to zero.

(c) *Digital Voltmeter.* The pulse-input rate can be made variable to provide analog-to-digital (A/D) conversion. For example, a digital voltmeter can be constructed as shown in Fig. 13-23. An *RC* charging circuit driving a voltage comparator produces a pulse output whose *rate* is proportional to the input voltage. With a constant input voltage (V_1), capacitor C charges from 0 V toward the input voltage. When the trigger level is reached, an output pulse is generated. This pulse discharges the capacitor, and the charging cycle starts again. The larger the input voltage V_1, the faster the voltage on the capacitor reaches the trigger voltage and the higher the output pulse rate. The output pulse rate is proportional to the input voltage. The capacitor and the resistor are usually adjusted to provide a convenient scale factor; a commonly used scale factor is 100-Hz full-scale frequency for 100-mV full-scale input voltage. If the gate time is 1 ms, the readout will be directly in mV.

(d) *Totalizers.* As the name implies, totalizers count and provide a readout of the *total* number of pulses received by the DCAs with no specific gate time being used. They can be used for counting anything, from the number of boxes coming off a production line to pulses from a nuclear particle detector.

Scalers are totalizers with some type of scale factor ahead of the readout. Scalers are particularly useful for converting units. For example, if we obtain one pulse for each egg that rolls down a chute and we want to know how many dozen eggs have rolled by, a scale factor of twelve is applied so that every count in the readout represents 1 dozen eggs.

The same applies to tachometers, where the total number of revolutions is required; the scale factor is the number of pulses from the tachometer-generator per revolution. Scaling is easily accomplished by the same technique employed in the generation of time bases, namely, by using *binary* dividers, *decade* dividers, or other types of *feedback* dividers.

(e) *Preset Counters.* An application of the totalizer is the *preset counter* (a specialized scaler) suited for process control. When the total number in the readout reads the same as the preset number (determined by switches), a pulse is generated and the unit stops counting until reset. A contact closure provided at the preset number can be used for machine control. For example, assume that we are winding a coil of wire and we have a pickoff that provides a pulse for every turn wound on the form. If fifty turns are needed, then the contact closure at the preset number can be used to control the winding mechanism and stop it after it has wound the required fifty turns. The same function can be useful in quality-control programs that require a sample from a given number of units. For example, by using a contact closure to drive an ejecting mechanism, every hundredth egg can be taken from the chute and held for inspection. Multiple functions can be performed by using more than one preset number and set of switches. Suppose we want taps on a coil at 10, 20, and 25 turns. By taking four preset numbers, we can command the machine to make taps as it reaches the first three preset numbers of turns, and to stop at the fourth. The counter would provide momentary contact closures but continue to count until it reached the fourth number.

References

1. Millman, Jacob, and Herbert Taub, *Pulse, Digital, and Switching Waveforms*. New York: McGraw-Hill Book Company, Inc., 1965.

2. Doyle, John M., *Pulse Fundamentals*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1963.

Problems

1. Draw a waveform chart illustrating the action in a chain of three binaries.

2. A chain of three binaries is connected as a *reversible* counter and the initial registration is 7. From this initial state draw the waveform chart showing how the state of each binary changes with each of the next three input pulses.

3. Draw the logic block diagram of a chain of four binaries with appropriate feedback loops to reduce the circuit to a 3/1 counter. Illustrate the operation with a waveform chart.

4. Explain how feedback may be used to convert a scale-of-16 counter to a decimal counter. Indicate two possible ways in which the original circuit may be modified.

5. Explain how two separate feedback paths may be used to convert a scale-of-16 counter to a binary counter. Illustrate the operation with a waveform chart.

6. A time-code generator delivers one output pulse per second. Design a circuit in block diagram form which delivers one output pulse per minute (a 60/1 divider).

7. Consider a digital counter-timer of the type illustrated in Fig. 13-14(b). Draw a simplified block diagram of this counter, explaining the function of each block.

8. Consider a digital counter consisting of three basic blocks: a time base, a control gate, and a decade-counting assembly. Indicate how these basic blocks are interconnected to perform the following measurements: (a) frequency; (b) period; (c) time interval; (d) total count.

9. A digital counter is used to measure a frequency of 2 MHz. The frequency selector of the instrument is set at the kHz position. For what time interval must the gate be open in order to display the correct count?

10. Draw simple circuit diagrams explaining the basic difference in operating principle between a diode OR gate and a diode AND gate.

11. Draw the schematic diagram of a diode OR circuit for use with negative input pulses.

FOURTEEN

TRANSDUCERS AS INPUT ELEMENTS TO INSTRUMENTATION SYSTEMS

14-1 Classification of Transducers

An electronic instrumentation system consists of a number of components which together are used to perform a measurement and record the result. An instrumentation system generally consists of three major elements: an *input* device, a *signal-conditioning* or processing device, and an *output* device. The input device receives the quantity under measurement and delivers a *proportional* electrical signal to the signal-conditioning device. Here the signal is amplified, filtered, or otherwise modified to a format acceptable to the output device. The output device may be a simple indicating meter, a CRO, or a chart recorder for visual display. It may be a magnetic tape recorder for temporary or permanent storage of the input data, or it may be a digital computer for data manipulation or process control. The kind of system depends on what is to be measured and how the measurement result is to be presented. (See also Chapter 15 for data-acquisition systems.)

The input quantity for most instrumentation systems is a *nonelectrical* quantity. In order to use electrical methods and techniques for measurement, manipulation, or control, the nonelectrical quantity is generally *converted* into an electrical signal by a device called a *transducer*. One definition states "a transducer is a device which, when actuated by energy in one transmission system, supplies energy in the same form or in another form to a second transmission system." This energy transmission may be electrical, mechanical, chemical, optical (radiant) or thermal.

This broad definition of a transducer includes, for example, devices which convert *mechanical* force or displacement into an electrical signal. These devices form a very large and important group of transducers commonly found in the industrial instrumentation area. The instrumentation engineer and technologist is therefore primarily concerned with this type of energy conversion. Many other *physical* parameters (such as heat, light intensity, humidity) may also be converted into electrical energy by means of transducers. These transducers provide an output signal when stimulated by a *nonmechanical* input: a thermistor reacts to temperature variations, a photocell to changes in light intensity, an electron beam to magnetic effects, and so on. In all cases, however, the electrical output is measured by *standard* methods, yielding the magnitude of the input quantity in terms of an *analog* electrical measure.

Transducers may be classified according to their application, method of energy conversion, nature of the output signal, and so on. All these classifications usually result in overlapping areas. A sharp distinction between, and classification of, types of transducers is difficult. Table 14-1 shows a classification of transducers according to the *electrical principles* involved. The first

TABLE 14-1

TYPES OF TRANSDUCERS*

Electrical Parameter and Class of Transducer	Principle of Operation and Nature of Device	Typical Application
	Passive Transducers (externally powered)	
Resistance		
Potentiometric device	Positioning of the slider by an external force varies the resistance in a potentiometer or a bridge circuit.	Pressure, displacement
Resistance strain gage	Resistance of a wire or semiconductor is changed by elongation or compression due to externally applied stress.	Force, torque, displacement
Pirani gage or hot wire meter	Resistance of a heating element is varied by convection cooling of a stream of gas.	Gas flow, gas pressure
Resistance thermometer	Resistance of pure metal wire with a large positive temperature coefficient of resistance varies with temperature.	Temperature, radiant heat
Thermistor	Resistance of certain metal oxides with negative temperature coefficient of resistance varies with temperature.	Temperature
Resistance hygrometer	Resistance of a conductive strip changes with moisture content.	Relative humidity
Photoconductive cell	Resistance of the cell as a circuit element varies with incident light.	Photosensitive relay

*Classified according to their electrical principles.

TABLE 14-1

TYPES OF TRANSDUCERS (Continued)

Electrical Parameter and Class of Transducer	Principle of Operation and Nature of Device	Typical Application
	Passive Transducers (externally powered)	
Capacitance		
Variable capacitance pressure gage	Distance between two parallel plates is varied by an externally applied force	Displacement, pressure
Capacitor microphone	Sound pressure varies the capacitance between a fixed plate and a movable diaphragm.	Speech, music, noise
Dielectric gage	Variation in capacitance by changes in the dielectric.	Liquid level, thickness
Inductance		
Magnetic circuit transducer	Self-inductance or mutual inductance of ac-excited coil is varied by changes in the magnetic circuit.	Pressure, displacement
Reluctance pickup	Reluctance of the magnetic circuit is varied by changing the position of the iron core of a coil.	Pressure, displacement, vibration, position
Differential transformer	The differential voltage of two secondary windings of a transformer is varied by positioning the magnetic core through an externally applied force.	Pressure, force, displacement, position
Eddy current gage	Inductance of a coil is varied by the proximity of an eddy current plate.	Displacement, thickness
Magnetostriction gage	Magnetic properties are varied by pressure and stress.	Force, pressure, sound
Voltage and current		
Hall effect pickup	A potential difference is generated across a semiconductor plate (germanium) when magnetic flux interacts with an applied current.	Magnetic flux, current
Ionization chamber	Electron flow induced by ionization of gas due to radioactive radiation.	Particle counting, radiation
Photoemissive cell	Electron emission due to incident radiation upon photoemissive surface.	Light and radiation
Photomultiplier tube	Secondary electron emission due to incident radiation on photosensitive cathode.	Light and radiation, photosensitive relays
	Self-generating Transducers (no external power)	
Thermocouple and thermopile	An emf is generated across the junction of two dissimilar metals or semiconductors when that junction is heated	Temperature, heat flow, radiation
Moving coil generator	Motion of a coil in a magnetic field generates a voltage.	Velocity, vibration
Piezoelectric pickup	An emf is generated when an external force is applied to certain crystalline materials, such as quartz.	Sound, vibration, acceleration, pressure changes
Photovoltaic cell	A voltage is generated in a semiconductor junction device when radiant energy stimulates the cell.	Light meter, solar cell

part of Table 14-1 lists transducers that require external power. These are the *passive* transducers, producing a variation in some electrical parameter, such as resistance, capacitance, and so on, which can be measured as a voltage or current variation. The second category of transducers in Table 14-1 are of the *self-generating* type, producing an analog voltage or current when stimulated by some physical form of energy. The self-generating transducers do *not* require external power. Although it would be almost impossible to classify all sensors and measurements, the devices listed in Table 14-1 do represent a valid cross section of commercially available transducers for application in instrumentation engineering. Some of the more common transducers and their application are discussed in the following sections.

14-2 Selecting a Transducer

Selection of the appropriate transducer for any given measurement is the first and perhaps most important step in obtaining *accurate* results. In a measurement system, the transducer is the input element with the critical function of transforming some physical quantity to a proportional electrical signal. A number of elementary questions must be answered before a transducer can be selected. These questions may be stated as follows:

(1) What is the physical quantity to be measured?

(2) Which transducer principles can be used to measure this quantity?

(3) What accuracy is required for this measurement?

The first question can be answered by determining the *type* and the *range* of the measurand. An appropriate answer to the second question requires that the *input* and *output* charactertistic of the transducer be compatible with the recording equipment and other elements used in the measurement system. In most cases, these two questions can be answered readily, implying that the proper transducer is selected simply by the addition of an *accuracy tolerance*. In practice, this is rarely true owing to the complexity and magnitude of the basic transducer parameters which affect the accuracy. Accuracy of measurement should be analyzed in terms of the *individual* factors contributing to it. These accuracy factors are listed in detail. The over-all conditions and accuracy requirements of the *system* determine the degree to which these factors must be considered.

(1) *Fundamental transducer parameters:* type and range of measurand, sensitivity, excitation

(2) *Physical conditions:* mechanical and electrical connections, mountint provisions, corrosion resistance

(3) *Ambient conditions:* nonlinearity effects, hysteresis effects, frequency response, resolution

(4) *Environmental conditions:* temperature effects, acceleration, shock and vibration

(5) *Compatibility of the associated equipment:* zero balance provisions, sensitivity tolerance, impedance matching, insulation resistance

Categories 1 and 2 are basic electrical and mechanical characteristics of the transducer. Transducer accuracy, as an independent component, is contained in categories 3 and 4. Category 5 considers the transducer's compatibility with its associated system equipment.

The *total measurement error* in a transducer-activated system may be reduced to fall within the required accuracy range by the following techniques:

(1) Using *in place* system *calibration* with corrections performed in the data reduction

(2) Simultaneously *monitoring* the environment and correcting the data accordingly

(3) Artificially controlling the environment to minimize possible errors

Some *individual* errors are predictable and can be calibrated out of the system. When the entire system is calibrated, these calibration data may then be used to correct the recorded data. The *environmental* errors can be corrected by data reduction if the environmental effects are recorded simultaneously with the actual data. Then the data are corrected using the known environmental characteristics of the transducers. These two techniques can provide a significant increase in system accuracy.

Another method to improve over-all system accuracy is to control *artificially* the environment of the transducer. If the environment of the transducer can be kept unchanged, these errors are reduced to zero. This type of control may require either physically moving the transducer to a more favorable position or providing the required isolation from the environment by a heater enclosure, vibration isolation, or similar means.

14-3 Strain Gages

(a) *Principle of Operation.* The strain gage is one example of a *passive* transducer, converting a mechanical displacement into a change in resistance. The gage is built of thin resistance wire of about 1 mil in diameter. The wire material used may be Constantan (60 per cent Cu and 40 per cent Ni) or Alloy

479 (92 per cent Pt and 8 per cent Wo). The wire is bonded to a supporting base to form a *bonded* strain gage or suspended on a frame structure to form an *unbonded* strain gage. In any case, the supporting structure is subjected to strain (lateral tension or pressure) causing the thin gage wire to change its length and its cross-sectional area. The deformation in the wire causes a change in its resistance (ΔR) which is a measure of the applied strain. The resistance change ΔR is usually measured with a *Wheatstone* bridge.

The characteristics of a strain gage are described in terms of its *sensitivity* (S), also called the *gage factor*. S is defined as the unit change in resistance per unit change in length and is expressed as

$$\text{sensitivity } S = \frac{\Delta R/R}{\Delta l/l} \tag{14-1}$$

where S = the sensitivity or gage factor
R = gage wire resistance
ΔR = change in gage wire resistance
l = length of the gage wire in the unstressed condition
Δl = change in length of the gage wire.

The term $\Delta l/l$ in the denominator of Eq. (14-1) is the strain σ, so that Eq. (14-1) can be rewritten as

$$S = \frac{\Delta R/R}{\sigma} \tag{14-2}$$

where σ = the strain in the lateral direction.

The resistance change ΔR of a conductor with length l can be calculated by using the expression for the resistance of a conductor of uniform cross section:

$$R = \rho \frac{\text{length}}{\text{area}} = \frac{\rho \times l}{(\pi/4)d^2} \tag{14-3}$$

where ρ = the specific resistance of the conductor material
l = length of the conductor
d = diameter of the conductor.

Tension on the conductor causes an increase Δl in its length and a simultaneous decrease Δd in its diameter. The resistance of the conductor then changes to

$$R_s = \rho \frac{(l + \Delta l)}{(\pi/4)(d - \Delta d)^2} = \rho \frac{l(1 + \Delta l/l)}{(\pi/4)d^2(1 - 2\,\Delta d/d)} \tag{14-4}$$

Equation (14-4) may be simplified by using Poisson's ratio, μ, which is defined as the ratio of strain in the lateral direction to strain in the axial direction. Therefore,

$$\mu = \frac{\Delta d/d}{\Delta l/l} \tag{14-5}$$

Substitution of Eq. (14-5) into Eq. (14-4) yields

$$R_s = \rho \frac{l}{(\pi/4)d^2}\left(\frac{1 + \Delta l/l}{1 - 2\mu\,\Delta l/l}\right) \tag{14-6}$$

Equation (14-6) may be simplified to

$$R_s = R + \Delta R = R\left[1 + (1 + 2\mu)\frac{\Delta l}{l}\right] \tag{14-7}$$

The increment of resistance ΔR as compared to the increment of length Δl may then be expressed in terms of the sensitivity S where

$$S = \frac{\Delta R/R}{\Delta l/l} = 1 + 2\mu \tag{14-8}$$

Poisson's ratio for *most* metals lies in the range of 0.25 to 0.35 and the gage factor would then be in the order of 1.5 to 1.7.

For strain-gage applications, a *high sensitivity* is of course very desirable. A high gage factor means a relatively large resistance change which can be more easily measured than a small resistance change. For Constantan wire, S is about 2, whereas Alloy 479 gives a value of about 4.

Semiconductor material, such as silicon and germanium, is also used for strain gages. By controlling the type and the amount of impurity (doping), the semiconductor properties can be optimized for specific applications. The sensitivity of a semiconductor strain gage is very much higher than a resistance wire strain gage and is typically in the order of 50 to 200. Unfortunately, semiconductor strain gages are very sensitive to temperature fluctuations and effective ways of *temperature compensation* must be provided.

It is interesting to carry out a simple calculation to find out what effect an applied stress has on the resistance change of a strain gage. Hooke's law gives the relationship between stress and strain for a linear stress-strain curve, in terms of the modulus of elasticity of the material under tension. Defining *stress* as the applied force per unit area and *strain* as the elongation of the stressed member per unit length, Hooke's law is written as

$$\sigma = \frac{s}{E}$$

where σ = strain, $\Delta l/l$ (no units)
s = stress, lb/in.^{2}
E = Young's modulus, lb/in.^{2}.

Example 14-1: A resistance strain gage with a gage factor of 4 is fastened to a steel member subjected to a stress of 15,000 lb/in.^{2}. The modulus of elasticity of steel is approximately 30×10^6 lb/in.^{2}. Calculate the change in resistance, ΔR, of the strain-gage element due to the applied stress.

SOLUTION: Hooke's law, Eq. (14-9), yields

$$\sigma = \frac{\Delta l}{l} = \frac{s}{E} = \frac{15{,}000}{30 \times 10^6} = \frac{1}{2{,}000}$$

The sensitivity of the strain gage $S = 2$. Therefore, from Eq. (14-2),

$$\frac{\Delta R}{R} = S\sigma = 2 \times \frac{1}{2{,}000} = \frac{1}{1{,}000} = 0.1\%$$

Example 14-1 illustrates that the very high stress of 15,000 lb/in.^{2} results in a resistance change of only 0.1 per cent, a very small change indeed. Actual measurements generally involve resistance changes of very much lower values, and the bridge measuring circuit must be very carefully designed to be able to detect these small changes in resistance.

(b) *Construction and Application of Strain Gages*. The resistance of bonded *strain gages* changes under stress application because the gages are made from fine wire looped back and forth to increase their total effective length. The resistance wire is cemented to a thin sheet of paper or plastic and covered with a protective coating of paper or felt. The size of the finished bonded gage, and the manner in which the wire pattern is arranged, varies with the application. Some bonded gages can be as small as $\frac{1}{8}$ in. by $\frac{1}{8}$ in. although they are generally somewhat larger and are manufactured to a maximum size of 1 in. long by $\frac{1}{2}$ in. wide. In the usual application, the bonded strain gage is cemented to the structure whose strain is to be measured. The problem of providing a good bonding between the gage and the structure is very difficult. The adhesive material must hold the gage firmly to the structure, yet it must have sufficient elasticity to give under strain without losing its adhesive properties. The adhesive should also be resistant to temperature, humidity, and other environmental conditions. Some of the patterns in which bonded strain gages are commercially available are arranged as shown in Fig. 14-1.

The *unbonded strain gage* consists of a stationary frame and an armature which is supported in the center of the frame. The armature can move only in one direction. Its travel in that direction is limited by four filaments of strain-sensitive wire, wound between rigid insulators which are mounted on the frame and on the armature. The filaments are of equal length and are arranged as shown in Fig. 14-2(a).

When an external force is applied to the strain gage, the armature moves in the direction indicated. The elements A and D increase in length, whereas the elements B and C decrease in length. The resistance change of the four filaments is proportional to their change in length, and this change can be measured with a Wheatstone bridge circuit, as shown in Fig. 14-2(b). The unbalance current, indicated by the current meter, is calibrated to read the magnitude of the displacement of the armature.

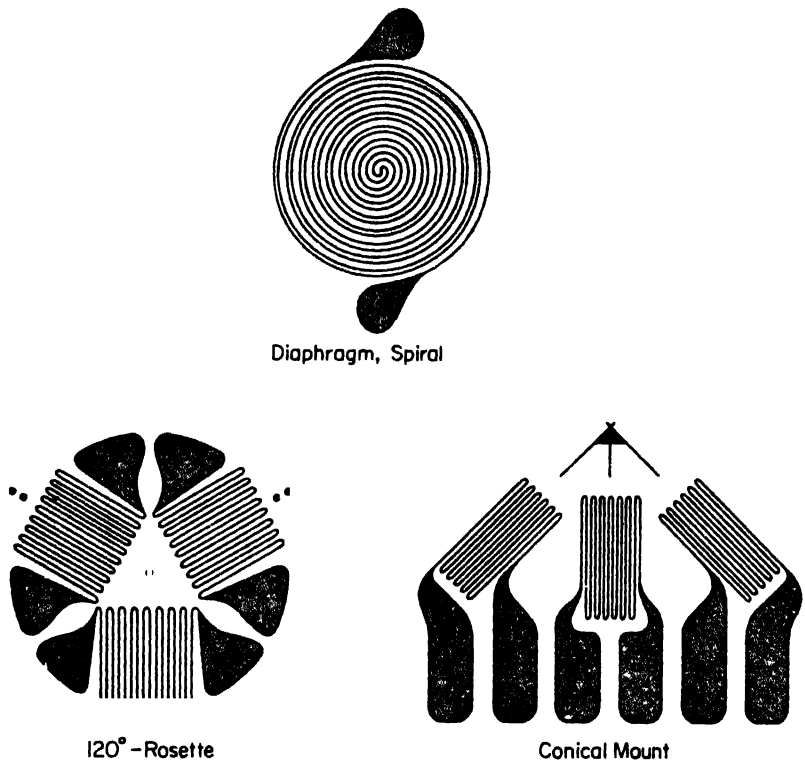


Figure 14-1
Strain gage patterns.

Sufficient initial tension is applied to the strain-sensitive resistance wires during the assembly of the transducer, to keep them under some residual tension when the armature is in either extreme position. The armature travel is limited mechanically to protect the wires from overload. The filaments must have stable elastic properties, high tensile strength, and high corrosion resistance. Certain alloys, possessing the right combination of gage factor and thermal coefficient of resistivity, are selected for optimum performance of the strain-sensitive wires.

Most specifications list the nominal resistance of the strain-gage bridge circuit at 350 Ω. This value is often selected because many measurements are made with recording galvanometers which require approximately 350 Ω of shunt resistance for optimum response. Bridge resistances from fifty to several thousand ohms are available to match a variety of receivers, such as magnetic tape recorders, cathode ray oscilloscopes, telemetering systems, and automated equipment.

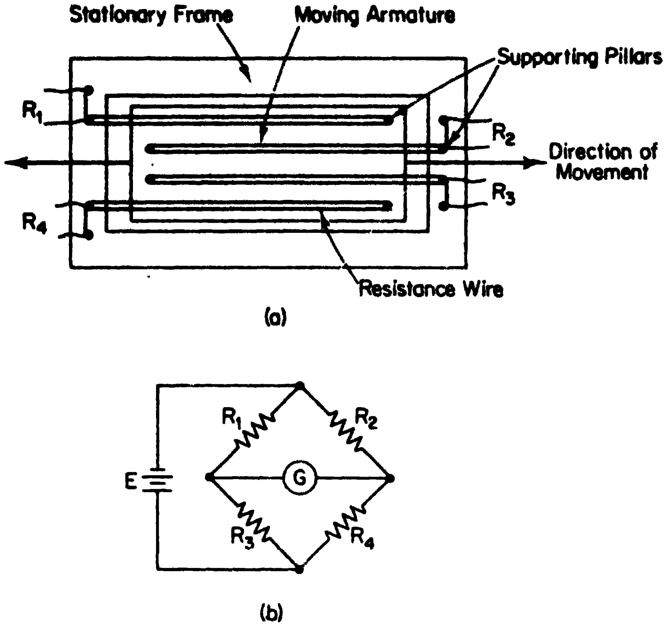


Figure 14-2
The unbonded strain gage. (a) Principle of construction. (b) Measuring circuit. When strain is applied to the gage, R_1 and R_4 both increase in resistance and R_2 and R_3 both decrease, or, with strain in the opposite direction, R_1 and R_4 both decrease while R_2 and R_3 both increase. This arrangement increases the sensitivity of the strain gage substantially.

The unbonded strain-gage transducer can be constructed in a variety of configurations, depending on the use required of the instrument. Its principal use is as a *displacement* transducer: a linkage pin can be attached to the armature in order to measure displacement directly. The unit of Fig. 14-2(a) allows an armature displacement of 0.0015 in. on each side of its center position. Using the same construction, the unit will function as a dynamometer, capable of measuring *force*. Depending on the number of turns and the diameter of the strain wires, the transducer will measure forces from ± 1.5 oz to ± 5 lb, full-scale.

The transducer becomes a *pressure* pickup when its armature is connected to a metallic bellows or diaphragm. When a bellows is used, force on the end of the bellows is transmitted by a pin to the armature, and the unit functions as a dynamometer. By applying pressure to one side of the bellows and venting the other side to the atmosphere, gage pressures may be read. If the bellows is evacuated and sealed, *absolute pressure* is measured.

Another modification is provided by two pressure connections, one to each side of the bellows or diaphragm, for the measurement of *differential* pressure. Finally, when a weight is fastened to the armature, the transducer becomes an *accelerometer*.

14-4 Displacement Transducers

(*a*) *Introduction*. The initial concept of converting an applied force into a displacement is basic to many types of transducers. The mechanical elements which are used to convert the applied force into a displacement are called *force-summing* devices. The most generally used force-summing members are

(1) Diaphragm, flat or corrugated
(2) Bellows
(3) Bourdon tube, circular or twisted
(4) Straight tube
(5) Mass-cantilever, single or double suspension
(6) Pivot torque

Examples of these force-summing devices are given in Fig. 14-3. *Pressure* transducers generally use one of the first four types of force-summing members; categories (5) and (6) may be found in *accelerometer* and *vibration* pickups.

The displacement created by the action of the force-summing device is converted into a change of some *electrical* parameter. The electrical principles most commonly used in measuring displacement are

(1) Capacitive
(2) Inductive
(3) Differential transformer
(4) Ionization
(5) Oscillation
(6) Photoelectric
(7) Piezoelectric
(8) Potentiometric
(9) Velocity generator

These principles are discussed and illustrated in the following section.

(*b*) *Capacitive Transducers*. The capacitance of a parallel-plate capacitor is given by the equation

$$C = \frac{kA\varepsilon_0}{d} \quad \text{(farads)} \tag{14-10}$$

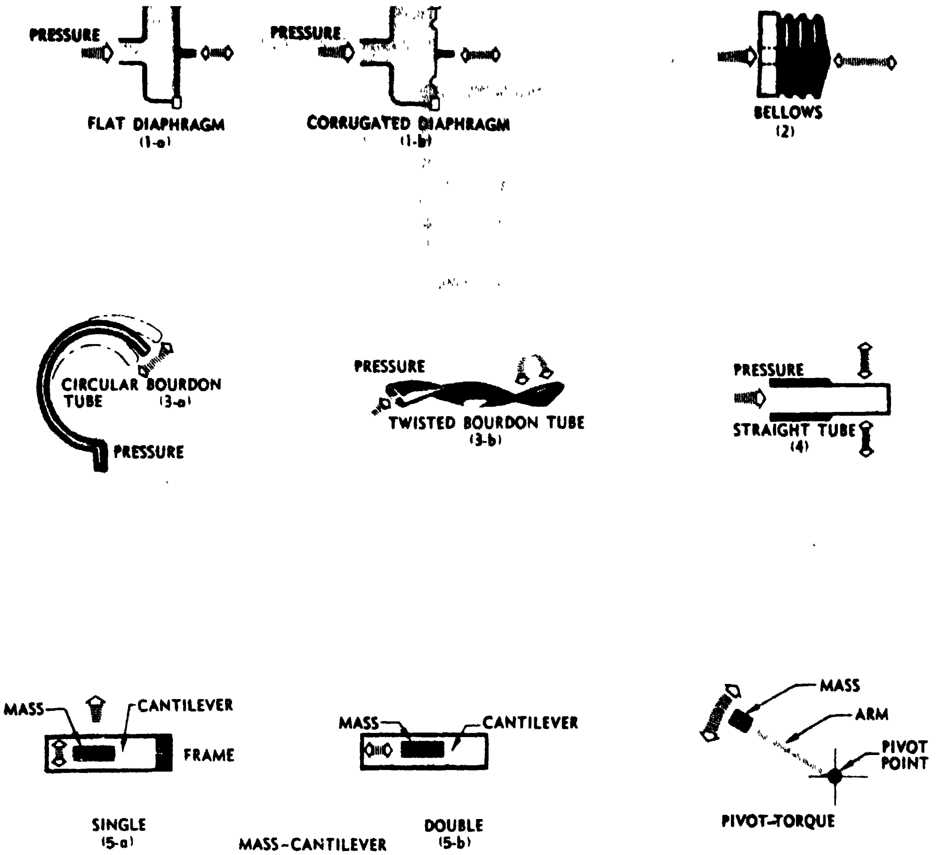


Figure 14-3
Force-summing devices. (*Courtesy Statham Instruments, Inc.*)

where A = the area of each plate, in m^2
d = the plate spacing, in m
$\varepsilon_0 = 9.85 \times 10^{-12}$, in F/m
k = dielectric constant.

Since the capacitance is inversely proportional to the spacing of the parallel plates, any variation in d causes a corresponding variation in the capacitance. This principle is applied in the capacitive transducer, indicated in Fig. 14-4. A force, applied to a diaphragm, which functions as one plate of a simple capacitor, changes the distance between the diaphragm and the static plate. The resulting change in capacitance may be measured with an ac bridge, but is usually measured with an oscillator circuit. The transducer, as part of the oscillatory circuit, causes a change in the frequency of the oscillator. The change in frequency is a measure of the magnitude of the original applied force.

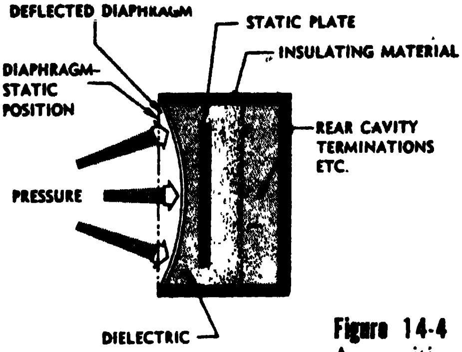


Figure 14-4
A capacitive transducer. (*Courtesy Statham Instruments, Inc.*)

The capacitive transducer has *excellent* frequency response and can measure either static or dynamic phenomena. Its disadvantages are sensitivity to temperature variations and possibility of erratic or distorted signals due to long lead length. Also, the receiving instrumentation may be large and complex and often includes a second fixed frequency oscillator for heterodyning purposes. The difference frequency, thus produced, can be read by an appropriate output readout device, such as an electronic counter.

(*c*) *Inductive Transducers*. The measurement of force is accomplished by the change in the *inductance ratio* of a pair of coils, or by the change of *inductance* in a single coil. The ferromagnetic armature is displaced by the force being measured, varying the *reluctance* of the magnetic circuit. Figure 14-5 shows how the air gap is varied by a change in position of the armature. The resulting

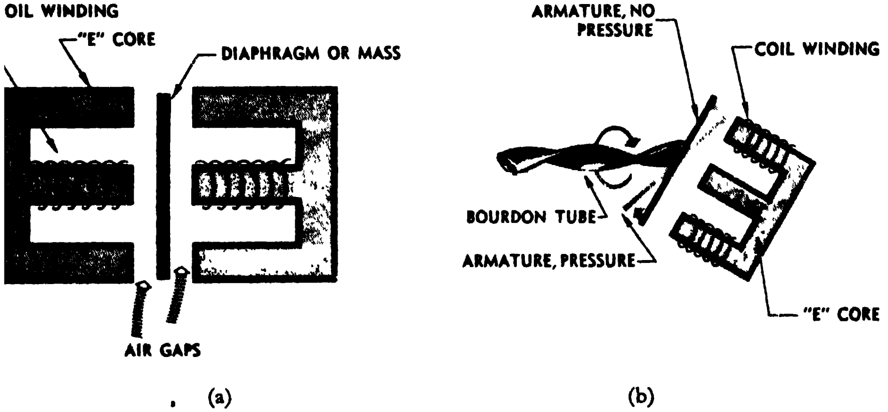


Figure 14-5
Inductive transducers. (*Courtesy Statham Instruments, Inc.*)

change in inductance of each coil is a measure of the magnitude of the applied force.

The coil can be used as a component of an LC oscillator circuit whose frequency then varies with applied force. This type of transducer is used extensively in telemetry systems, with a single coil which modulates the frequency of a local oscillator.

Hysteresis errors of the transducer are almost entirely limited to the mechanical components. When a diaphragm is used as the force-summing member, as shown in Fig. 14-5(a), it may form part of the magnetic circuit. In this arrangement, the over-all performance of the transducer is somewhat degraded because the desired mechanical characteristics of the diaphragm must be compromised to improve the magnetic performance.

The inductive transducer responds to static and dynamic measurements, it has continuous resolution and a fairly high output. Its disadvantages are that the frequency response (variation of the applied force) is limited by the construction of the force-summing member. In addition, external magnetic fields may cause erratic performance.

(*d*) *Differential Transformer Transducers.* The differential transformer transducer measures force in terms of the displacement of the ferromagnetic core of a transformer. The basic construction of the linear variable differential transformer (LVDT) is given in Fig. 14-6. The transformer consists of a *single* primary winding and *two* secondary windings, which are placed on either side of the primary. The secondaries have an equal number of turns but they are connected in series opposition so that the emf's induced in the coils oppose each other. The position of the movable core determines the flux linkage between the ac-excited primary winding and each of the two secondary windings.

With the core in the center or *reference* position, the induced emf's in the secondaries are equal and since they oppose each other, the output voltage will be 0 V. When an externally applied force moves the core to the left-hand position, more magnetic flux links the left-hand coil than the right-hand coil. The induced emf of the left-hand coil is therefore larger than the induced emf of the right-hand coil. The magnitude of the output voltage is then equal to the *difference* between the two secondary voltages and it is *in phase* with the voltage of the left-hand coil. Similarly, when the core is forced to move to the right, more flux links the right-hand coil than the left-hand coil and the resulting output voltage is now in phase with the emf of the right-hand coil, while its magnitude again equals the difference between the two induced emf's—see Fig. 14-6(b).

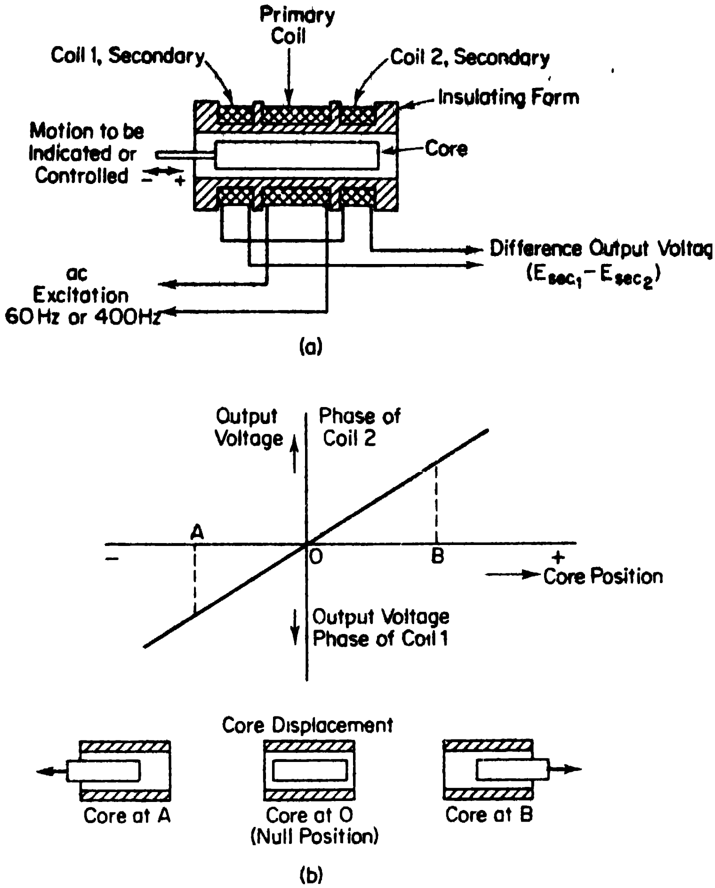


Figure 14-6
The linear variable differential transformer (LVDT). (a) The essential components of the LVDT. (b) The relative positions of the core generate the indicated output voltages. The linear characteristics impose limited core movements, which are typically up to 0.2 in. from the null position.

The output of the differential transformer may serve as a component in force-balancing *servo systems*. This is indicated schematically in Fig. 14-7. The output terminals of an *input* transformer and a *balancing* transformer are connected in series opposition. The algebraic sum of the two voltages is fed to an amplifier which drives a two-phase motor. When the two transformers are in their reference positions, the sum of their output voltages is zero and no voltage is delivered to the servo motor. When the core of the input transformer is moved away from its reference position by an externally applied displacement input, an output voltage is delivered to the amplifier and the motor

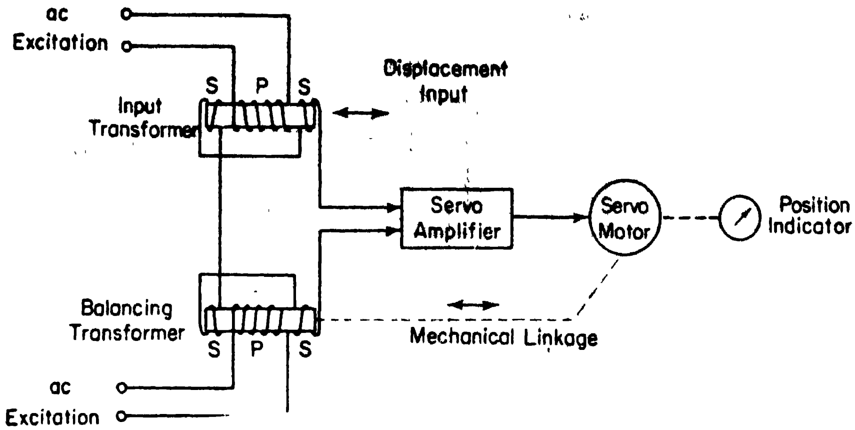


Figure 14-7
Displacement measurement using two differential transformers in a closed-loop servo system.

rotates. The motor shaft is mechanically coupled to the core of the balancing transformer. Since the output of the balancing transformer opposes the output of the input transformer, the motor continues to rotate until the outputs of the two transformers are equal. The indicator on the motor shaft is calibrated to read the displacement of the balancing transformer and, indirectly, the displacement of the input transformer.

A variation of the moving-core differential transformer is given in Fig. 14-8. Here the primary winding is wound on the center leg of an E core, and the secondary windings are wound on the outer legs of the E core. The armature is rotated about a pivot point above the center leg of the core by the externally applied force. When the armature is displaced from its balance or reference position, the reluctance of the magnetic circuit through one secondary coil is decreased, while, simultaneously, the reluctance of the magnetic circuit through the other secondary coil is increased. The induced emf's in the secondary windings, which are equal in the reference position of the armature, will now be different in magnitude as a result of the applied displacement. The induced emf's again oppose each other and the transformer operates in the same manner as the moving-core transformer of Fig. 14-6.

The differential transformer provides continuous resolution and shows low hysteresis. Relatively large displacements are required, and the instrument is sensitive to vibration. The receiving instrument must be selected to operate on ac signals or a demodulator network must be used if a dc output is required.

(*e*) *Oscillation Transducers.* This class of transducer uses the force-summing member to change the capacitance or the inductance in an LC oscillator cir-

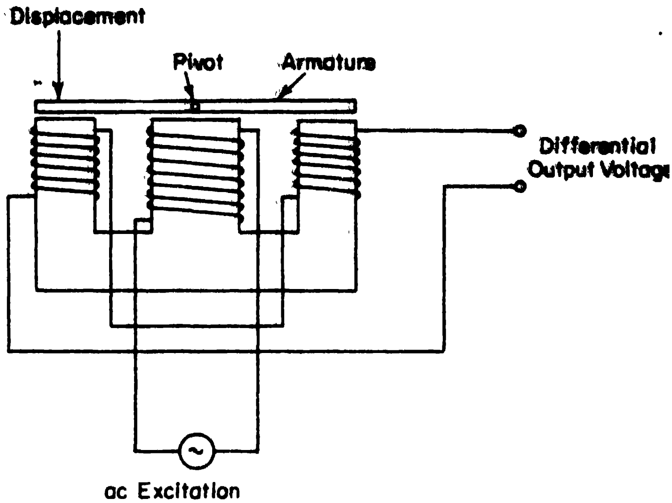


Figure 14-8
A differential transformer using an E-core and pivoted armature.

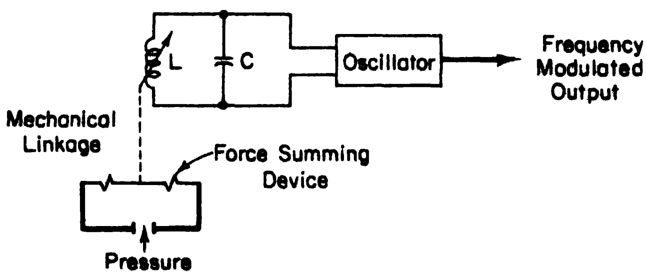


Figure 14-9
Basic elements of an oscillation transducer.

cuit. Figure 14-9 shows the basic elements of an LC transistor oscillator whose frequency is affected by a change in the inductance of the coil. The stability of the oscillator must be excellent in order to detect changes in oscillator frequency caused by the externally applied force.

This transducer measures both static and dynamic phenomena and is convenient for use in telemetry applications. Its limited temperature range, poor thermal stability, and low accuracy restrict its use to low-accuracy applications.

(*f*) *Photoelectric Transducers.* The photoelectric transducer makes use of the properties of a photoemissive cell or phototube. The ***phototube*** is a radiant energy device which controls its electron emission when exposed to incident

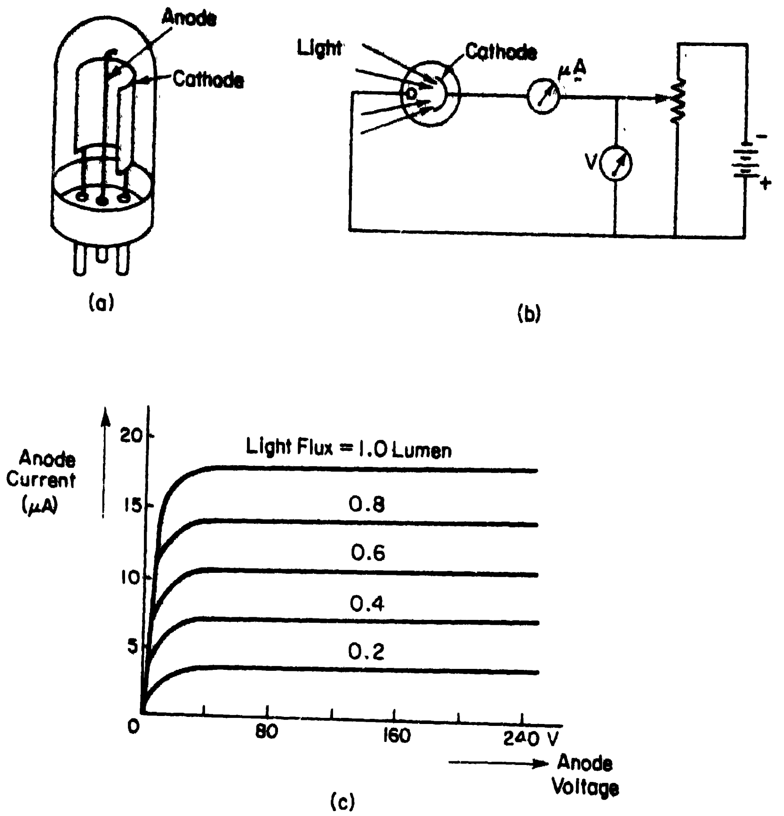


Figure 14-10
(a) Construction of a photoemissive tube. (b) Test circuit for a phototube. (c) Characteristic curves for a high-vacuum phototube.

light. The construction of a phototube is shown in Fig. 14-10(a); its symbol is given in the schematic diagram of Fig. 14-10(b).

The large semicircular element is the photosensitive cathode and the thin wire down the center of the tube is the anode. Both elements are placed in a high-vacuum glass envelope. When a constant voltage is applied between the cathode and the anode, the current in the circuit is directly proportional to the amount of light, or light intensity, falling on the cathode. The curves of Fig. 14-10(c) show the anode characteristics of a typical high-vacuum phototube.

Notice that for voltages above approximately 20 V the output current is nearly independent of applied anode voltage but depends entirely on the amount of incident light. The current through the tube is extremely small, usually in the range of a few microamperes. In most cases therefore, the phototube is connected to an amplifier to provide a useful output.

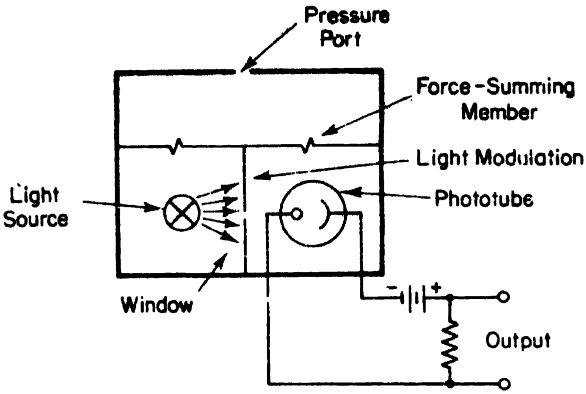


Figure 14-11 Elements of a photoelectric transducer.

The photoelectric transducer of Fig. 14-11 uses a phototube and a light source separated by a small window, whose aperture is controlled by the force-summing member of the pressure transducer. The displacement of the force-summing member modulates the quantity of incident light on the photosenstitive element. According to the curves of Fig. 14-10(c), a change in light intensity varies the photoemissive properties at a rate approximately linear with displacement. This transducer can use either a stable source of light or an ac-modulated light.

The advantages of this type of transducer are its high efficiency and its adaptability to measuring both static and dynamic conditions. The device may have poor long-term stability, does not respond to high-frequency light variations, and requires a large displacement of the force-summing member.

(*g*) *Piezoelectric Transducers.* Asymmetrical crystalline materials, such as quartz, Rochelle salt, and barium titanite, produce an emf when they are placed under stress. This property is used in piezoelectric transducers, where a crystal is placed between a solid base and the force-summing member, as shown in Fig. 14-12. An externally applied force, entering the transducer through its pressure port, applies the pressure to the top of a crystal. This produces an emf across the crystal, proportional to the magnitude of the applied pressure.

Since the transducer has a very good high-frequency response, its principal use is in high-frequency accelerometers. In this application, its output voltage is typically in the order of 1 to 30 mV per g of acceleration. The device needs no external power source and is therefore self-generating. The principal disadvantage of this transducer is that it cannot measure static conditions. The output voltage is also affected by temperature variations of the crystal.

(*h*) *Potentiometric Transducers.* A potentiometric transducer is an electromechanical device containing a resistance element which is contacted by a

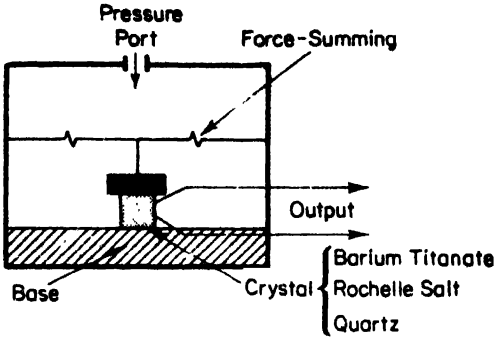


Figure 14-12 Elements of piezoelectric transducer.

movable slider. Motion of the slider results in a resistance change that may be linear, logarithmic, exponential, and so on, depending on the manner in which the resistance wire is wound. In some cases deposited carbon, platinum film, and other techniques are used to provide the resistance element. The basic elements of the potentiometric transducer are given in Fig. 14-13.

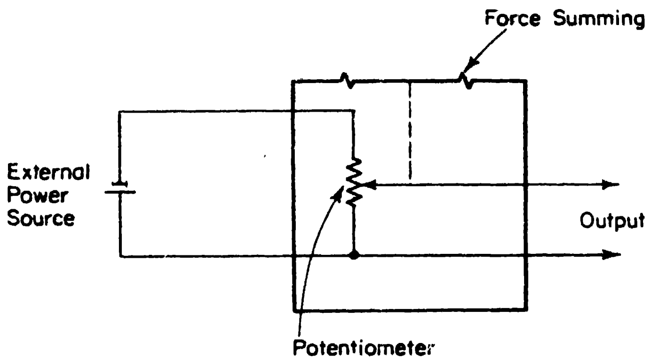


Figure 14-13
Principle of the potentiometric transducer.

The potentiometric principle is widely used despite its many limitations. Its electrical efficiency is very high and it provides a sufficient output to permit control operations without further amplification. The device may be ac- or dc-excited and can therefore serve a wide range of functions. Owing to the mechanical friction of the slider against the resistance element, its life is limited and noise may develop as the element wears out. Large displacements are usually required to move the slider along the entire working surface of the potentiometer.

(*i*) *Velocity Transducers.* The velocity transducer essentially consists of a moving coil suspended in the magnetic field of a permanent magnet, as shown

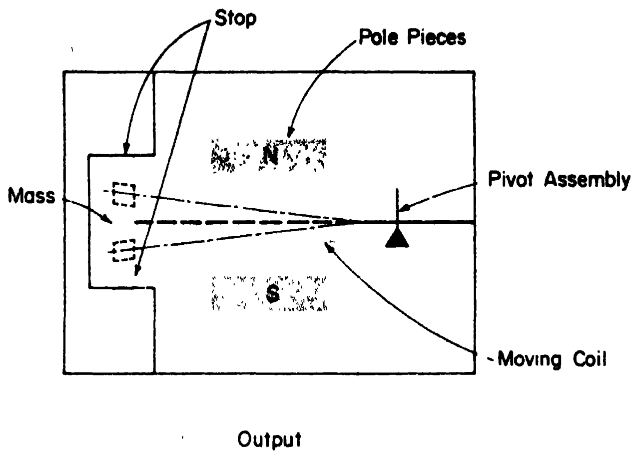


Figure 14-14
Elements of a velocity transducer.

in Fig. 14-14. A voltage is generated by the motion of the coil in the field. The output is proportional to the velocity of the coil, and this type of pickup is therefore generally used for the measurement of velocities developed in a linear, sinusoidal, or random manner. Damping is obtained electrically, thus assuring high stability under varying temperature conditions.

14-5 Temperature Measurements

(*a*) *Resistance Thermometers*. The resistance thermometer makes use of the fact that almost all pure metals undergo a change in resistance which is directly proportional to temperature change. The metals used in the sensing element of a resistance thermometer are *copper, nickel,* and *platinum,* which exhibit resistance-temperature characteristics as shown in Fig. 14-15. The figure indicates that the resistance of some pure metals increases almost linearly with increasing temperature.

The resistance of a conductor at some temperature above or below a specified reference temperature is given by.

$$R_t = R_{ref}(1 + \alpha\Delta t) \tag{14-11}$$

where R_t = resistance of the conductor at temperature t
R_{ref} = resistance of the conductor at reference temperature, usually 20°C
α = average temperature coefficient of resistance of the conductor material

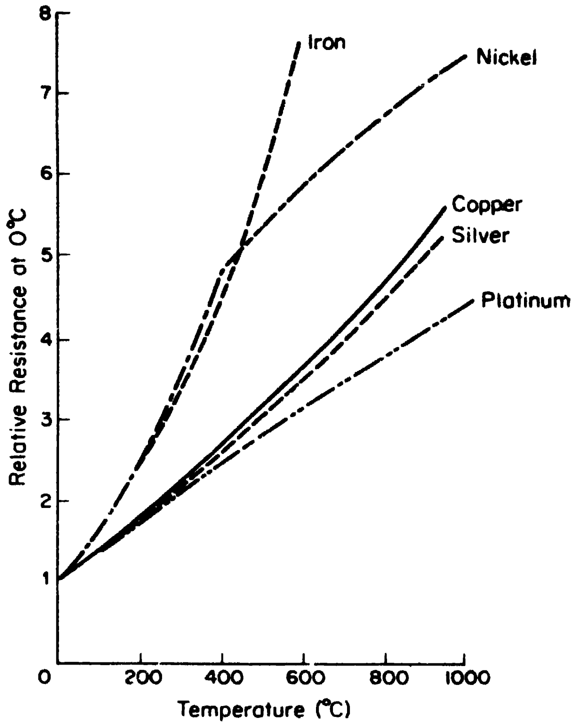


Figure 14-15
Relative resistance (R_t/R_{ref}) versus temperature for some pure metals.

Δt = difference between the reference temperature and the temperature of measurement.

Rewriting Eq. (14-11) and solving for t yields

$$R_t = R_{ref} + \alpha \Delta t R_{ref}$$

and

$$\Delta t = \frac{R_t - R_{ref}}{\alpha R_{ref}} = \frac{\Delta R}{\alpha R_{ref}} \tag{14-12}$$

where ΔR = the resistance change of the conductor.

The resistance change in the sensing element of the resistance thermometer is directly proportional to the change in its temperature. A Wheatstone bridge may be calibrated to indicate the *temperature* which created the resistance change rather than the resistance change itself.

In a typical resistance thermometer, the temperature responding wire is wound around a solid silver core which has a high thermal conductivity. When the tip of the silver core is in metallic contact with the well or housing, heat from the immediate surrounding is transferred rapidly by the metal-to-metal contact to the winding. The response is therefore fast and measurement accuracy is not affected significantly by changes in ambient temperature.

A typical bridge circuit with the resistance thermometer in the unknown

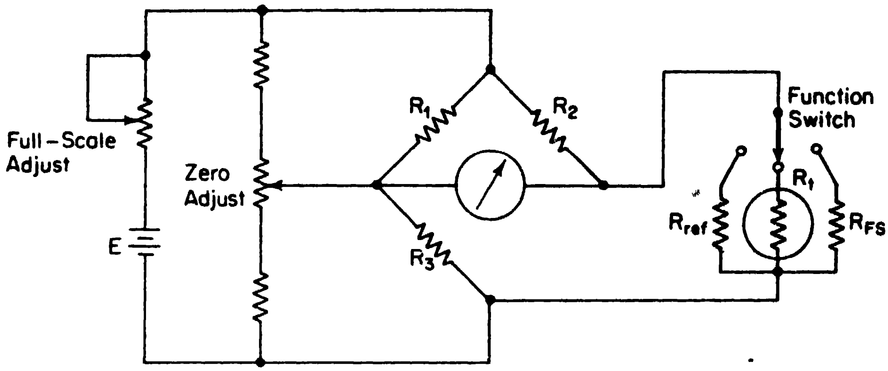


Figure 14-16
A bridge circuit with a resistance thermometer as one of the bridge elements.

position is shown in Fig. 14-16. The function switch connects three different resistors in the circuit. R_{ref} is a fixed resistor whose resistance is equal to that of the thermometer element at the reference temperature (say, 20°C). With the function switch in the "ref" position, the zero-adjust resistor is varied until the bridge indicator reads zero. R_{fs} is another fixed resistor, whose resistance equals that of the thermometer element for full-scale reading of the current indicator. With the function switch in the "fs" position, the full-scale-adjust resistor is varied until the indicator reads full scale. The function switch is then set to the "measure" position, connecting the resistance thermometer in the circuit. When the resistance-temperature characteristic of the thermometer element is linear, the galvanometer indication may be interpolated linearly between the set values of reference temperature and full-scale temperature.

The Wheatstone bridge has certain disadvantages when used to measure the resistance variations of the resistance thermometer. These are the effect of contact resistances of connections to the bridge terminals, heating of the elements by the unbalance current, and heating of the wires connecting the thermometer to the bridge. Slight modifications of the Wheatstone bridge, such as the double slidewire bridge, eliminate most of these problems. Despite these measurement problems, the resistance thermometer method is so accurate that it is one of the standard methods of temperature measurement within the range of −297°F to 1167°F.

(*b*) *Thermocouples.* When two dissimilar metal wires are twisted together and heated, a sensitive meter connected to the other end of the pair will indicate an emf which is very nearly proportional to the *difference* in temperature between the heated or *hot junction* and the other end, called the *cold junction.*

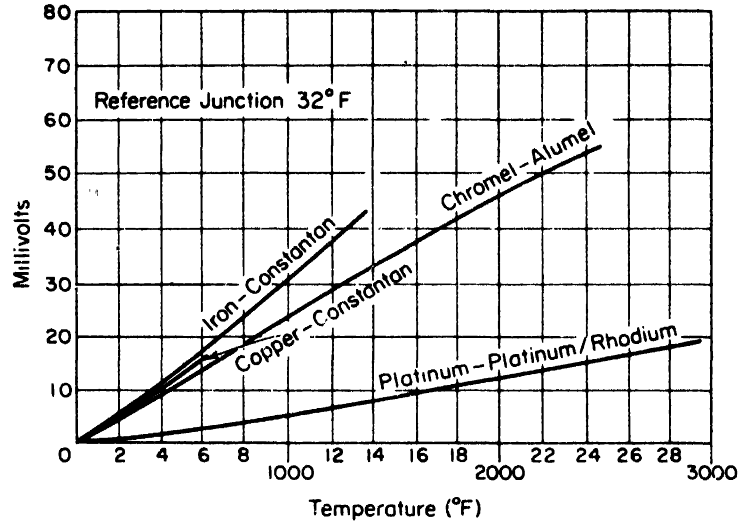


Figure 14-17
Response of some common industrial thermocouples, indicating a substantial linear rise in emf with increasing temperature.

Figure 14-17 shows the reponse of some common thermocouple combinations. In order to have a valuable industrial thermocouple, a rise in temperature should always produce a substantial *linear* rise in emf.

To insure long life in its operating environment, a thermocouple is protected in an open- or closed-end metal protecting tube or *well.* To prevent contamination of the couple when precious metals are used (platinum and its alloys), the protecting tube is both chemically inert and vacuum tight. The thermocouple is usually in a location *remote* from the measuring instrument. Connections are then made by means of special extension wires called *compensating* wires. Maximum accuracy of measurement is assured when t[1] compensating wires are of the same material as the thermocouple wires.

Thermocouples are calibrated with their cold junction maintained at some *reference* temperature, usually 32°F. Thermocouples are available from manufactures with an NBS certificate of calibration or with a certificate of test based on precision comparison against NBS-certified couples.

The simplest type of *temperature measurement* using a thermocouple is by connecting a sensitive millivoltmeter directly across the cold junction. The deflection of the meter is then almost directly proportional to the difference in temperature between the hot junction and the reference junction. This simple instrument has several serious limitations, mainly because the thermocouple can only supply very limited power to drive the meter movement.

The method most commonly used in thermocouple temperature measurements involves a *potentiometer*. A representative example of this instrument, specifically designed to measure thermocouple voltages, is shown in the photograph of Fig. 6-7 and its circuit diagram is given in Fig. 6-8 (see Sec. 6-3 for a description of this typical instrument).

Many types of *automatic* potentiometers have been developed both for automatic recording of temperatures on chart recorders and for automatic process control. One common system uses a *photoelectric* device, as shown in the simplified schematic diagram of Fig. 14-18. The photocell controls the position of the moving contact of a slidewire potentiometer. The advantage of this system is that the galvanometer is not subjected to a physical load since it is used only to direct *light* onto a photocell. The photocell receives its light

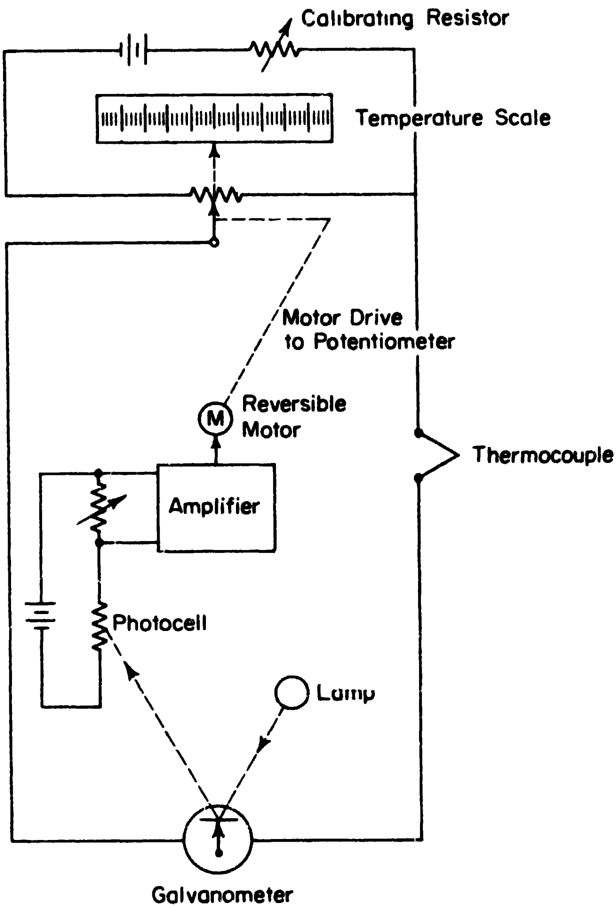


Figure 14-18
Automatic balancing potentiometer using a photocell to provide two-directional positioning of the potentiometer.

by reflection from the galvanometer mirror whose angular position is a measure of the unbalance voltage in the potentiometer circuit. The photocell is part of the input circuit to the amplifier, and its resistance controls the input voltage to the amplifier. The amplifier drives a reversible motor which controls the movement of the slider. When the potentiometer is initially balanced, for instance at the reference temperature, there will be a definite amount of light reflected to the photocell. If the light striking the photocell changes because of some change in temperature at the thermocouple junction, the resistance of the photocell changes and affects the input to the amplifier. The amplifier then drives the reversible motor in a direction which tends to restore balance in the potentiometer circuit.

A fairly standard method of dealing with small dc error signals is shown in Fig. 14-19. The thermocouple is again placed in a potentiometer circuit. The

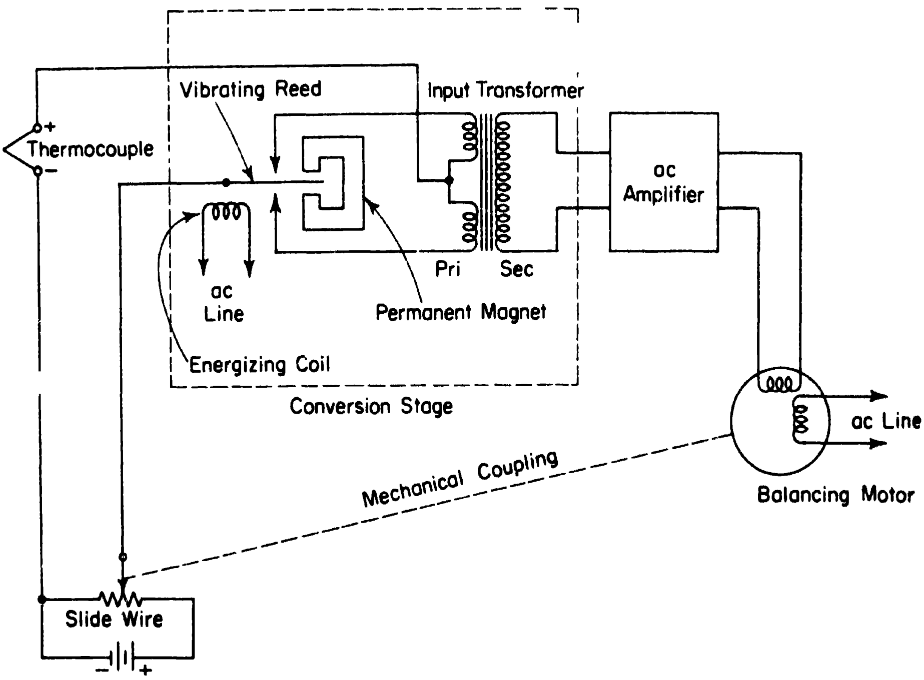


Figure 14-19
Automatic balancing potentiometer using a converter to change the dc error signal into ac for amplification and use.

unbalance voltage in the potentiometer, caused by temperature variations at the thermocouple hot junction, is applied to a converter. When power is applied to the driving coil of the converter (usually at 60 Hz or 400 Hz), the magnetic armature vibrates in synchronism with the frequency of the coil voltage.

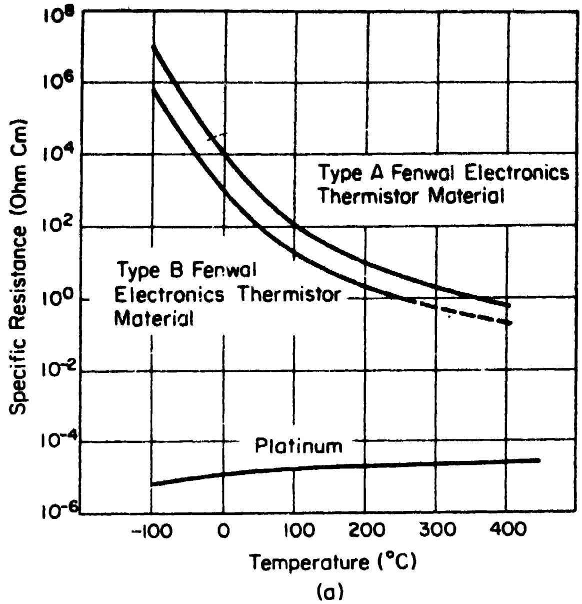
The armature alternately connects opposite ends of the transformer primary to the unbalance voltage. The pulsating unbalance voltage in the transformer primary is transferred to the secondary where it is amplified by an ac amplifier and applied to the balancing motor. The polarity of the error signal originating from the potentiometer circuit determines whether the pulse going into the transformer is of one polarity or of the opposite polarity. The polarity of the amplified error signal with respect to the instantaneous line voltage determines the direction of rotation of the balancing motor. The motor is mechanically coupled to the slider of the potentiometer and drives it in a direction to restore circuit balance.

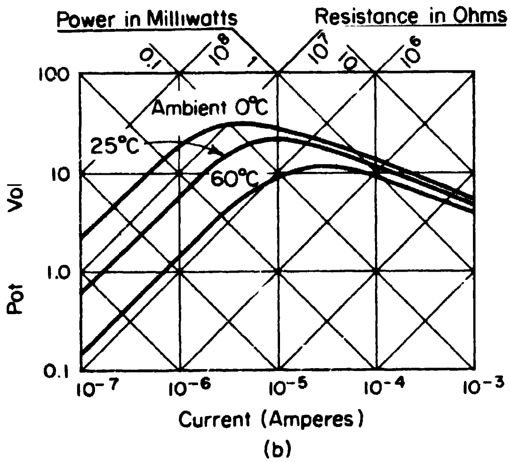
(*c*) *Thermistors*. Thermistors or *thermal resistors* are essentially semiconductors which behave as resistors with a high, usually *negative,* temperature coefficient of resistance. In some cases the resistance of a thermistor at room temperature may decrease as much as 6 per cent for each 1°C rise in temperature. This high sensitivity to temperature change makes the thermistor extremely well suited to precision *temperature measurement, control,* and *compensation.* Thermistors are therefore widely used in such applications, especially in the lower temperature range of −100°C to 300°C.

Thermistors are composed of a sintered mixture of metallic oxides, such as manganese, nickel, cobalt, copper, iron, and uranium. Their resistances range from 0.5 Ω to 75 MΩ and they are available in a wide variety of shapes and sizes. Smallest in size are the beads with a diameter of 0.006 in. to 0.05 in. Beads may be sealed in the tips of solid glass rods to form probes which are somewhat easier to mount than beads. Disks and washers are made by pressing thermistor material under high pressure into flat cylindrical shapes with diameters from 0.1 in. to 1 in. Washers can be stacked and placed in series or in parallel for increased power dissipation.

Three important characteristics of thermistors make them extremely useful in measurement and control applications: (1) the *resistance-temperature* characteristic, (2) the *voltage-current* characteristic, (3) the *current-time* characteristic. Representative examples of these characteristic curves are shown in Fig. 14-20.

The *resistance-temperature* characteristic of Fig. 14-20(a) shows that the thermistor has a very high negative temperature coefficient of resistance, making it an ideal *temperature transducer*. The resistance variations with temperature of the two industrial materials are compared to the characteristics for platinum (a widely used resistance thermometer material). Between the temperatures of −100°C and +400°C, the resistance of type A thermistor material changes from 10^7 to 1 Ω-cm, while the resistance of the platinum varies only by a factor of approximately 10 over the same temperature range.





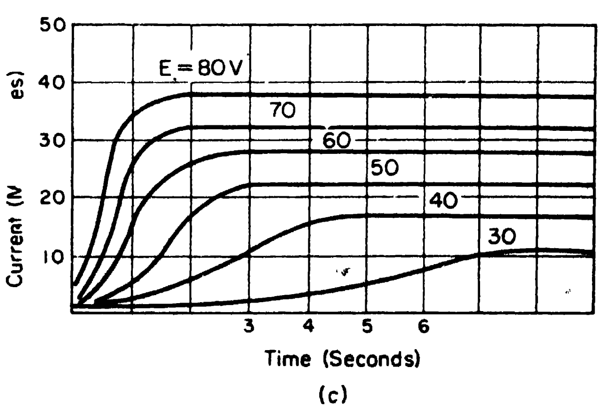


Figure 14-20

Characteristic curves of thermistors. (a) Resistance-temperature characteristic. (b) Voltage-current characteristic. (c) Current-time characteristic. (*Courtesy Fenwal Electronics, Inc.*)

The *voltage-current characteristic* of Fig. 14-20(b) shows that the voltage drop across a thermistor increases with increasing current until it reaches a peak value beyond which the voltage drop decreases as the current increases. In this portion of the curve, the thermistor exhibits a *negative resistance* characteristic. If a very small voltage is applied to the thermistor, the resulting small current does not produce sufficient heat to raise the temperature of the thermistor above ambient. Under this condition, Ohm's law is followed and the current is proportional to the applied voltage. Larger currents, at larger applied voltages, produce enough heat to raise the thermistor above the ambient temperature and its resistance then decreases. As a result, more current is then drawn and the resistance decreases further. The current continues to increase until the heat dissipation of the thermistor equals the power supplied to it. Therefore, under any fixed ambient conditions, the resistance of a thermistor is largely a function of the power being dissipated within itself, provided that there is enough power available to raise its temperature above ambient. Under such operating conditions, the temperature of the thermistor may rise 100°C or 200°C and its resistance may drop to one-thousandth of its value at low current.

This characteristic of *self-heat* provides an entirely new field of uses for the thermistor. In the self-heat state, the thermistor is sensitive to anything that changes the *rate* at which heat is conducted away from it. It can so be used to measure *flow, pressure, liquid level, composition of gases,* etc. If, on the other hand, the rate of heat removal is fixed, then the thermistor is sensitivie to power input and can be used for *voltage* or *power-level* control.

The *current-time characteristic* curve of Fig. 14-20(c) indicates the time delay to reach maximum current as a function of the applied voltage. When the self-heating effect just described occurs in a thermistor network, a certain finite time is required for the thermistor to heat and the current to build up to a maximum steady-state value. This time, although fixed for a given set of circuit parameters, may easily be varied by changing the applied voltage or the series resistance of the circuit. This time-current effect provides a simple and accurate means of achieving time delays from milliseconds to many minutes.

Although thermistors are probably best known for their function in the measurement and control of temperature, they may be used in a variety of other applications. A few of the more common applications are described in the following sections.

(1) Temperature Measurement. The thermistor's relatively large resistance change per degree change in temperature (called the *sensitivity*) provides good accuracy and resolution. A typical industrial-type thermistor with a 2,000-Ω resistance at 25°C and a temperature coefficient of 2.9 per cent/°C, will exhibit a resistance change of 78 Ω per degree C change in temperature.

When this thermistor is connected in a simple series circuit consisting of a battery and a microammeter, any variation in temperature causes a change in thermistor resistance and a corresponding change in circuit current. The meter can be calibrated directly in terms of temperature and may be able to resolve temperature variations of 0.1°C. Higher sensitivity may be obtained by using the bridge circuit of Fig. 14-21. The 4-kΩ thermistor will readily indicate a temperature change of as little as 0.005°C.

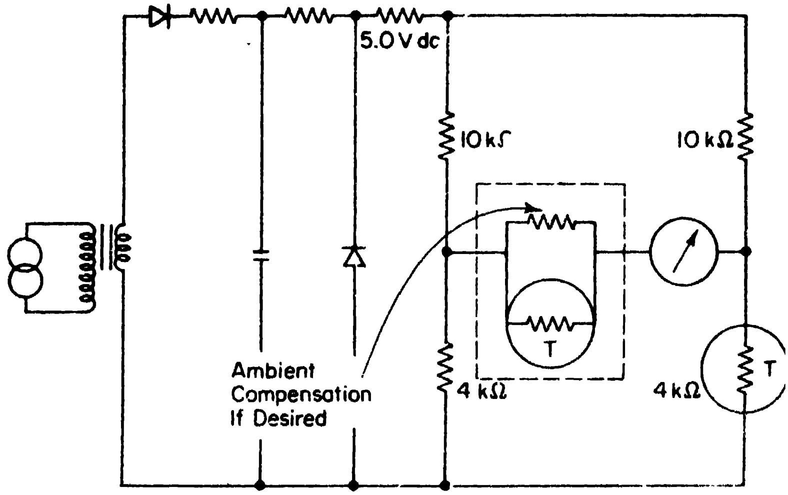


Figure 14-21
Typical thermistor circuit using a bridge configuration.

This high sensitivity, together with the relatively high thermistor resistance which may be selected (for example, 100 kΩ), makes the thermistor ideal for *remote* measurement or control, since changes in contact or transmission line resistance due to ambient temperature effects are negligible. For example, 400 ft of No. 18 AWG copper wire transmission line, subjected to a 25°C temperature change, will affect the accuracy of measurement or control approximately by 0.05°C.

(2) Temperature Control. A simple temperature control circuit may be constructed by replacing the microammeter in the bridge circuit of Fig. 14-21 with a relay. This is indicated in the typical thermistor temperature control circuit of Fig. 14-22 where a 4-kΩ thermistor is connected in an ac-excited bridge. The unbalance voltage is fed to an ac amplifier whose output drives a relay. The relay contacts are used to control the current in the circuit which generates the heat. These control circuits can be operated to a precision of 0.0001°F.

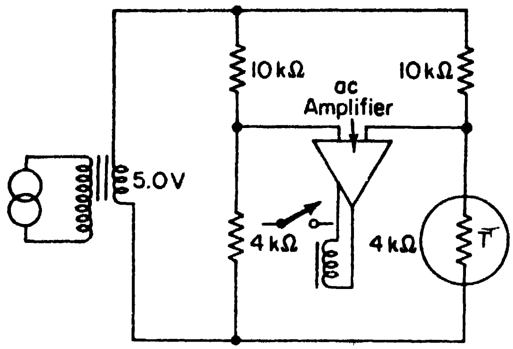


Figure 14-22
Typical thermistor temperature control circuit.

Thermistor control systems are inherently sensitive, stable, and fast acting and require relatively simple circuitry. The voltage output of the standard thermistor bridge circuit at 25°C will be approximately 18 mV/°C using a 4,000-Ω thermistor in the configuration of Fig. 14-21.

(3) Temperature Compensation. Because thermistors have a negative temperature coefficient of resistance—opposite to the positive coefficient of most electrical conductors and semiconductors—they are widely used to compensate for the effects of temperature on both component and circuit performance. Disk-type thermistors are used for this purpose where the maximum temperature does not exceed 125°C. A properly selected thermistor, mounted against or near a circuit element, such as a copper meter coil, and experiencing the same ambient temperature changes, can be connected in such a way that the total circuit resistance is constant over a wide range of temperatures. This is shown in the curves of Fig. 14-23, which illustrates the effect of a compensation network.

The compensator consists of a thermistor, shunted by a resistor. The negative temperature coefficient of this combination equals the positive coefficient of the copper coil. The coil resistance of 5,000 Ω at 25°C, varies from approximately 4,500 Ω at 0°C to 5,700 Ω at 60°C, representing a change of about ± 12 per cent. With a single thermistor compensation network, this variation is reduced to about ± 15 Ω or $\pm \frac{1}{4}$ per cent. With double or triple compensation networks, variations can be reduced even further.

(4) Thermal Conductivity Measurements. In this application, two thermistors are normally connected in adjacent legs of a Wheatstone bridge, as shown in Fig. 14-24. The bridge supply voltage is high enough to raise the thermistors above ambient temperature, typically to about 150°C. One thermistor is mounted in a static area to provide temperature compensation while the other is placed in the medium to be measured. Any change in the thermal

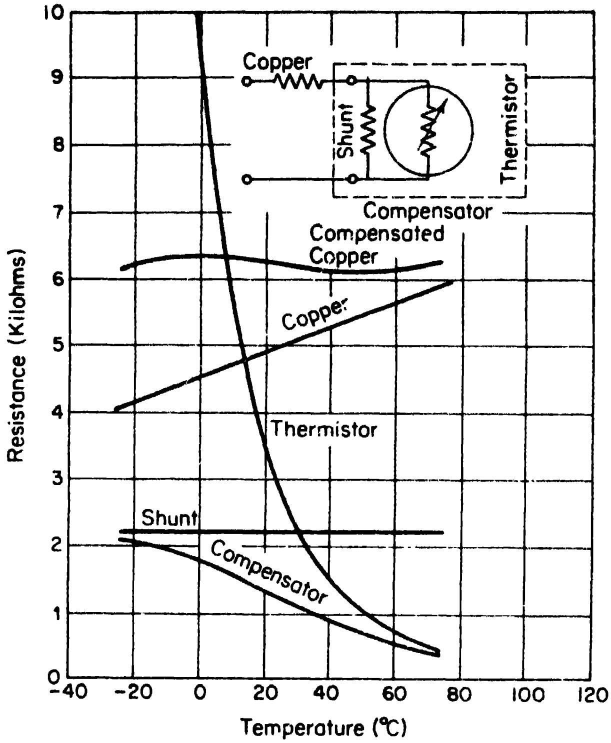


Figure 14-23
Temperature compensation of a copper conductor by means of a thermistor network.

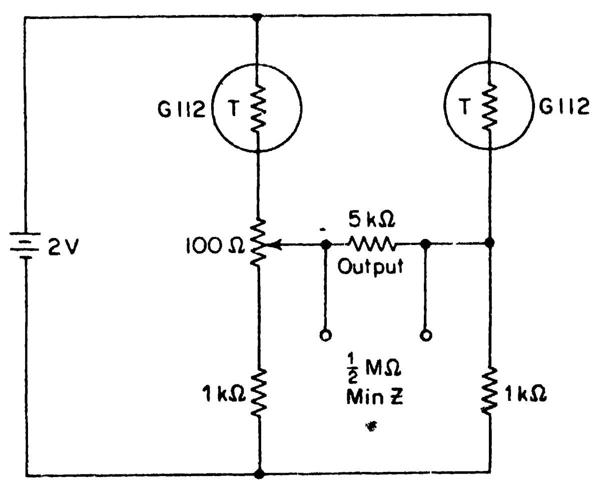


Figure 14-24
Typical thermistor thermal conductivity measurement.

conductivity of this medium will change the rate at which heat is dissipated from the sensing thermistor, thus changing its temperature. This results in bridge unbalance, which can be calibrated in appropirate units.

In a typical application, two thermistors are placed in separate cavities in a brass block. With air in both cavities, the bridge is balanced. When the air in one cavity is replaced by pure carbon dioxide, which has a lower conductivity than air, the bridge will be unbalanced because that thermistor becomes hotter and lower in resistance. This amount of unbalance represents 100 per cent CO_2 in the analyzer; 50 per cent CO_2 gives just half the meter reading, and the instrument may be calibrated with a linear scale to read per cent CO_2 in air. Similar calibration may be made for any other mixture of two gases.

If the same bridge uses one thermistor sealed in a cavity in a brass block and the other thermistor mounted in a small pipe, it may be used as a *flow meter*. When no air is flowing through the pipe, the bridge may be balanced. When air flows through the pipe, the thermistor is cooled, and its resistance increases which unbalances the bridge. The amount of cooling is proportional to the rate of flow of the air and the meter may be calibrated in terms of flow in the pipe. Such instruments have been made to measure flow rates as low as 0.001 cm^3 per minute.

14-6 Photosensitive Devices

(*a*) *General Information*. Photosensitive elements are versatile tools for detecting radiant energy or light. They exceed the sensitivity of the human eye to all the colors of the spectrum and operate even into the ultraviolet and infrared regions.

The photosensitive device has found practical use in many engineering applications. This section deals with the following devices and their applications:

(1) *Vacuum-type phototubes,* used to best advantage in applications requiring the observation of light pulses of short duration, or light modulated at relatively high frequencies.

(2) *Gas-type phototubes,* in extensive use in the motion picture industry as a sound-on-film sensor.

(3) *Multiplier phototubes,* with tremendous amplifying capability, are used extensively in photoelectric measurement and control devices and also as scintillation counters.

(4) *Photoconductive cells,* also known as *photoresistors* or *light-dependent resistors,* find wide use in industrial and laboratory control applications.

(5) *Photovoltaic cells* are semiconductor junction devices used to convert radiation energy into electrical power. A fine example is the *solar cell,* used in space engineering applications.

(*b*) *The Vacuum Photobube.* This device was described briefly in Sec. 14-4(f), where an application of the phototube as a pressure transducer was given. The construction details are therefore omitted here.

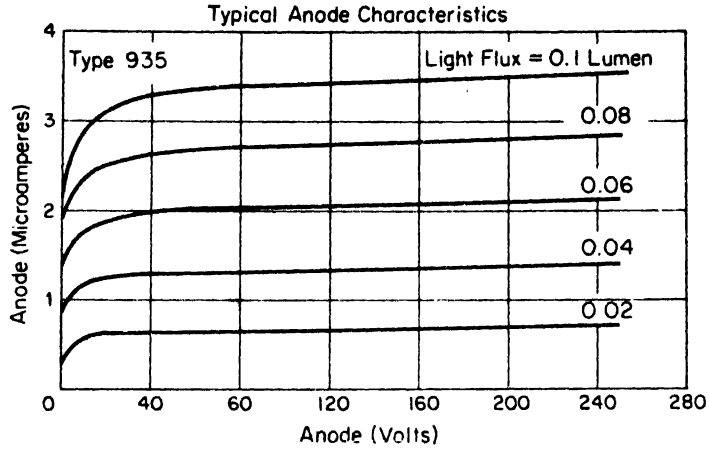
The photocathode emits electrons when stimulated by incident radiant energy. The most important photocathode now used in vacuum phototubes is the cesium-antimony surface, which is characterized by high sensitivity in the visible spectrum. The type of glass used in the glass envelope determines mainly the sensitivity of the device at other wavelengths. Usually the glass cuts off the transmitted radiation in the ultraviolet region.

Typical voltage-current characteristics are shown in Fig. 14-25(a). When sufficient voltage is applied between the photocathode and the anode, the collected current is almost dependent only on the amount of incident light. Vacuum phototubes are characterized by a photocurrent response which is linear over a wide range, so much so that these tubes are frequently used as standards in light-comparison measurements. Figure 14-25(b) shows the linear current-light relationship.

(*c*) *The Gas-Filled Phototube.* This phototube has the same general construction as the vacuum phototube, except that the envelope contains inert gas (usually argon) at a very low pressure. Electrons are emitted from the cathode by photoelectric action and accelerate through the gas by the applied voltage at the anode. If the energy of the electrons exceeds the ionization potential of the gas (15.7 V for argon), the collision of an electron and a gas molecule can result in ionization, i.e., the creation of a positive ion and a second electron. As the voltage is further increased above the ionization potential, the current collected by the anode increases owing to the higher number of collisions between photoelectrons and gas molecules. If the anode voltage is raised to a very high value, the current becomes uncontrolled; all the gas molecules are then ionized and the tube exhibits a glow discharge. This condition should be avoided as it may permanently damage the phototube. Typical current-voltage characteristics for various light levels are given in Fig. 14-26.

(*d*) *Multiplier Phototubes.* To detect very low light levels, special amplification of the photocurrent is necessary in most applications. The multiplier phototube or photomultiplier uses secondary emission to provide current amplification in excess of a factor 10^6 and then becomes a very useful detector for low light levels.

In a photomultiplier, the electrons emitted by the photocathode are elec-



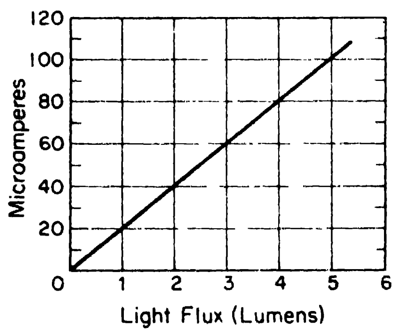


Figure 14-25

Characteristics of a vacuum-type phototube. (a) Typical anode characteristics. (b) Linearity of current as a function of light.

trostatically directed toward a secondary emitting surface, called a *dynode*. When the proper operating voltage is applied to the dynode, three to six secondary electrons are emitted for every primary electron striking the dynode. These secondary electrons are focused to a second dynode, where the process is repeated. The original emission from the photocathode is therefore *multiplied* many times.

In Fig. 14-27 a photomultiplier with ten dynodes is shown. The last dynode (10) is followed by the anode (11) which collects the electrons and serves as the signal output electrode in most applications. The dynodes may be arranged in different patterns. One of the early designs is the circular-cage type of Fig. 14-28, a compact unit having nine dynodes. The photocathode being rather

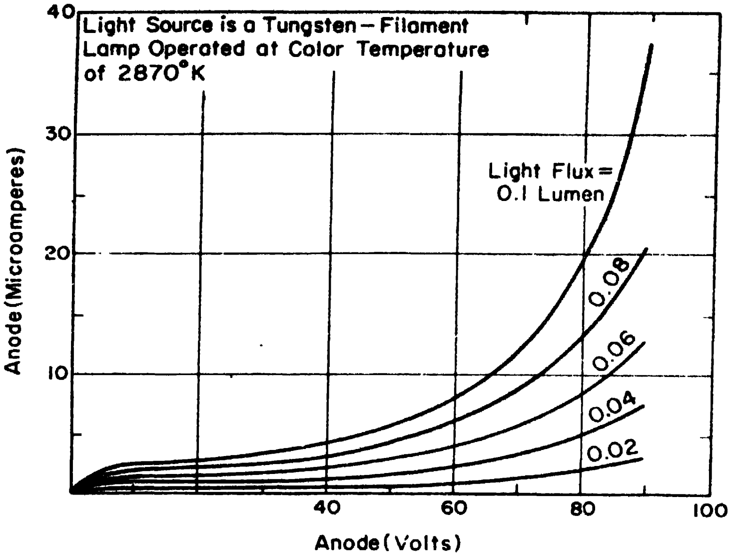


Figure 14-26
Characteristics of a gas-type phototube.

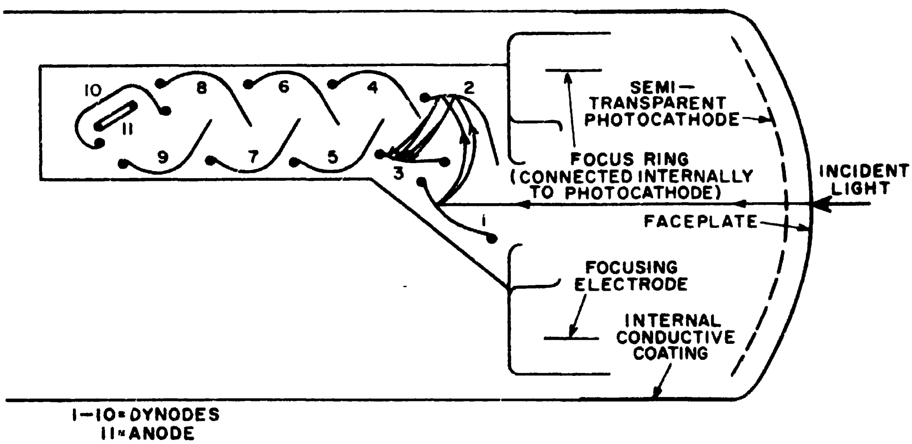


Figure 14-27
Schematic view of a linear photomultiplier. The focusing electrodes direct the photocathode emission toward the first dynode. Details of this focusing arrangement are given in Fig. 14-29. (*Courtesy Radio Corporation of America.*)

small limits its use somewhat. For scintillation counting, for instance, it is desirable to have a large flat photocathode for efficient coupling of the tube to the scintillation crystal.

The linear type of photomultiplier (also known as the Matheson tube) of Fig. 14-27 has a specially designed focused cage structure with a large effec-

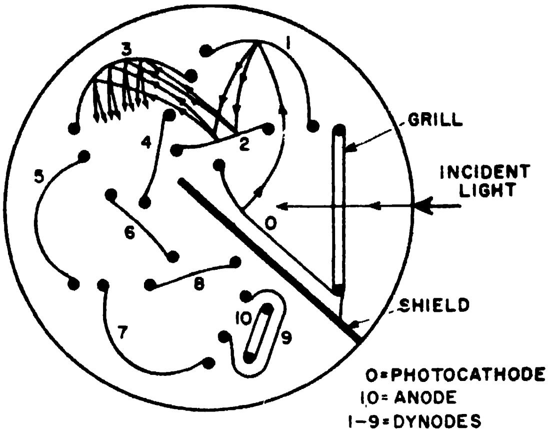


Figure 14-28
A photomultiplier of circular compact construction. (*Courtesy Radio Corporation of America.*)

tive area for the collection of photoelectrons on the first dynode. The Matheson solution, shown in Fig. 14-29, uses a curved cathode and annular rings for electrostatic focusing of the photoelectrons. This construction results in very effective collection of photoelectrons and also in very short transit times (high-frequency response).

The gain of the photomultiplier depends on the number of dynodes used and the properties of the dynode material. For a typical ten-dynode tube, such as shown in Fig. 14-27, the gain would be in the order of 10^6 with an applied voltage of 100 V per stage (a 1,000-V supply source would be needed in this case). The spectral response may be controlled by the material of which the cathode and the dynodes are made. The output of the multiplier is linear, similar to the output of the vacuum phototube.

Magnetic fields affect the gain of the photomultiplier because some electrons may be deflected from their normal path between stages and therefore never reach a dynode nor, eventually, the anode. In scintillation-counting applications this effect may be disturbing, and mu-metal magnetic shields are often placed around the photomultiplier tube.

(*e*) *Photoconductive Cells*. Photoconductive cells are elements whose conductivity is a function of the incident electromagnetic radiation. Many materials are photoconductive to some degree, but the commercially more important ones are cadmium sulfide, germanium, and silicon. The spectral response of the cadmium sulfide cell closely matches that of the human eye, and the cell is therefore often used in applications where human vision is a factor, such as street light control or automatic iris control for cameras.

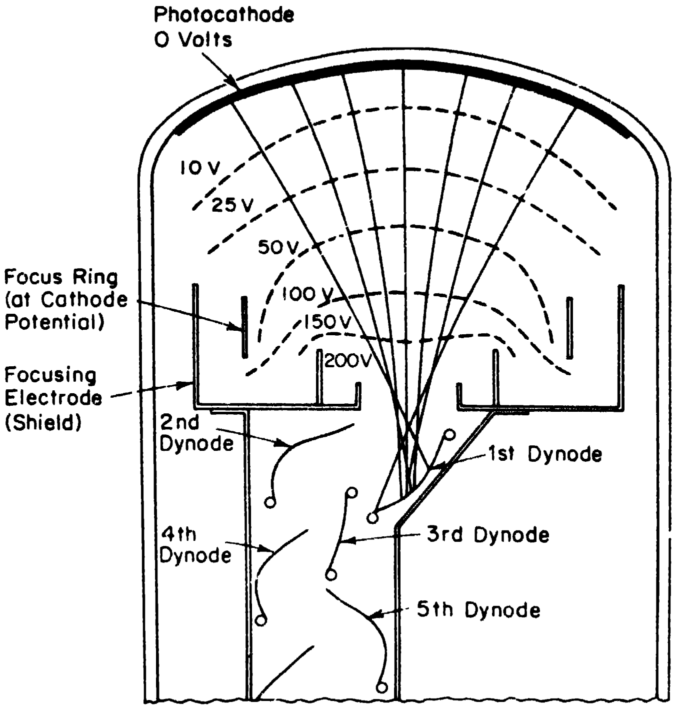


Figure 14-29
The Matheson-type front end configuration, showing equipotential lines and electron trajectories, feeding into a linear dynode cage. (*Courtesy Radio Corporation of America.*)

The essential elements of a photoconductive cell are the ceramic substrate, a layer of photoconductive material, metallic electrodes to connect the device into a circuit, and a moisture-resistance enclosure. A cutaway view of a typical photoconductive cell is shown in Fig. 14-30.

A typical application of a practical *on-off* photocell control circuit is given in Fig. 14-31. Resistors R_2, R_3, and R_4 are chosen so that the emitter-to-base bias of Q_2 is sufficiently positive to allow Q_2 to conduct. As a result, the relay in Q_2 collector circuit will be energized. When configuration A is used as the control circuit, the relay is energized when the light on the photocell is below a predetermined level. When the photocell is illuminated, the emitter-to-base bias of Q_1 becomes sufficiently positive to allow Q_1 to conduct. Its collector potential becomes considerably less positive, decreasing the bias on Q_2 and Q_2 cuts off, deenergizing the relay. When configuration B is used, the relay will be energized when the light incident on the photocell is above a predetermined level.

Semiconductor junction photocells are used in some applications. The

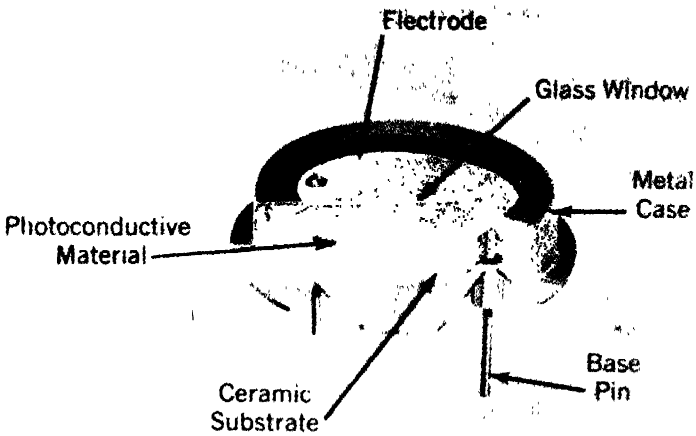


Figure 14-30
Cutaway view of a photoconductive cell. (*Courtesy Radio Corporation of America.*)

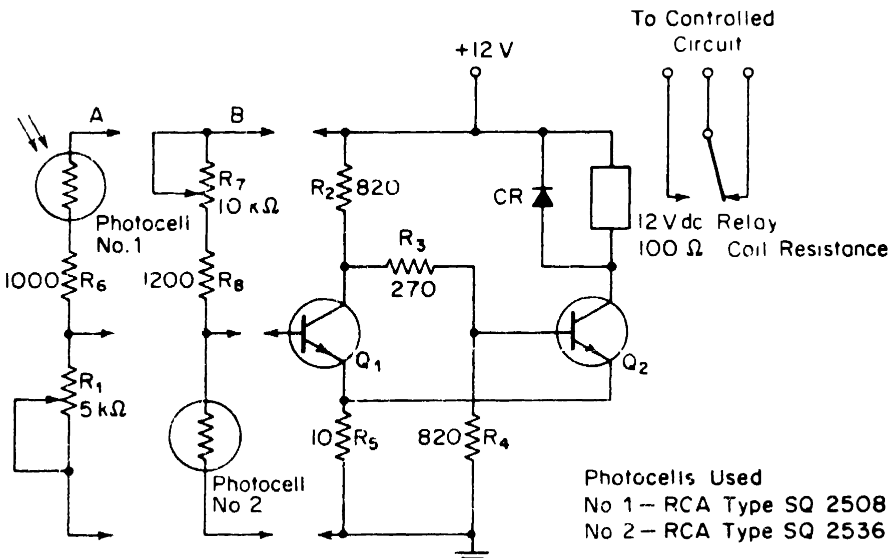


Figure 14-31
12-V photocell control circuit. (*Courtesy RCA, Electronic Components and Devices Division.*)

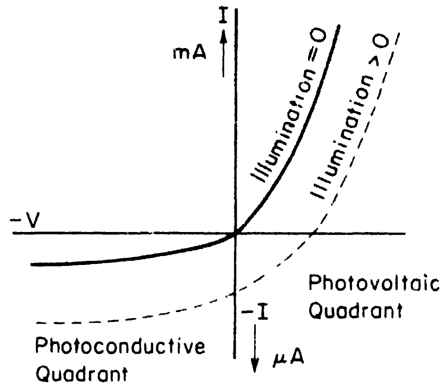


Figure 14-32
Current-voltage characteristics of a photojunction diode.

volt-ampere characteristics of a p-n junction may appear as the solid line in Fig. 14-32, but when light is applied to the cell, the curve shifts downward, as shown by the dashed line.

In photoconductive applications, the cell is biased in the reverse direction. When the cell is illuminated, the reverse current increases and an output voltage can be developed across an output resistor. This output voltage is then proportional to the amount of incident light. A typical order of magnitude for the increase in output current is approximately 0.7 μA for each 1 footcandle increase in illumination. This increase in photocurrent is linear with an increase in illumination.

The time constant of p-n junction photocells is relatively fast, making this device useful for optical excitation frequencies well above the audio range.

(*f*) *Photovoltaic Cells.* Photovoltaic cells may be used in a number of applications. 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 square, into which a very thin (0.5 micron) layer of *n* type material is diffused. The conversion efficiency depends on the spectral content and the intensity of the illumination.

Multiple-unit silicon photovoltaic 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 photovoltaic devices in the infrared region of the spectrum and may be used as infrared detectors.

14-7 Magnetic Measurements

(*a*) *The Ballistic Galvanometer.* The deflection of a ballistic galvanometer is directly proportional to the electric charge flowing through its coil. Since charge and flux are related by a constant of proportionality, the galvanometer deflection is a measure of the flux so that

$$\phi = K\theta \quad \text{(weber)} \tag{14-13}$$

where ϕ = magnetic flux, in webers
K = constant of proportionality
θ = angular deflection of the galvanometer, in radians.

The instrument must be calibrated in the circuit in which the flux measurement is to be made. One calibration technique is described in Sec. 3-5, and the actual flux measurement setup is shown in Fig. 3-8. (See Sec. 3-5 for further information about the mathematical considerations of this flux measurement.)

To examine the properties of magnetic materials, one single flux measurement is often insufficient. The measurement setup of Fig. 14-33 allows deter-

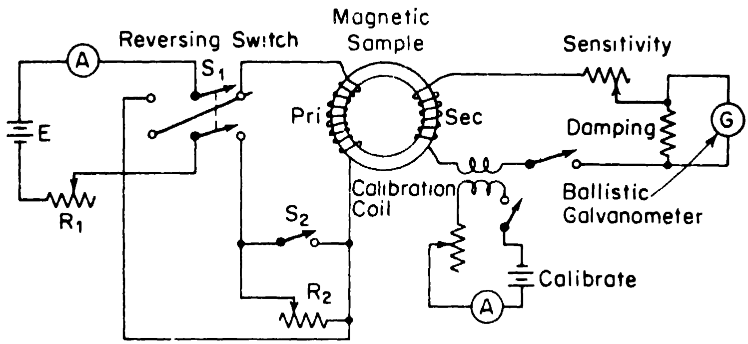


Figure 14-33
A circuit using a ballistic galvanometer to determine the hysteresis curve of a magnetic sample.

mination of a *hysteresis loop* of a ring sample of magnetic material by measuring the flux with a ballistic galvanometer at different values of the magnetizing force. The calibration coil of Fig. 14-33 is used to determine the galvanometer sensitivity K of Eq (3-9). The hysteresis loop or BH curve is plotted by measuring B with the galvanometer for different values of H. In the metric system (SI) H is measured in AT/m. The current through the primary winding on the ring sample is controlled by rheostat R_1. Since the number of primary turns and the mean circumference of the sample are known, H can be set to any desired value with R_1. The hysteresis loop is measured in the following manner: Switch S_2 is initially closed and the current in the primary winding is set by R_1 to the desired value of maximum H. Switch S_1 is reversed several times so that the sample is in a cyclic condition and the deflections of the ballistic galvanometer are read. The average value of the repeated measurements yields the value for maximum flux density B. Switch S_2 is now opened, which shunts R_2 into the current circuit and reduces the magnetizing force by a small amount. The reduction in flux density ΔB is obtained from the galvanometer deflection and the new value of H from the meter reading. Several measurements are made by manipulation of the reversing switch S_1, and the average value of ΔB is found. Now S_2 is closed again and the sample is returned to its initial position of maximum magnetization. Rheostat R_2 is now slightly adjusted to decrease the total magnetization current and a new set of measurements is made, starting from the initial point of maximum H.

(*b*) *The Flux Meter and the Gaussmeter.** The *flux meter* uses a special moving coil movement which lacks an internal magnet and pole pieces. This meter is placed in the unknown magnetic field and a current is passed through the meter. The deflection of the flux meter then depends on the amount of current and the strength of the unknown magnetic field. The amount of current can be controlled with a rheostat and read on another meter whose scale is calibrated in gauss, or weber. The amount of current for a standard deflection on the flux meter is then directly proportional to the magnetic fieldstrength and the current reading is a direct indication of the magnetic fieldstrength.

The *gaussmeter* operates on a different principle. The torque exerted on a small magnet by magnetic induction is balanced by the restoring torque of a spiral spring. The small magnet is brought into the influence of the unknown magnetic field and rotated for maximum indication of a pointer connected to the restoring spiral spring. The instrument scale is calibrated to read magnetic fieldstrength directly in gauss, or weber. (For detailed descriptions and discussion of these special instruments, see appropriate texts.)

(*c*) *Magnetic Transducers. Bismuth* and *mu-metal* have the property of changing their resistance or impedance if placed in a *transverse* magnetic field. This effect may be used to measure flux density with a conventional Wheatstone bridge. The method is indicated in Fig. 14-34 where two mu-metal wires are placed in the unknown magnetic field. The impedance of the wires is a function of the strength of the magnetic field and is measured with an ac bridge. This transducer produces an acceptable emf for the input to an instrumentation system and is capable of measuring fieldstrength in the order of milligauss.

A slight variation of the circuit of Fig. 14-34 uses two bismuth spirals in opposite arms of the bridge, The resistance of the bismuth changes with the strength of the magnetic field and this change is measured with a dc bridge.

The *Hall effect* transducer uses a strip of semiconductor material which is exposed to the unknown magnetic field. When a strip of conductor or semicon-

*Stout, *op. cit.*, pp. 442–446, 459.

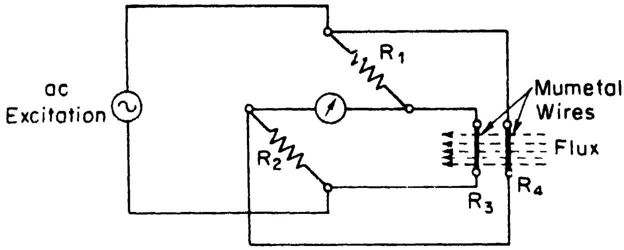


Figure 14-34
An ac mu-metal bridge to measure small current magnetic fields.

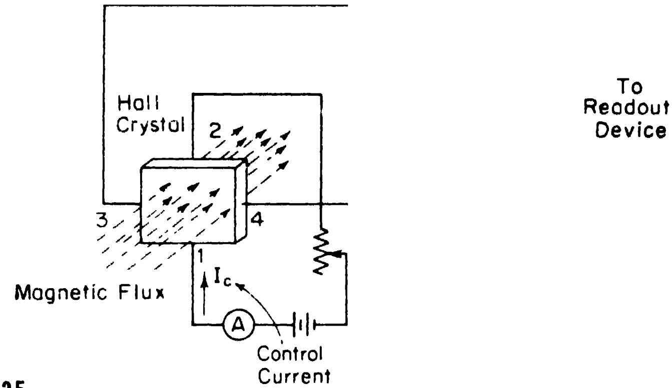


Figure 14-35
A circuit for measuring magnetic flux by using the Hall effect of a germanium crystal.

ductor material carries current in the presence of a transverse magnetic field, an emf is produced between the opposite edges of the semiconductor strip, as shown in Fig. 14-35. Leads 1 and 2 conduct current through a strip of germanium. Leads 3 and 4 are at the same potential when there is no transverse magnetic field passing through the strip. When there is a magnetic flux through the strip, the voltage between leads 3 and 4 is proportional to the product of the current and the fieldstrength. A number of materials exhibit the Hall effect but in many cases it is so small that it has no practical value. Germanium can be manufactured with a very high Hall coefficient and germanium probes to measure magnetic flux are used for small flux densities.

References

1. Stout, Melville B., *Basic Electrical Measurements*, 2nd ed., Chap. 16. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

2. Bartholomew, Davis, *Electrical Measurements and Instrumentation,* Chap. 11. Boston: Allyn and Bacon, Inc.,

3. Partridge, G. R., *Principles of Electronic Measurements,* Chap. 13. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1958.

4. Perry, C. C., and H. R. Lissner, *The Strain Gage Primer,* 2nd ed. New York: McGraw-Hill Book Company

5. Lion, Kurt S., *Instrumentation in Scientific Research.* New York: McGraw-Hill Book Company, 1959.

6. Fribance, Austin E., *Industrial Instrumentation Fundamentals,* Chaps. 10, 12, 15, 16. New York: McGraw-Hill Book Company,

7. Minnar, E. J. (ed.), *Instrument Society of America Transducer Compendium.* New York: Plenum Press,

8. *Capsule Thermistor Course,* Fenwal Electronics, Inc., Framingham, Mass., n. d.

9. *Phototubes and Photocells,* Technical Manual PT 60, Radio Corporation of America, Electronic Components and Devices, Lancaster, Pa., n. d.

10. *Introduction to Transducers for Instrumentation,* Statham Instruments, Inc., Los Angeles, Calif., n. d.

Problems

1. Name four types of electrical pressure transducers and describe one application of each type.

2. Under what conditions is a "dummy" strain gage used and what is the function of that gage?

3. What is the difference between a photoemissive, a photoconductive, and a photovoltaic cell? Name one application for each cell.

4. A resistance strain gage with a gage factor of 2.4 is mounted on a steel beam whose modulus of elasticity is 30×10^6 lb/in.^{2}. The strain gage has an unstrained resistance of 120.0 Ω which increases to 120.1 Ω when the beam is subjected to a stress. Calculate the stress at the point where the strain gage is mounted.

5. The unstrained resistance of each of the four elements of the unbonded strain gage of Fig. 14-2 is 120 Ω. The strain gage has a strain factor of 3 and is subjected to a strain ($\Delta l/l$) of 0.0001. If the indicator is a high-impedance voltmeter, calculate the reading of this voltmeter for a battery voltage of 10 V.

6. The high-impedance voltmeter of Problem 5 is replaced with a 200-Ω

galvanometer with a current sensitivity of 0.5 mm/μA. Calculate the galvanometer indication, in millimeters for the situation described in Problem 5.

7. The linear variable differential transformer (LVDT) of Fig. 14-6 produces an output of 2 V rms for a displacement of 20×10^{-6} in. Calculate the sensitivity of the LVDT in μV/mm. The 2-V output of the LVDT is read on a 5-V voltmeter which has a scale with 100 divisions. The scale can be read to 0.2 division. Calculate the resolution of the instrument in terms of displacement in inches.

FIFTEEN

ANALOG AND DIGITAL DATA-ACQUISITION SYSTEMS

15-1 Instrumentation Systems

Data-acquisition systems are used to measure and record analog signals obtained in basically two different ways: (a) Signal originating from *direct measurement* of electrical quantities; these signals may include dc and ac voltages, frequency, or resistance and are typically found in such areas as electronic component and subassembly testing or environmental and quality analysis work. (b) Signals originating from *transducers*; some of the more common transducers are strain gages and thermocouples (see Chapter 14).

Instrumentation systems can be categorized into two basic classes: digital systems and analog systems. *Analog systems* deal with measurement information in analog form. An *analog signal* may be defined as a continuous function, such as a plot of voltage versus time or displacement versus pressure. *Digital systems* handle information in digital form. A *digital quantity* may consist of a number of discrete and discontinuous pulses whose time relationship contains information about the magnitude or the nature of the quantity.

An *analog data-acquisition system* typically consists of some or all of the following elements:

(1) *Transducers* for translating physical parameters into electrical signals.

(2) *Signal conditioners* for amplifying, refining, or selecting certain portions of these signals.

(3) *Visual display devices* for continuous monitoring of the input signals. These devices may include single or multichannel CROs, storage CROs, panel meters, numerical displays, and so on.

(4) *Graphic recording instruments* for obtaining permanent records of the input data. These instruments include heated stylus and ink recorders providing continuous records on paper charts, optical recording systems, such as mirror galvanometer recorders, and ultraviolet recorders.

(5) *Magnetic tape instrumentation* for acquiring all input data, preserving their original electrical form and reproducing them at a later date for more detailed analysis. [Standards for instrumentation magnetic tape recorders were established within the field of telemetry by the Inter Range Instrumentation Group (IRIG) and these standards are very rigidly adhered to throughout the instrumentation industry to provide compatibility and exchange of recorded data between various recording systems.]*

A *digital data-acquisition system* may include some or all of the elements shown in Fig. 15-1. The essential functional operations within a digital system include handling analog signals, making the measurement, converting and handling digital data, and internal programming and control. The function of each of the system elements of Fig. 15-1 is listed below.

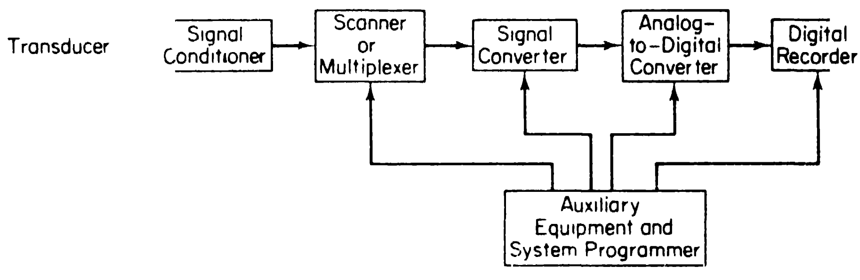


Figure 15-1
Elements of a digital data-acquisition system.

Transducer. Translates physical parameters to electrical signals acceptable by the acquisition system. Some typical parameters include temperature, pressure, acceleration, weight displacement, and velocity (see Chapter 14). Electri-

*The latest information about IRIG standards is given in "IRIG Telemetry Standards," Document No. 106-66, March, 1966. Copies of this document are available from the Defense Documentation Center for Scientific and Technical Information, Cameron Station, Alexandria, Virginia 22314.

cal quantities, such as voltage, resistance, or frequency, also may be measured directly.

Signal conditioner. Generally includes the supporting circuitry for the transducer. This circuitry may provide excitation power, and balancing and calibration elements. An example of a signal conditioner is a strain-gage bridge balance and power supply unit.

Scanner or multiplexer. Accepts multiple analog inputs and sequentially connects them to one measuring instrument.

Signal converter. Translates the analog signal to a form acceptable by the analog-to-digital converter. An example of a signal conditioner is an amplifier for amplifying low-level voltages generated by thermocouples or strain gages.

Analog-to-digital (A/D) converter. Converts the analog voltage to its equivalent digital form. The output of the A/D converter may be displayed visually and is also available as voltage outputs in discrete steps for further processing or recording on a digital recorder.

Auxiliary equipment. This section contains instruments for system-programming functions and digital data processing. Typical auxiliary functions include linearizing and limit comparison. These functions may be performed by individual instruments or by a digital computer.

Digital recorder. Records digital information on punched cards, perforated paper tape, magnetic tape, typewritten pages, or a combination of these systems. The digital recorder may be preceded by a coupling unit which translates the digital information to the proper form for entry into the particular digital recorder selected.

Data-acquisition systems are used in a large and ever-increasing number of applications in a variety of industrial and scientific areas, including the biomedical, aerospace, and telemetry industries. The type of data-acquisition system, whether analog or digital, depends largely on the intended use of the recorded input data. In general, analog data systems are used when wide bandwidth is required or when lower accuracy can be tolerated. Digital systems are used when the physical process being monitored is slowly varying (narrow bandwidth) and when high accuracy and low per-channel cost is required. Digital systems range in complexity from single-channel dc voltage measuring and recording systems to sophisticated automatic multichannel systems that measure a large number of input parameters, compare against preset limits or conditions, and perform computations and decisions on the input signal. Digital data-acquisition systems are in general more complex than analog systems, both in terms of the instrumentation involved and the volume and complexity of input data they can handle.

Data-acquisition systems often use magnetic tape recorders, both analog and digital. (The principles of magnetic tape recording are discussed in Sec. 15-2.) Digital systems require converters to change analog voltages into discrete digital quantities or numbers. Conversely, digital information may have to be converted back into analog form, such as a voltage or a current, which may then be used as a feedback quantity controlling an industrial process. (Conversion techniques are discussed in Secs. 15-3, 15-4.) Scanning or multiplexing equipment is described in Sec. 15-5; Sec. 15-6 briefly introduces the spatial or shaft encorder.

15-2 Magnetic Tape Recorders

A basic magnetic tape recorder consists of the following elements:

(a) The *record head* which responds to an electrical signal and creates a magnetic pattern on a magnetizable medium.

(b) The *magnetic tape,* as the magnetizable medium, which conforms to and retains the magnetic pattern.

(c) The *reproduce head* which detects a magnetic pattern on the tape and converts it into an electrical signal.

(d) *Conditioning devices,* such as amplifiers and filters, to modify the input signal to a format that can be properly recorded on the tape.

(e) A *tape-transport mechanism* which moves the tape along the record and reproduce heads at a constant speed.

Three recording methods are specified to meet the various industrial requirements: (a) *direct recording,* (b) *frequency modulation (FM) recording,* (c) *pulse modulation (PM) recording.* Direct recording provides the greatest bandwidth and requires only relatively simple electronic components. FM recording overcomes some of the basic limitations of the direct recording process although sacrificing high-frequency bandwidth. Direct and FM recording are commonly used in industrial instrumentation systems, and they are discussed briefly in this chapter.

(a) *The Direct Recording Process.* A recording head is similar to a toroidal transformer with a single winding as shown in Fig. 15-2. The core is made in the form of a closed ring but it has a short *nonmagnetic gap* in it. Signal current in the winding causes magnetic flux which detours around the gap through the magnetic material on the tape. Magnetic tape is simply a ribbon of plastic with tiny particles of magnetic material deposited on it. When the tape is moved across the record head gap, the magnetic tape material is subjected to a flux pattern proportional to the signal current in the head winding. As it leaves the

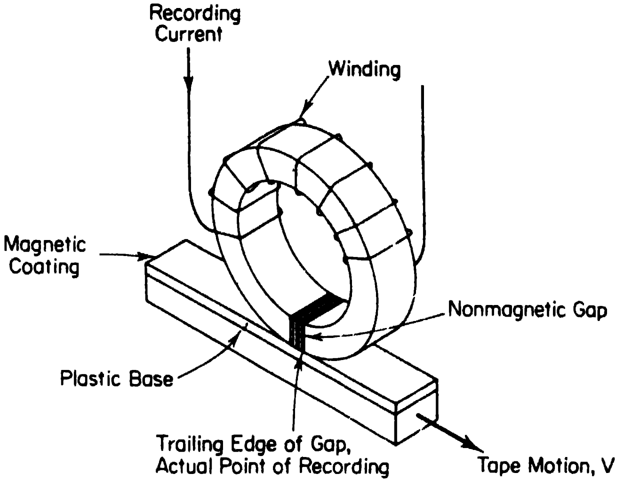


Figure 15-2
Simplified diagram of a magnetic recording head.

gap, each tiny magnetic particle retains the state of magnetization that was last imposed upon it by the protruding flux. The actual recording, therefore, takes place at the *trailing* edge of the record head gap.

The magnetic pattern on the tape is *reproduced* by moving the tape across a reproduce head. This head is similar in construction to the record head (in fact, it may be the same head). The magnetic oxide on the moving tape bridges the small nonmagnetic gap in the head and magnetic lines of flux are shunted through the core. The induced voltage in the head winding is proportional to the *rate of change* of flux, and the reproduced signal is actually the *derivative* of the recorded signal and not the signal itself. Consider the following discussion:

Assuming that the recording current in the head of Fig. 15-2 is sinusoidal, the instantaneous value of this current is given by

$$I_1 = I_m \sin(2\pi ft) \tag{15-1}$$

where I_m = peak value of the recording current
f = frequency of the sinusoidal waveform.

The magnetic flux in the recording head is proportional to the recording current, and we can write

$$\begin{aligned}\phi_1 &= k_1 I_1 \\ &= k_1 I_m \sin(2\pi ft) \\ &= \phi_m \sin(2\pi ft)\end{aligned} \tag{15-2}$$

where ϕ_1 = recording flux
k_1 = proportionality constant
f = frequency of the recording current.

In the reproduce mode, the flux through the reproduce head is proportional to the flux deposited on the magnetic tape and

$$\begin{aligned} \phi_2 &= k_2\phi_1 \\ &= k_2\phi_m \sin(2\pi f t) \end{aligned} \tag{15-3}$$

where ϕ_2 = reproduce flux

k_2 = proportionality constant.

The emf induced in the winding of the reproduce head equals

$$\begin{aligned} e &= -N\frac{d\phi_2}{dt} \\ &= -Nk_2\phi_m 2\pi f \cos(2\pi f t) \\ &= k_3 f \phi_m \cos(2\pi f t) \end{aligned} \tag{15-4}$$

Equation (15-4) indicates that the output from the reproduce head is not only proportional to the flux recorded on the tape, but also to the *frequency* of the recording current. Since the output voltage doubles for every octave rise in frequency of the recording current, it is said that the reproduce head output is subject to a 6 dB/octave rise. The 6 dB/octave rise must be compensated for in the reproduce amplifier by a process called *amplitude equalization,* which simply means that the amplifier has an amplitude response which drops at the rate of 6 dB/octave thereby compensating for the 6 dB/octave rise in the reproduce head. This is indicated in Fig. 15-3.

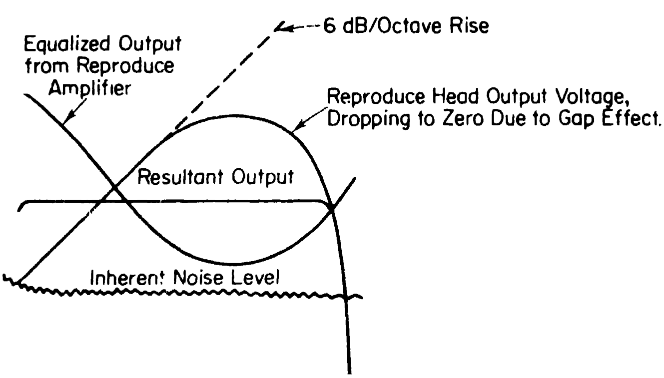


Figure 15-3
Reproduce characteristics.

Furthermore, Eq. (15-4) shows that, as the recorded frequency approaches zero, the output voltage falls below the inherent noise level of the over-all recording system. The direct recording process *cannot* be used for recording dc signals.

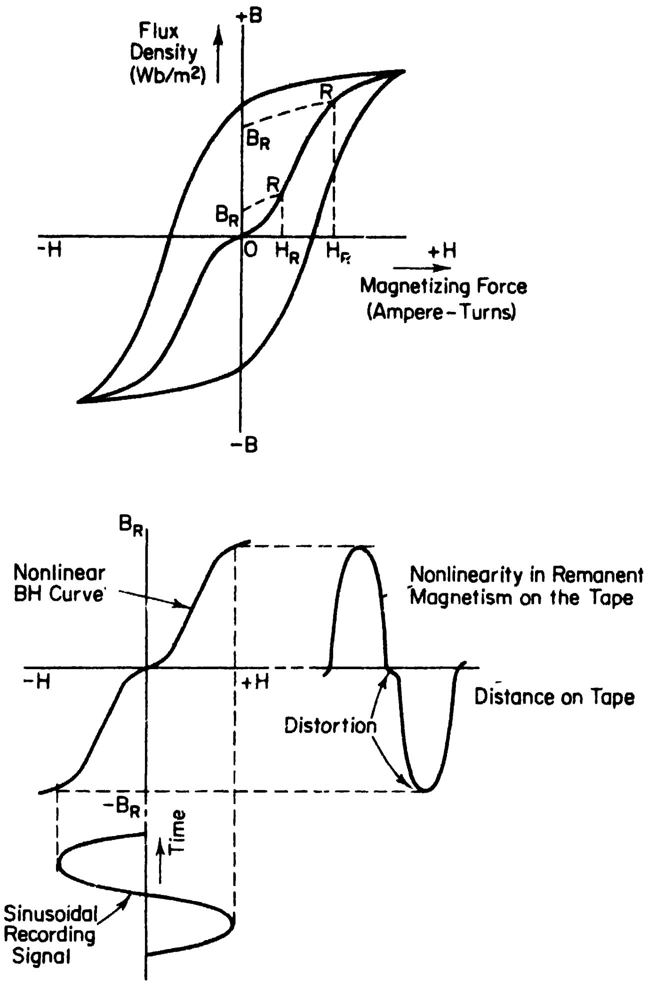


Figure 15-4

(a) A typical magnetization curve. (b) Head-to-tape transfer characteristics.

It has been assumed that the hysteresis or *BH* curve of the magnetic tape material is linear so that the magnetization (*B*) on the tape increases linearly with the magnetizing force (*H*) or number of ampere-turns. Figure 15-4(a) shows a typical *BH* curve with nonlinear characteristics. Tape magnetization proceeds as follows: As a demagnetized particle on the tape approaches the record head gap, it carries no residual magnetism and is therefore at point *O* on the *BH* curve of Fig. 15-4. Assuming that one period of the recorded signal is very long compared to the gap length, the particle will pass under the gap area through an essentially constant magnetizing force H_R. This force carries the particle up the *BH* curve to point *R*. As the particle leaves the area of the gap, the magnetizing force H_R drops to zero and the particle follows a minor hysteresis loop RB_R, retaining a residual or remanent magnetism B_R. Since the *BH* curve is nonlinear, the remanent magnetism of all the particles on the tape follows essentially the same nonlinear pattern, and the entire magnetization process is therefore inherently nonlinear. High distortion in the reproduced signal results unless corrective action is taken—see Fig. 15-4(b).

Nonlinearity may be greatly reduced by applying an *ac bias current* in *series* with the recording signal current so that only the linear sections of the magnetization curve are used. Figure 15-5 shows how the recording function is biased

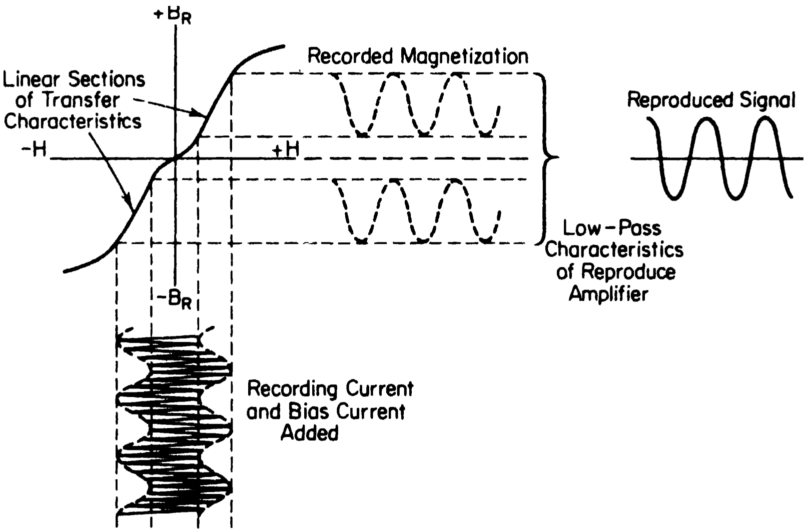


Figure 15-5
Graphical representation of high-frequency bias in direct recording.

into the linear region of the *BH* transfer characteristics simply by *adding* or *mixing* a high-frequency (approximately four times the highest signal frequency) bias signal to the recording signal. The correct amplitude of the bias current depends on the exact *BH* transfer characteristics of the magnetic tape and should be adjusted to reach from center to center of the linear regions, as indicated in Fig. 15-5. Too large bias currents saturate the tape and result in loss of high-frequency response; too low bias currents cause distortion at the low-frequency end. In practice, bias currents are from 1 mA to 20 mA, generally five to thirty times the recording signal current, depending on the tape and head characteristics.

If a sine-wave signal is recorded on the tape, the magnetic intensity of the recorded track varies sinusoidally. The distance on the tape required to record one complete cycle is called the ***recorded wavelength*** λ, and

$$\lambda = \frac{v}{f} \tag{15-5}$$

where λ = recorded wavelength
v = tape speed
f = frequency of the recorded signal.

A tape recorder with a frequency response specified as 1.2 MHz at a tape speed of 120 ips (inch per second), is capable of a ***packing density*** of 10,000 cycles per inch. The recorded wavelength of the magnetic pattern on the tape is then 0.0001 in. or $\frac{1}{10}$ mil, which is the limit of the recorder's ***resolution***. Both ***packing density*** and ***resolution*** can be used to describe the response of a recorder, independent of tape speed, and they are more definitive of the recorder's capability than a simple frequency specification at a given speed.

The ***high-frequency response*** of the direct recording system is limited by several factors, the most important of which is the ***gap length*** of the reproduce head. As shown in Fig. 15-6, the output voltage of the reproduce head increases with frequency, simply because $d\phi/dt$ increases. This is true up to the point where the dimension of the head gap equals one-half the recorded wavelength. Beyond this point the output voltage decreases rapidly and drops to zero when the gap dimension equals the recorded wavelength. At this point there is no magnetic gradient spanned by the gap and therefore no output voltage. This phenomena is called the ***gap effect*** and is the most serious restriction on the high-frequency response of a tape recorder.

The ***dynamic range*** or signal-to-noise (S/N) ratio is usually quoted in decibels and is defined as the ratio of the maximum signal to the minimum signal which can be recorded at a certain amount of total harmonic distortion (THD). In general, instrumentation recorders have a S/N ratio of 22 dB to 30 dB at 1 per cent THD and a frequency response down to 400 Hz. Response standards for the direct recording process are established by the IRIG standards (see Sec. 15-1).

(b) *Frequency Modulation Recording.** In FM recording, the carrier or center frequency is deviated in response to the amplitude of a modulating or data signal. An ac data signal alternately increases and decreases the carrier frequency above and below its center value at a rate equal to the frequency of the data signal. In the reproduce process, the amplitude instability of the carrier is virtually eliminated by limiting, and the data signal is reconstructed by detect-

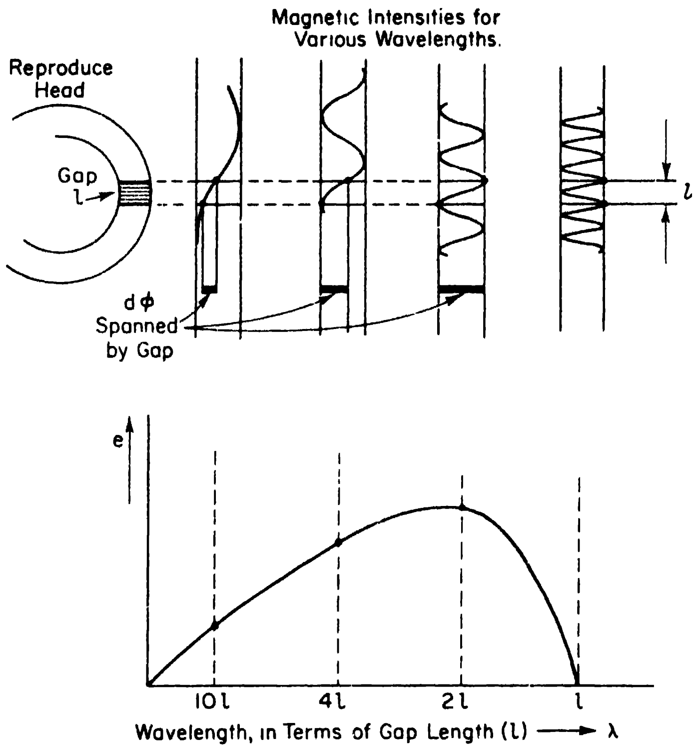


Figure 15-6
Illustrating gap effect.

ing zero crossings in the FM demodulator. Residual carrier signals and any out-of-band noise are removed by a low-pass filter. FM recording is used primarily when the dc component (an undeviated carrier) of the input signal is to be preserved or when the amplitude variations of the direct recording method cannot be tolerated. FM recording is extremely sensitive to tape-speed fluctuations (*flutter*), since, in either the record or the reproduce mode, tape-speed variations introduce unwanted apparent modulation of the carrier (noise), reducing the dynamic range of the system.

Two important factors in FM recording are *deviation ratio* and *percentage deviation*. *Deviation ratio* is defined as the ratio of carrier deviation from center frequency to the signal or modulating frequency, or

$$\delta = \frac{\Delta f}{f_m} \tag{15-6}$$

where δ = deviation ratio
Δf = carrier deviation from center frequency
f_m = data signal frequency.

A system with a high deviation ratio generally has a low noise figure. However, Δf is limited by the recorder bandwidth and f_m must be kept high to accommodate all the data signals.

Percentage deviation is defined as the ratio of carrier deviation to center frequency, or

$$m = \frac{\Delta f}{f_0} \tag{15-7}$$

where m = percentage deviation or modulation index
Δf = carrier deviation from center frequency
f_0 = center frequency.

With a low percentage deviation, such as 7.5 per cent in FM telemetry subcarrier systems, the effect of flutter becomes noticeable. For example, a 1 per cent deviation caused by flutter in a system with 7.5 per cent deviation corresponds to $1/7.5 \times 100$ per cent = 13.3 per cent noise. The same flutter imposed on a 40 per cent deviation system, causes only $1/40 \times 100$ per cent = 2.5 per cent noise signal. Systems with a higher percentage deviation are generally less influenced by tape-speed variations.

A well-designed FM carrier recording system will give reasonably good amplitude accuracy, dc response and dc linearity, and low distortion. Frequency response for a given tape speed is greatly reduced, and the record-reproduce electronics is much more complex than for a direct recording system. To keep flutter figures low, the tape transport mechanism must be of good design supplying constant tape speed. FM instrumentation recorders generally adhere to IRIG standards where tape speeds, center frequencies, deviations, etc., are well defined.

(c) *Digital Magnetic Tape Recording.* Digital magnetic tape units are generally used as storage devices in digital data-processing applications. Digital tape units are generally of two types: *synchronous* and *incremental.* The *incremental* digital recorder is commanded to step ahead (increment) for each digital character to be recorded. The input data may then be at a relatively slow or even a discontinuous rate. In this way, each character is equally and precisely spaced along the tape.

In the *synchronous* digital recorder, the tape moves at a constant speed (say 30 ips) 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 a *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-bits in the appropriate tracks across the tape width. The recording technique used in most instrumentation tape recorders is the industry-accepted IBM format of NRZ recording. In this system, the tape is magnetically *saturated at all times*, either in the positive or in the negative direction. The NRZ (non-return-to-zero) method uses the *change* in flux direction on the tape to indicate a 1-bit and *no change* in flux direction as a 0-bit. The method is illustrated in Fig. 15-7, where the binary number 1100101 is represented by a flux pattern in the NRZ system.

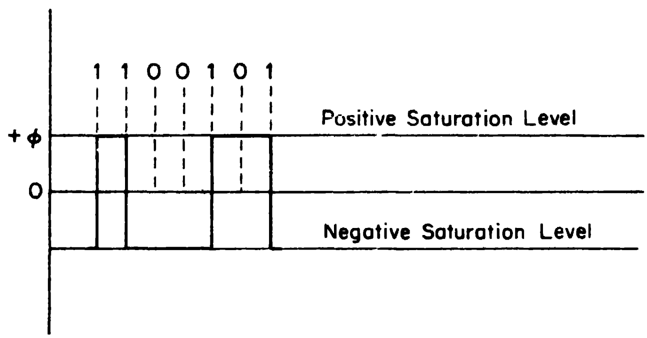


Figure 15-7
NRZ (non-return-to-zero) recording method for a digital tape recorder. Changes in flux direction indicate a 1-bit and no changes indicate a 0-bit.

Since tape magnetization is independent of frequency and amplitude but relies only on the *polarity* of the recording current, the usual problems of nonlinearity and distortion found in direct and FM recording are nonexistent. The write coils of the tape heads require only sufficient current of the correct polarity to saturate the tape. Some of the problems encountered in digital tape recording are signal *dropout* 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 so-called parity check. This check involves keeping track of the number of 1-bits initially recorded on the tape by writing a parity check pulse on an extra tape track if the number of 1s recorded is even (even parity) or if the number of 1s recorded is odd (odd parity). 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.

Some of the advantages of digital tape recording are high accuracy and relative insensitivity to tape speed, simple electronic conditioning devices, and direct compatibility with computer systems allowing direct transfer of data between tape unit and computer.

15-3 Digital-to-Analog Conversion

The problem of converting a digital number to an analog voltage can be solved in several ways. The more sophisticated converters use diodes, transistors, and other elements, but the principle of conversion is best illustrated by the simple passive resistor network of Fig. 15-8.

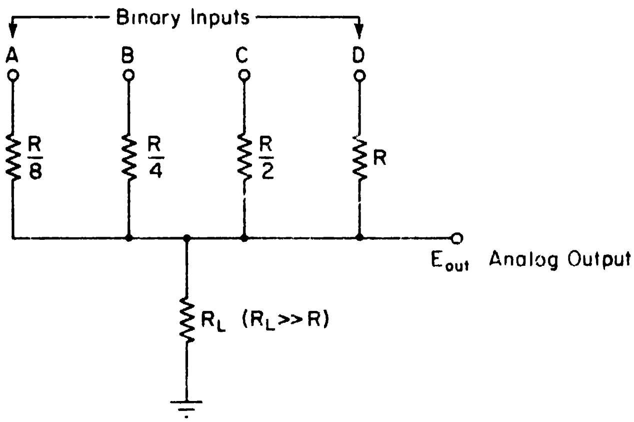


Figure 15-8
Basic digital-to-analog converter using a resistive divider network. The binary inputs are at 0 V or + E V, assuming positive logic.

Assume a logic system where a binary 0 is represented by a voltage level of 0 V and a binary 1 by a voltage level of +E V. The binary number, represented by its corresponding combination of voltage levels, is applied to the input terminals of the resistive divider, with the least significant bit (LSB) connected to the terminal marked D. The four input resistors are weighted so that bit 1 (the LSB) has an input resistor of value R, bit 2 has an input resistance of value $R/2$, bit 3 has a resistor of value $R/4$, and so on. The value of R_L, the load resistor, is very large compared to the input resistors. The output voltage E_{out} will be a dc voltage between the values 0 V and +E V, depending on the value of the binary number represented by the four inputs.

The binary number 0001, applied to the input of the converter of Fig. 15-8, applies 0 V to the A, B, and C inputs and +E to the D input. The input resistors act as a *voltage divider*, connected between 0 and +E V and consisting of R_D

in series with the parallel combination of R_A, R_B, and R_C. The output voltage E_{out}, therefore, equals $\frac{1}{16}$E V. If the input equals the binary number 0010, the voltage E_{out} will be $\frac{2}{16}$E V and if the input is 0011, the output voltage will be $\frac{3}{16}$E V. In other words, an increase of one bit at the input to the converter causes an increase in output voltage of $\frac{1}{16}$E V. When the input reaches its maximum number of 1111, the full output voltage of + E V is obtained. The digital input signal is therefore converted, in discrete steps of $\frac{1}{16}$E V, to an *analog output* voltage.

The accuracy of conversion depends on the accuracy of the resistors and the voltage levels of the binary inputs. The resistors are usually carefully selected precision resistors, and the voltage levels of the binary inputs are controlled by a reference supply in order to increase conversion accuracy. The input circuit should also be able to deliver the necessary current without affecting the dc input level.

In a practical circuit, the resistive network—sometimes called a DAC, (*digital-to-analog conversion* module)—is connected to a flip-flop register which *holds* the digital number. (See Fig. 15-9.) Since the divider is simply a passive network, the digital input voltage (the *on* and *off* levels) determines the output voltage. Since digital voltage levels are usually not as precise as required in an analog system, level *amplifiers* may be placed between the flip-flop register and the divider network. These amplifiers switch the inputs to the divider network between ground and a reference voltage supplied by a precision reference supply. The analog output voltage then falls between these two levels.

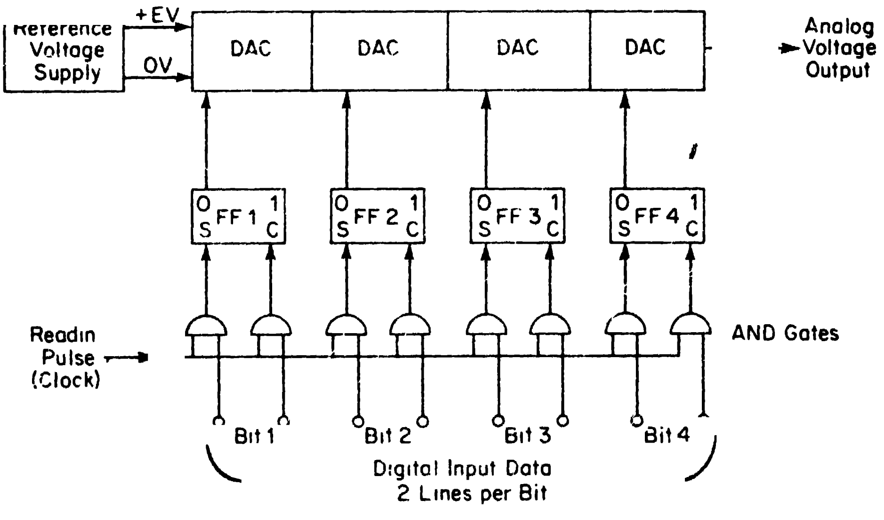


Figure 15-9
A digital-to-analog converter.

A *practical* D-to-A converter is shown in Fig. 15-9. The basic components are recognized as the flip-flop *register, DAC* modules which include the level amplifiers, and a *reference* supply. The digital signals are dropped into the register by a *drop-in* pulse (usually a clock pulse) and are automatically converted by the DAC divider network to the appropriate analog voltage.

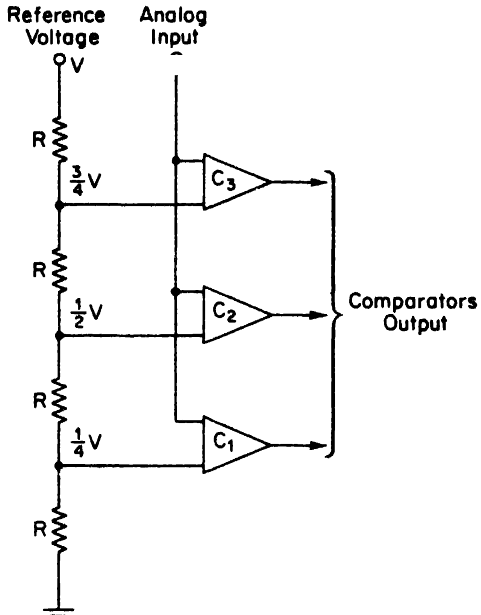
It always takes some time for the conversion to be completed after the digital signals are dropped into the register. The *settling* time depends on the number of flip-flops in the register that change state and also on the voltage difference between the original output voltage and the new output voltage. For instance, when the digital input changes from the binary number 0111 to the new binary number 1000, all the flip-flops change their state. The output voltage, however, changes only by $\frac{1}{16}E$ V. *Transients* may occur at the analog output owing to variations between transition times of the different flip-flops, and transient current may be drawn from the reference supply. These transients are usually of very short duration (typically in the order of 2 μs) and can be neglected since the load cannot respond within this time.

15-4 Analog-to-Digital Conversion

Analog-to-digital conversion is slightly more complex than digital-to-analog conversion and a number of different methods may be used. The four most common methods are described in this section. Of these, the successive-approximation counter is the most widely used A/D converter because it provides excellent performance for a wide range of applications at a reasonable cost.

The *comparator circuit* forms the basis of all A/D converters. This circuit compares an *unknown* voltage with a *reference* voltage and indicates which of the two voltages is larger. A comparator is essentially a multistage, high-gain differential amplifier, where the *state* of the output is determined by the *relative polarity* of the two input signals. If, for instance, input signal A is greater than input signal B, the output voltage is a maximum and the comparator is *on*. If input signal A is smaller than input signal B, the output voltage is zero or minimum and the comparator is *off*. Since the amplifier has a very high gain, it either saturates or cuts off at relatively low differential input levels, so that it acts as a *binary* device.

(a) *The Simultaneous Method.* A simple yet effective digital-to-analog converter can be built using several comparator circuits. This is shown in the circuit of Fig. 15-10, where three comparator circuits are used. Each of the three comparators has a *reference input* voltage, derived from a precision reference voltage source. A resistive divider consisting of four equal precision resistors



C_1	C_2	C_3	Input Voltage
off	off	off	0 to $\frac{1}{4}$V
on	off	off	$\frac{1}{4}$V to $\frac{1}{2}$V
on	on	off	$\frac{1}{2}$V to $\frac{3}{4}$V
on	on	on	$> \frac{3}{4}$V

Figure 15-10
Simultaneous analog-to-digital converter.

is connected across the reference supply and provides output voltages of $\frac{3}{4}$ V, $\frac{1}{2}$ V, and $\frac{1}{4}$ V volts, where V is the reference output voltage. The other input terminal of each comparator is driven by the unknown analog voltage.

In this example, the comparator is *on* (providing an output), if the analog voltage is larger than the reference voltage. If none of the comparators are on, the analog input must be smaller than $\frac{1}{4}$ V. If comparator C_1 is on and both C_2 and C_3 are off, the analog voltage must be between $\frac{1}{4}$ V and $\frac{1}{2}$ V. Similarly, if C_1 and C_2 are both on and C_3 is off, the analog voltage must be between $\frac{1}{2}$ V and $\frac{3}{4}$ V; if all the comparators are on, the analog voltage must be greater than $\frac{3}{4}$ V. In total, four different output conditions may exist: from *no* comparators on to *all* comparators on. The analog input voltage can therefore be *resolved* in four equal steps. These four output conditions can be coded to give two binary bits of information. This is shown in the table alongside the diagram of Fig. 15-10. Seven comparators would give three binary bits of information, fifteen comparators would give four bits, etc.

The advantage of the simultaneous system of A/D conversion is its simplicity and speed of operation, especially when low resolution is required. For a high-resolution system (a large number of bits), this method requires so many comparators that the system becomes bulky and very costly.

(b) *The Counter Method.* If the reference voltage, to which the analog input is to be compared, were *variable,* the number of comparators could be reduced to only *one.* If, for instance, the reference voltage is a linearly increasing voltage (a ramp) which is continuously applied to the comparator input, coincidence of the reference voltage and the unknown voltage could be determined in terms of the *time elapsed* since the ramp started. But, a *digitally controlled variable reference* already exists in the form of the simple digital-to-analog converter of Fig. 15-9. This D/A converter can be used to convert a digital number in its DAC register into an analog voltage which can be compared to the unknown analog input by a comparator circuit. If the two voltages are not equal, the digital number in the DAC register is modified and its output compared again. This is exactly the operation of the circuit of Fig. 15-11.

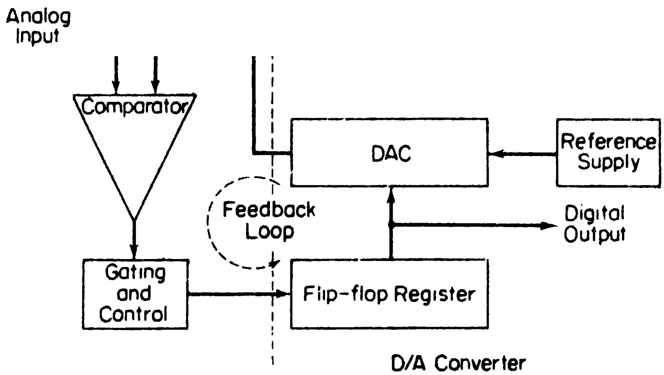


Figure 15-11

Analog-to-digital converter using a digital-to-analog converter to provide the comparison voltage. The contents of the flip-flop register provide the digital output.

The generalized analog-to-digital converter of Fig. 15-11 is actually a closed-loop *feedback* system, where the main components are the DAC, the comparator, and some control logic circuitry.

Various methods may be used to *control* the conversion taking place in the D/A converter. One of the simplest ways is to start the DAC at zero and *count* the number of input pulses required to give an output voltage which equals the analog input.

The *counter-type analog-to-digital converter* of Fig. 15-12 contains a D/A converter section consisting of the resistive-divider network (DAC), the reference supply, and a six-stage counter which replaces the DAC register of Fig. 15-11. The comparator again receives the unknown analog input for comparison against the generated DAC output voltage. The control circuitry consists

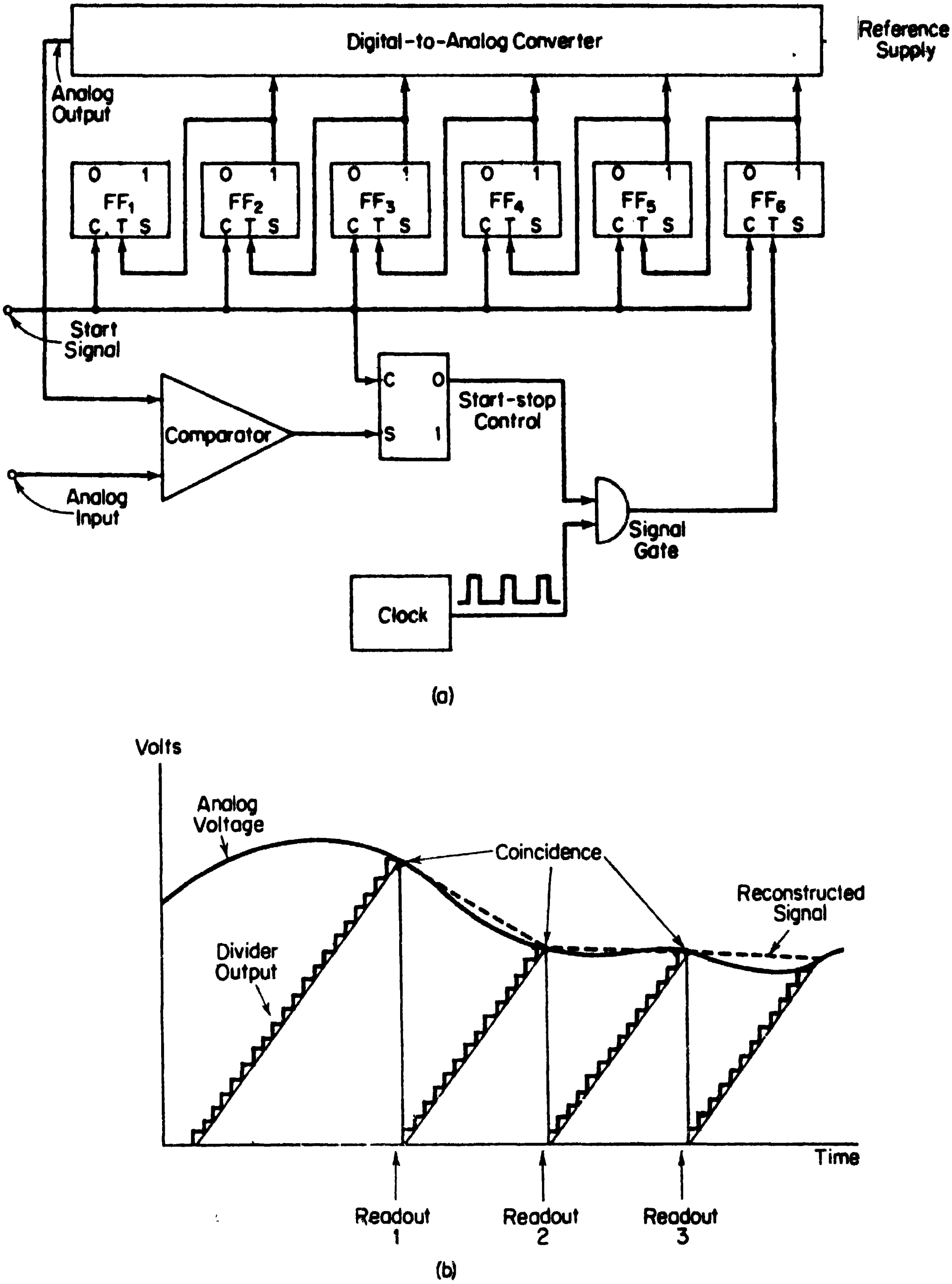


Figure 15-12
(a) Simplified logic diagram of a counter converter. (b) Waveform diagram of analog input and divider output showing points of coincidence when readout occurs.

of a pulse generator or *clock*, a signal *gate* which steers the clock pulses to the counter, and a *control* flip-flop for starting and stopping the conversion.

When a start signal is given, all the counter flip-flops are cleared and the start-stop flip-flop is reset. This flip-flop provides a gating level (positive logic) to the signal gate, allowing the clock pulses to be applied to the counter register. The clock pulses are propagated through the counter, and the DAC divider output increases in steps toward the top of the reference voltage. When the divider output is equal to the analog input, the comparator switches, delivering an output signal to the start-stop flip-flop. This flip-flop sets and its output drops to zero, blocking the clock pulses at the signal gate. At this instant, the counter stores the number of clock pulses that were required to raise the reference voltage to the level of the analog input voltage. The contents of the counter is then the binary equivalent of the analog input.

The *conversion time* is measured from the moment a request is given to the moment a digital output is available. For the countertype A/D converter the conversion time depends on the magnitude of the analog voltage and is therefore not constant. If the input signal is variable, it is also important to know when the input signal had the value given by the digital output. The uncertainty of this time measure is called the *aperture* time (sometimes also called *window* or *sample time*). The aperture occurs at the end of the conversion, as shown in the waveform diagram of Fig. 15-12(b).

The resolution of the counter-type converter is improved by adding extra bits (extra counter stages). This addition can be done at small extra cost. The conversion time, however, increases rapidly with the number of bits used, since an N-bit converter needs time for 2^N counts to accumulate. Higher resolution is therefore obtained at the cost of increased conversion time.

One method of decreasing the conversion time is to divide the counter into sections. For instance, a ten-bit converter could be divided into two sections of five bits each. At the start of the conversion, the least significant section of the counter is preset to all 1s and counts are inserted only into the most significant section. When the comparator indicates that the analog input level has been exceeded, the least significant section of the counter is cleared, reducing the DAC divider output, and pulses are then inserted into the least significant section until the correct value is reached. The maximum number of steps required to complete a conversion is 2^5 for the most significant counter and 2^5 for the least significant counter, giving a total of 2^6 steps. This is a maximum of 64 counts (2^6) versus 1,024 counts (2^{10}) for the standard counter.

The section counter technique is frequently used in digital voltmeters, where the output is to be in decimal notation. Each section of the counter then represents a sectional digit.

(c) *The Continuous Converter.* The big disadvantage of the counter converter is that the entire comparison process starts from the beginning after a coincidence has been detected by the comparator. This means low resolution and low speed.

A slight modification of the counter method involves replacing the simple counter with a *reversible counter* or *up-down counter.* Then the converter can continuously follow the analog input voltage whatever the direction in which this voltage changes. Once the converter starts running, the digital equivalent of the input voltage can be sampled at any time, and an extremely rapid readout is possible.

The simplified logic block diagram of Fig. 15-13 represents the continuous converter. The illustration contains four basic parts: (1) the up-down counter, (2) the D/A converter, (3) the comparator, (4) the synchronization and control logic.

An ordinary binary counter counts in the forward direction (*up*) when the trigger input of the succeeding binary is connected to the 1 output of the preceding binary (see Sec. 13-2). The count will proceed in the reverse direction (*down*) if the coupling is made instead to the 0 output of the preceding binary. The two methods may be used simultaneously to produce an up-down counter. In Fig. 15-13 additional AND gates are used in the trigger circuits of the binaries to make sure that counts are accumulated only at the desired moment; *synchronization of the count and the comparison* must take place.

The D/A converter section is identical to the basic resistive divider of Fig. 15-8; the reference supply provides the required precision voltage for accurate conversion. The 0 outputs of the binaries are connected directly to the DAC, but it should be understood that appropriate level conversion takes place between the binaries and the DAC input terminals.

The comparator again compares the analog input voltage with the DAC output voltage, providing two possible output voltage levels. When the analog input voltage is larger than the feedback voltage (DAC output), the appropriate comparator output terminal is connected to the set input of the *up* flip-flop via a gating element. Similarly, when the analog input voltage is smaller than the feedback voltage, the comparator provides an output voltage at its other terminal which is then connected, via a gate, to the set terminal of the *down* flip-flop. The actual transfer of the comparator output signals to the up and down flip-flops is controlled by gating pulses from the clock, which controls the synchronization of the entire measurement cycle. The outputs of the up and down flip-flops are *exclusive OR'd* together to make sure that no count takes place when both flip-flops are set (a safety precaution).

At the start of a measuring cycle, when all the flip-flops are cleared, the clock generates a pulse which samples the comparator output. If the analog

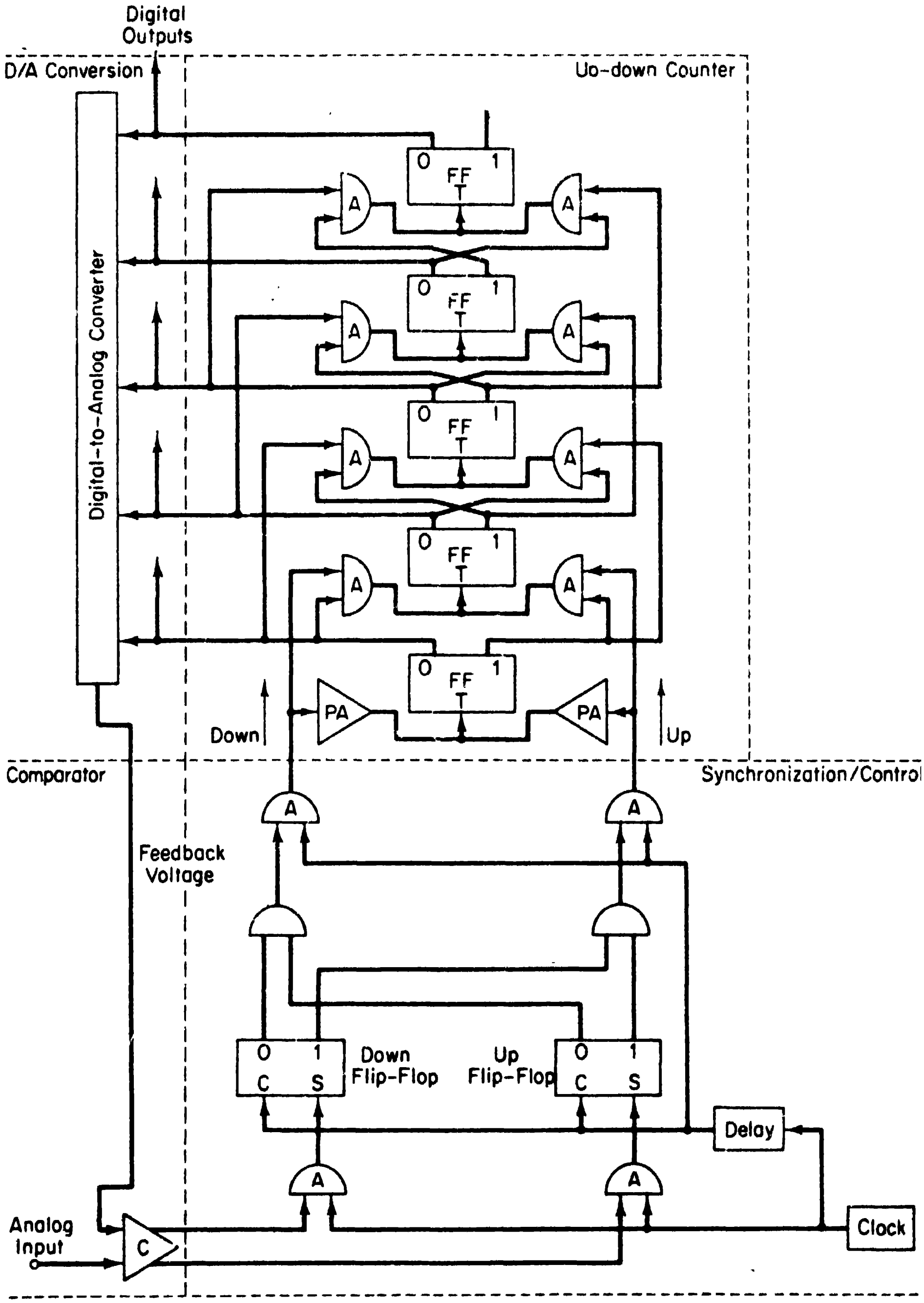


Figure 15-13
Simplified logic block diagram of the continuous converter.

input is larger than the feedback voltage (as it usually is when a measurement first starts), the *up* flip-flop is set. The delayed clock pulse is then allowed to trigger the first binary while, at the same time, it also conditions the trigger gate of the succeeding binary. The 0 output of the first binary is connected to the DAC, and the resulting DAC output is compared to the analog input by the comparator. If the analog input is still larger, the next clock pulse sets the *up* flip-flop again, allowing the delayed clock pulse to trigger the first binary back to its original state and also to trigger the next binary. The count then has advanced by one and the corresponding DAC output is used for comparison against the analog input. The procedure repeats until the feedback voltage equals the analog input, at which time the comparator output is zero and the count is stopped.

If the analog input changes to a lower value, the next clock pulse detects this at the comparator output and sets the *down* flip-flop. Now the delayed clock pulse is allowed to enter the binary counter at the trigger input of the first

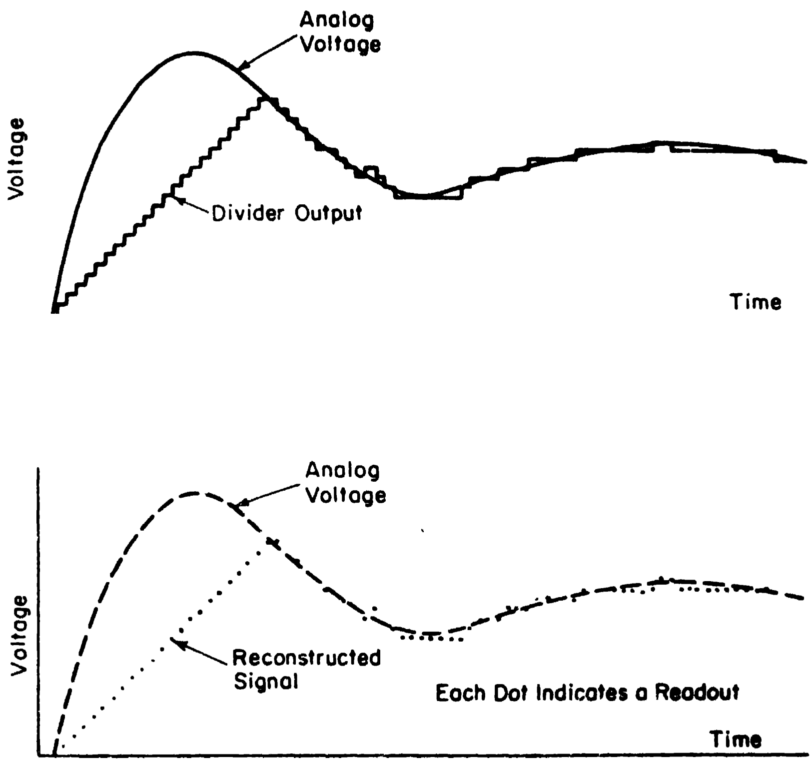


Figure 15-14
Waveform diagrams illustrating the action of the continuous A/D converter.

binary, but the count is carried from stage to stage at the 0-output side of the binaries of Fig. 15-13, so that the contents of the counter is reduced by one. The DAC output then also drops the appropriate amount and the next comparison determines whether the *up* flip-flop or the *down*-flip-flop will be set. The counter therefore *continuously follows* the analog input voltage.

The waveform diagrams of Fig. 15-14 illustrate the action of the continuous converter. The *aperture* is the time for the last step. The assumption is made that the analog input voltage does not change more than ± 1 LSB (the smallest increment of the DAC) between conversion steps. To meet this requirement, the maximum rate of change of the input voltage must not exceed the maximum rate of change of the converter.

(d) *The Successive-Approximation Converter.* The successive-approximation analog-to-digital converter compares the analog input to a DAC reference voltage which is repeatedly divided in half. The process is illustrated by Fig. 15-15, where a four-digit binary number (1000), representing the full reference supply voltage V, is divided in half (binary number 100), corresponding to $\frac{1}{2} V$. A comparison between this reference voltage ($\frac{1}{2} V$) and the analog input is made. If the result of this comparison shows that this first approximation was too small ($\frac{1}{2} V$ is smaller than the analog input), then the next comparison

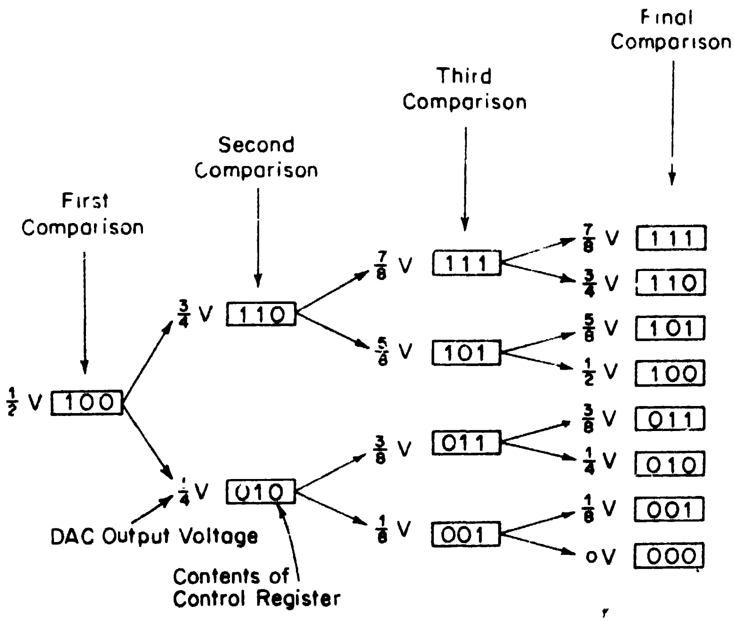


Figure 15-15
Operation of the successive approximation A/D converter.

will be made against $\frac{3}{4}$ V (binary number 110). If the comparison showed that the first approximation was too large ($\frac{1}{2}$ V larger than the analog input), then the next comparison will be made against $\frac{1}{4}$ V (binary number 010). After four successive approximations, the digital number is resolved. A six-digit number will be resolved in six successive approximations. This compares very favorably with the sixty-four (2^6) comparisons needed with a conventional counter-type converter.

The successive approximation method is a little more elaborate than the previous methods since it requires a special *control register* to gate pulses to the first bit, then to the second bit, and so on. The additional cost of the control register however is small and the converter can handle continuous and discontinuous signals with large and small resolutions at moderate speed and moderate cost.

The generalized block diagram of Fig. 15-16 shows the basic successive

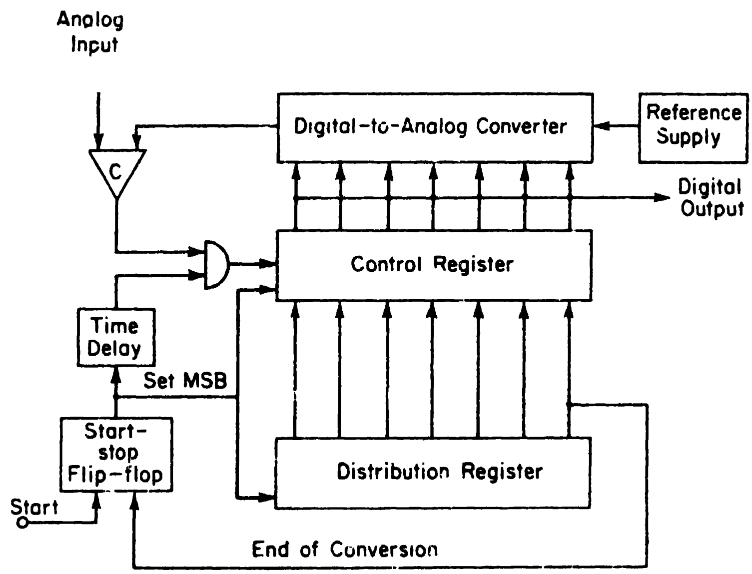


Figure 15-16
Simplified block diagram of the successive approximation A/D converter.

approximation converter. The converter uses a digital control register with gateable 1 and 0 inputs, a digital-to-analog converter with reference supply, a comparison circuit, a control timing loop, and a distribution register. The distribution register is like a ring counter with a single 1 circulating in it to determine which step is taking place.

At the beginning of the conversion cycle, both the control register and the distribution register are set with a 1 in the *most significant bit* (MSB) and a 0 in all bits of less significance. The distribution register therefore registers that the cycle has started and that the process is in its first phase. The control register, which now reads 1000 . . . , causes an output voltage at the digital-to-analog converter section of *one-half* of the reference supply. At the same instant, a pulse enters the timing delay chain. By the time that the D/A converter and the comparator have settled, this delayed pulse is gated with the comparator output. When the *next most significant bit* is set in the control register by the action of the timing chain, the *most significant* bit either remains in the 1 state or it is reset to the 0 state, depending on the comparator output. The single 1 in the distribution register is shifted to the next position and keeps track of the number of comparisons made.

This procedure repeats, following the diagram of Fig. 15-15, until the final approximation has been corrected and the distribution register indicates the end of the conversion. Synchronization is not required in this system because the comparator controls only one flip-flop at a time.

For a successive-approximation converter, the digital output corresponds to some value that the analog input had during the conversion. Thus, the aperture time is equal to the total conversion time. This is illustrated in the waveform reconstruction of Fig. 15-17. Aperture time of this converter can be reduced by using redundancy techniques or by using a *sample-and-hold circuit* (Sec. 15-14(e)).

(e) *Sample-and-Hold.* A *sample-and-hold* circuit is used with an analog-to-digital converter when it is necessary to convert a high-frequency signal which is varying too rapidly to allow an accurate conversion. The sample-and-hold circuit is basically an operational amplifier which charges a capacitor during the *sample mode* and retains the value of the charge of the capacitor during the *hold mode*. The sample-and-hold circuit can be represented by the simple switch and capacitor of Fig. 15-18.

When the switch is first closed, the capacitor charges to the value of the input voltage and then follows the input (assuming a low driving source impedance). When the switch is opened, the capacitor holds the voltage which it had at the time that the switch was opened (assuming a high-impedance load).

The *acquisition time* of the sample-and-hold is the time required for the capacitor to charge up to the value of the input signal after the switch is first shorted. The *aperture time* is the time required for the switch to change state and the uncertainty in the time that this change of state occurs. The *holding time* is the length of time the circuit can hold the charge without dropping more than a specified percentage of its initial value.

It is possible to build a sample-and-hold circuit exactly as shown in Fig.

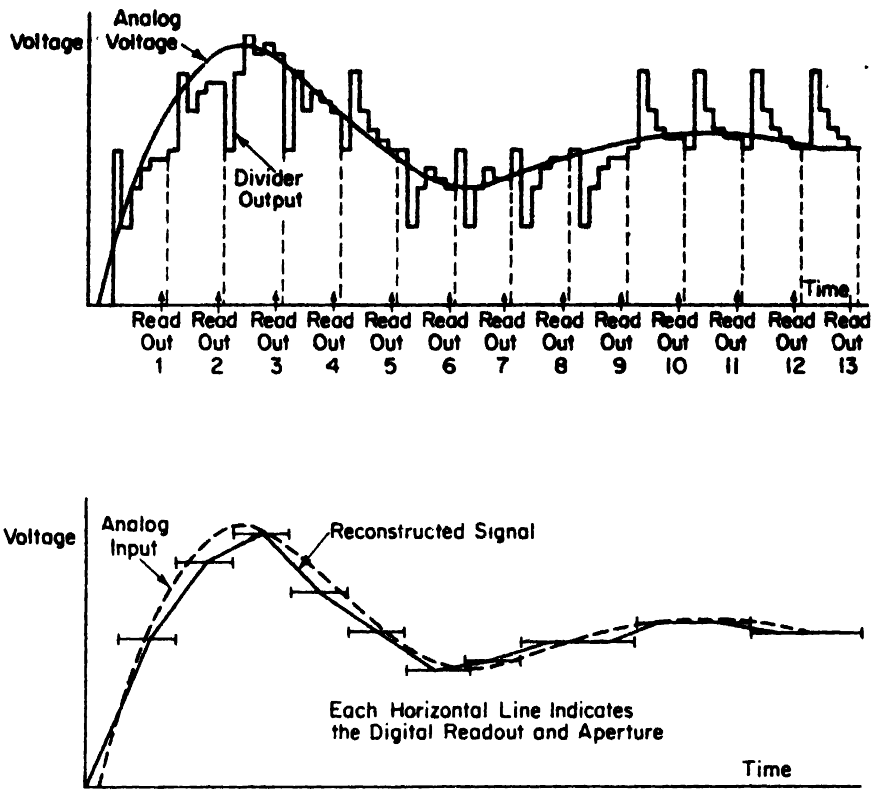


Figure 15-17
Waveform diagrams illustrating the operation of the successive approximation A/D converter.

15-18. Often, however, the circuit is built with fast-acting transistor switches and an operational amplifier to increase the available driving current into the capacitor or to isolate the capacitor from an external load on the output. However the sample-and-hold circuit is built, it always acts as the simple switch and capacitor shown.

An actual sample-and-hold circuit is shown in the schematic diagram of Fig. 15-19. The *sample* pulse operates switches 1 and 3, the *hold* pulse operates switches 2 and 4. The sample-and-hold control pulses are complementary. In the sample mode, the hold capacitor is charged up by the operational amplifier. In the hold mode, the capacitor is switched into the feedback loop, while the input resistor R_i and the feedback resistor R_f are switched to ground. Since the input to the amplifier remains within a few μV to ground (except during switching), the input impedance is 10 kΩ in both the sample and the hold modes.

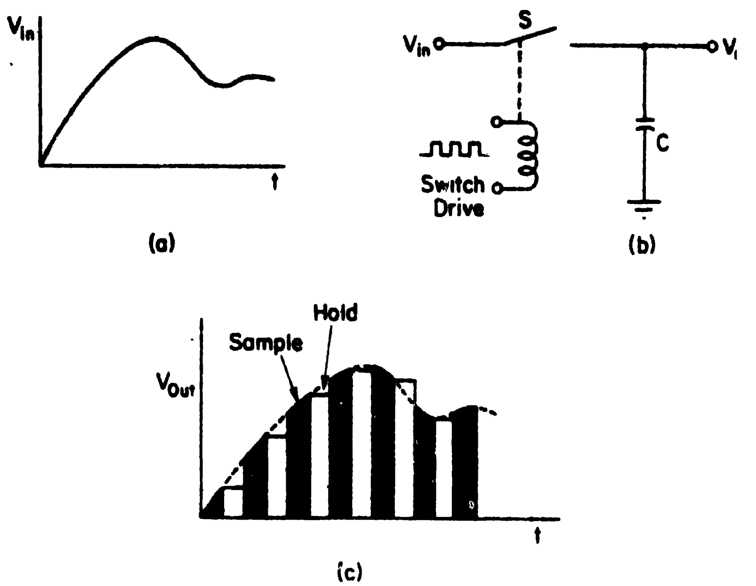


Figure 15-18
Basic sample-and-hold circuit. (a) Input waveform. (b) Circuit. (c) Output waveform, showing the varying waveform during the sample interval and the fixed amplitude during the hold interval.

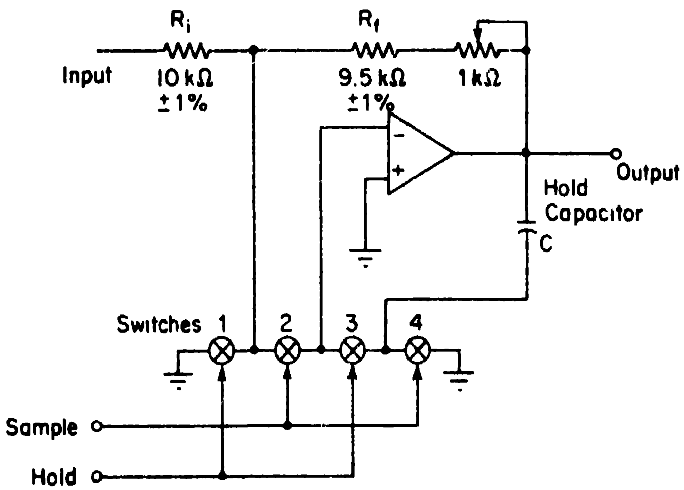


Figure 15-19
Sample-and-hold system.

15-5 Multiplexing

It is often necessary or desirable to combine or *multiplex* a number of analog signals into a single digital channel or, conversely, a single digital channel into a number of analog channels. Both digital signals and analog voltages can be multiplexed.

(a) *Digital-to-Analog Multiplexing.* In digital-to-analog conversion, a very common application of multiplexing is found in computer technology, where digital information, arriving sequentially from the computer, is distributed to a number of analog devices, such as a CRO, a pen recorder, an analog tape recorder, and so on. There are two ways to accomplish multiplexing: The first method uses a *separate* digital-to-analog converter for each channel, as shown in Fig. 15-20(a). The second method uses one *single* digital-to-analog converter, together with a set of analog multiplexing switches and sample-and-hold circuits on each analog channel, as shown in Fig. 15-20(b).

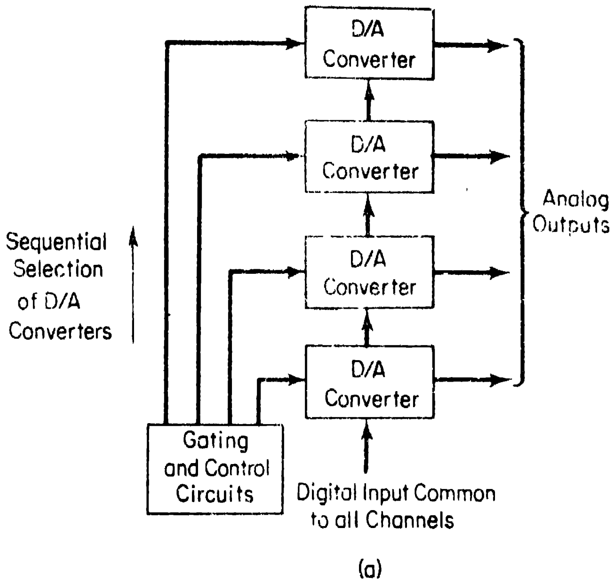
In the system of Fig. 15-20(a), the digital information is applied simultaneously to all channels, and channel selection is made by gating clock pulses to the appropriate output channels. One D/A converter is required per channel, so that the initial cost may be somewhat higher than the second system, but the advantage is that the analog information is available at the DAC output for an indefinite period of time (as long as the contents of the DAC flip-flop register are gated to the DAC).

The second method, illustrated in Fig. 15-20(b), uses only one digital-to-analog converter and is therefore slightly lower in initial cost. The multiple sample-and-hold technique, however, requires that the signal on the sample-and-hold circuits be renewed at periodic intervals (the capacitors do not hold their charge indefinitely).

(b) *Analog-to Digital Multiplexing.* In analog-to-digital conversion, it is convenient to multiplex the analog inputs rather than the digital outputs. A possible system is given in Fig. 15-21, where switches, either solid-state or relays, and used to connect the analog inputs to a *common bus*. This bus then goes into a single analog-to-digital converter which is used for all channels.

The analog inputs are switched *sequentially* to the bus by the channel selector control circuitry. If simultaneous time samples from all channels are required, a sample-and-hold circuit may be used ahead of each multiplexer switch. In this manner, all channels would be sampled simultaneously and then switched to the converter sequentially.

It is also possible to multiplex by using a separate comparator for each



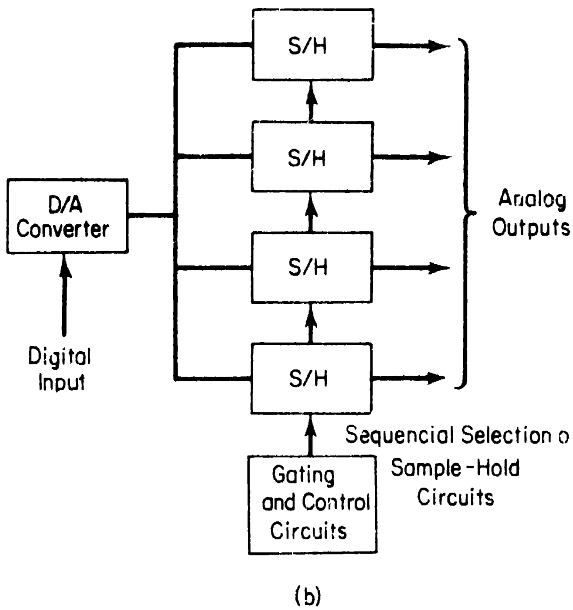


Figure 15-20

(a) Digital-to-analog multiplexing, using several converters. (b) Digital-to-analog multiplexing, using one converter and several sample-and-hold circuits.

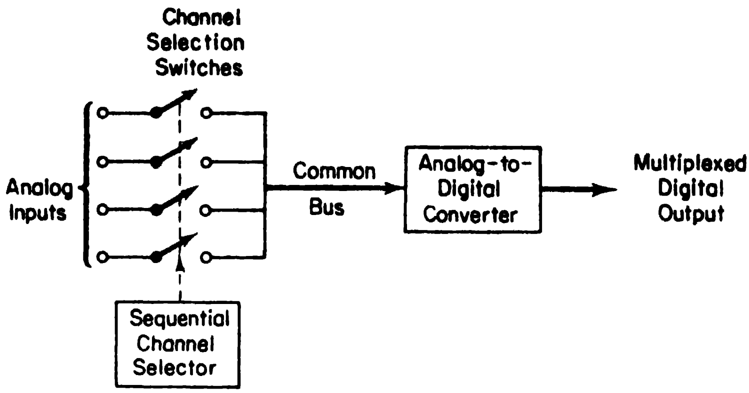


Figure 15-21
A multiplexed A/D conversion system.

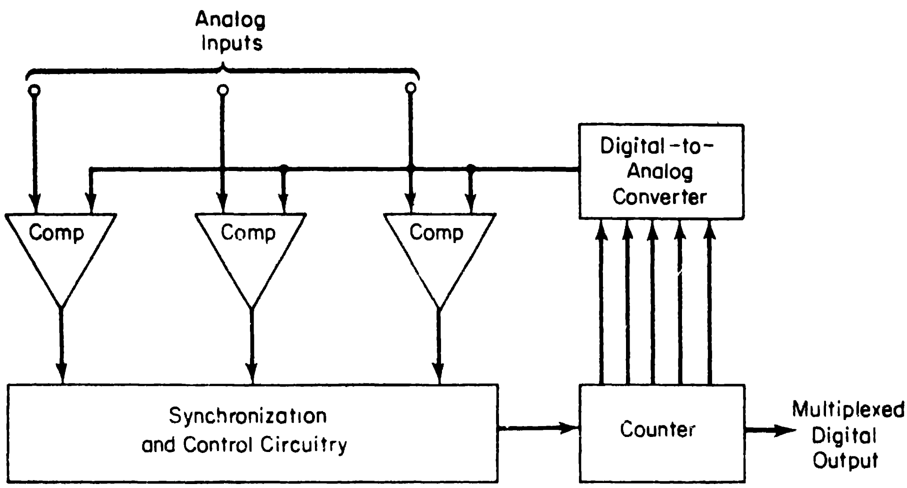


Figure 15-22
A counter type A/D converter with multiplexed input.

analog channel. This system is shown in Fig. 15-22, where it is used with a counter-type analog-to-digital converter. The input of each comparator is connected to the output of the DAC. The other input to each comparator is connected to the separate analog input channels. Synchronization and control circuitry is required to operate the counter and sample the comparators. At the start of the multiplexing process, the counter is cleared and count pulses are applied to the counter. The digital-to-analog converter translates the counter output and provides an analog output voltage, which is fed to all the comparators. When one of the comparators indicates that the digital-to-analog output is greater than the input voltage on that channel, the contents of the counter are read out. Counting is then resumed until the next signal is received, when the correct comparator is identified and the counter contents read out again.

15-6 Spatial Encoders

A *spatial encoder* is a mechanical converter which translates the angular position of a shaft into a digital number. It is therefore an analog-to-digital converter, where the analog quantity is *nonelectrical.* An important application of this type of encoder is in the self-balancing potentiometer (Sec. 6-7), where the encoder is mounted on the shaft of the self-balancing motor and therefore *reads* the balancing emf applied to the measuring circuit.

The encoder consists of a cylindrical disk with the coding patterns arranged in concentric rings on one side of the disk. This is shown in Fig. 15-23.

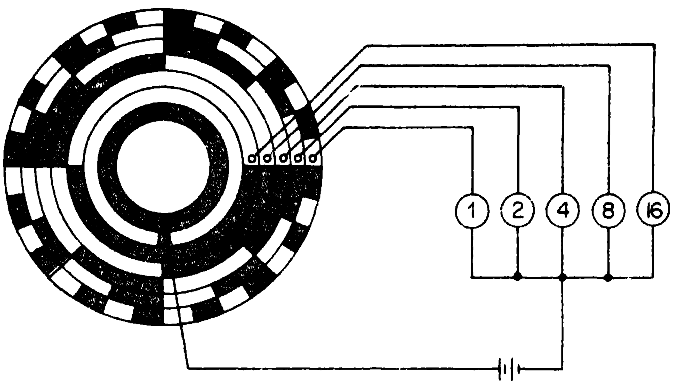


Figure 15-23
A spatial encoder using a binary counting system.

The patterns are alternating segments of conducting (black) and nonconducting (white) material. The nonconductive areas are formed by depositing a thin layer of insulating material on the conductive disk. The disk is contacted by a set of brushes, one for each ring, arranged radially from the center outward. The number of segments on the concentric rings decrease, in a binary count (32-16-8-4-2), from a total of thirty-two (sixteen conductive and sixteen nonconductive) on the outside ring to two on the inside ring. With a battery and a lamp matrix connected as shown in Fig. 15-23, each angular position of the driving shaft would have a different combination of brush contacts bearing on the conductive surface areas. The lamp matrix would light in a binary-coded pattern.

Since the outer ring or *commutator* of this disk encoder has 32 distinct areas, the resolution would be $\frac{1}{32} \times 360° = 11\frac{1}{4}°$ (1 digit). The resolution can be improved by increasing the number of commutators, thereby decreasing the angle.

Spatial encoders can be made to produce any desired digital number system. The encoder shown in Fig. 15-24 is a binary-coded decimal converter

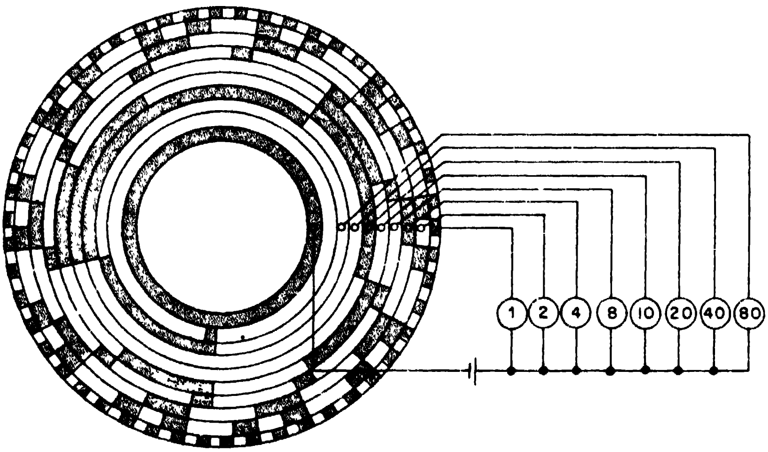


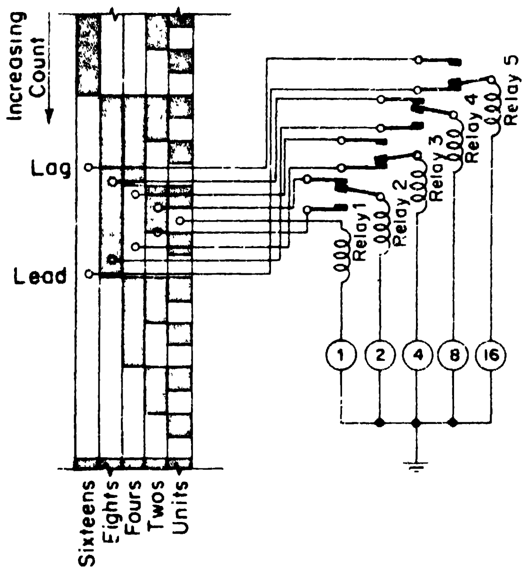
Figure 15-24
A binary-coded-decimal disk converter capable of a readout from 0 to 99. Since the outer commutator is divided into 100 segments, the angular position of the disk can be resolved to 3.6°.

where the eight coded commutators produce a readout from 0 to 99. In the position shown, the commutator segments make contact with the brushes connected to readout circuits 20, 4, and 1, giving a total indication of 25 on the readout panel. If the disk were to be rotated clockwise for one division of the outer ring, the commutators would make contact with readout circuits 20, 4, and 2, for a totalized display of 26.

Suppose now that the disk is rotated for only half a division from its original position in Fig. 15-24. The possibility exists that the brush on the outer ring would still make contact with the commutator segment before the brush on the second ring leaves its commutator area. The readout circuit would then indicate a reading of 20, 4, 2, and 1, for a total reading of 27, whereas the correct reading should be either 25 or 26. The possibility of ambiguous readings with this type of encoder has led to the development of several other systems to overcome the problem. One of these systems uses an extra outer commutator, with the conducting segments placed in regions where ambiguities may occur. When the brush on this outer ring makes contact, it automatically adds the proper increment of driving voltage to move the disk to a region of nonambiguity.

One common method in the binary number system uses two sets of brushes, arranged in a V-shaped pattern. This type of encoder is called the *V-brush binary encoder*. Its operation is based on the natural progression of the digits in the binary number system, as given in Table 12-1. Note that, when the digit in the "unit" column changes from a 0 to a 1, none of the digits in the other columns change. Furthermore, when the digit in the unit column changes from a 1 to a 0 the digit in the twos column changes. Similarly, when the digit in the twos column changes from a 0 to a 1, the digits in the more significant columns are unchanged and when the digit in the twos column goes from a 0 to a 1, the digit in the fours column changes. This relationship is true for a change of digits in any column.

This logic is shown in diagrammatic form in Fig. 15-25, which represents



(a)

Decimal	Eights	Fours	Twos	Ones
0				0
1				1
2			1	0
3			1	1
4		1	0	0
5		1	0	1
6		1	1	0
7		1	1	1
8	1	0	0	0
9	1	0	0	1
10	1	0	1	0
11	1	0	1	1
12	1	1	0	0
13	1	1	0	1
14	1	1	1	0
15	1	1	1	1

(b)

Figure 15-25

(a) The V-brush binary encoder. (b) The binary number system.

a section of a binary-coded disk converter. Each of the five commutators represents a digit in a five-digit binary number. The right-hand commutator represents the least significant digit, corresponding to the units column of Table 12-1 which is reproduced alongside the disk section in Fig. 15-25 for convenience. The next commutator to the left represents the next least significant digit, corresponding to the twos column, and so on. The shaded portions of the commutator are connected to the battery circuit, similar to the connections of Fig. 15-24. In Fig. 15-25, one brush is placed on the least-significant-digit, or units, commutator. On the next commutator, two brushes are placed: one brush is *leading* the brush on the units commutator by half a division (half a count); the other brush is *lagging* the brush on the units commutator by half a count. The next two brushes, placed on the fours commutator, are displaced by half a count with respect to the two brushes on the previous commutator. All the other commutators have a leading and a lagging brush, displaced by the same half-count distance.

When the single brush on the units commutator reads a 0, as is indicated in Fig. 15-25, the lagging brush of the twos commutator reads correctly, while the leading brush may give an ambiguous reading. When the disk advances one division, the brush on the units container reads a 1, but now the leading brush on the next commutator reads correctly whereas the lagging brush may give the ambiguous reading. Similarly, when the output of the twos commutator is a 0, the lagging brush on the next commutator reads correctly; and when the output of the twos commutator is a 1, the leading brush on the next commutator reads correctly.

It follows, then that the single brush on the first ring could be used to control some logic circuitry which selects the leading or the lagging brush on the other commutating segments. Such a logic system can be provided by electromechanical relays, such as shown in Fig. 15-25, or by transistor circuitry.

If the brush on the units ring reads a 0, relay 1 is not energized, and the lagging brush on the twos ring is connected to relay No. 2 and indicator No. 2. When the brush on the units ring reads a 1, relay No. 1 is energized, removing the lagging brush and connecting the leading brush of the twos ring into the No 2 display circuit. Similarly, when the brush on the twos commutator reads a 1, it would energize the relay in circuit No. 2, connecting circuit No. 3 to the leading brush on the fours commutator, and so on. In other words, automatic selection of the proper brushes prevents ambiguities in the readings.

INDEX

A

N

O

P

Q

R

S

T