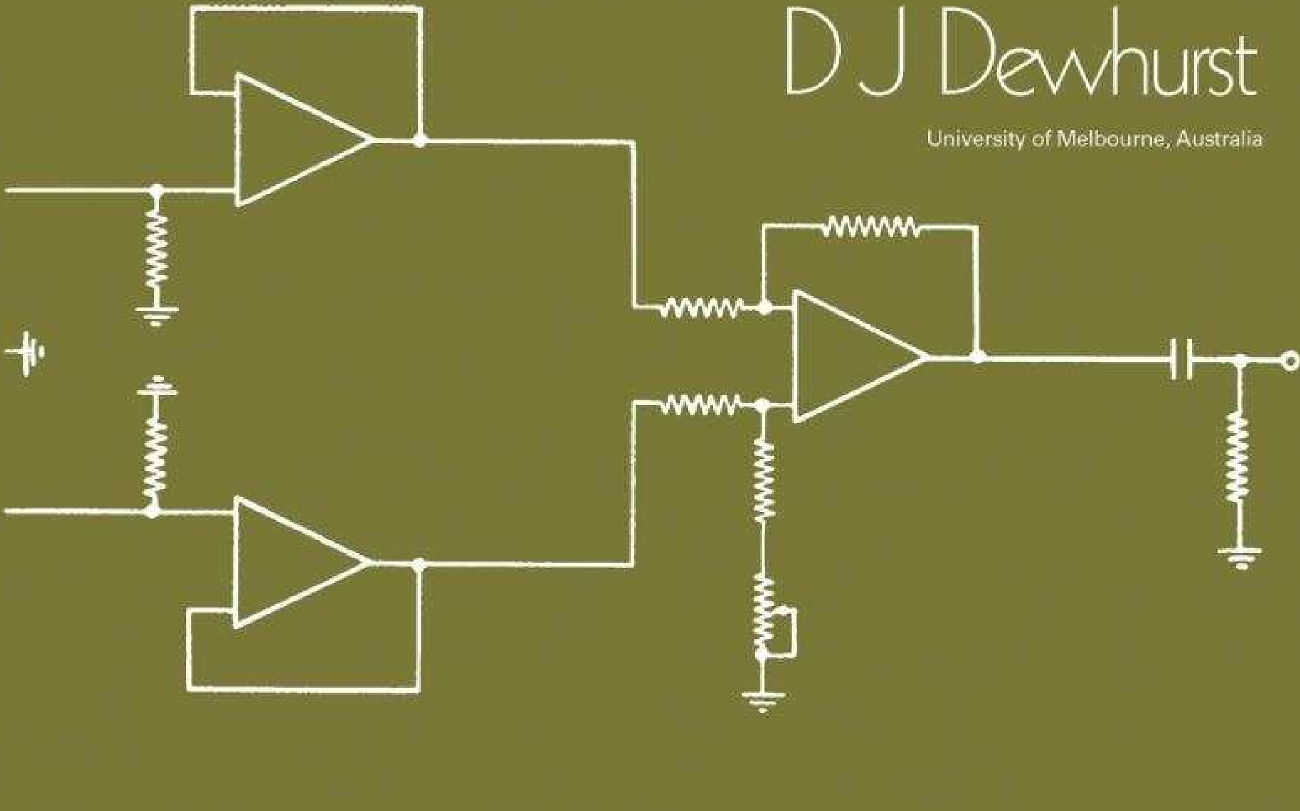
AN INTRODUCTION TO BIOMEDICAL INSTRUMENTATION



D J Dewhurst

University of Melbourne, Australia

Pergamon International Library of Science, Technology, Engineering and Social Studies. The 1000-volume original paperback library in aid of education, industrial training and the enjoyment of leisure.

An Introduction to Biomedical Instrumentation

BY

D. J. DEWHURST

Reader in Biomedical Engineering
University of Melbourne
Australia

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO
SYDNEY · PARIS · BRAUNSCHWEIG

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada Ltd., 207 Queen's Quay West, Toronto 1, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
WEST GERMANY	Pergamon Press GmbH, 3300 Braunschweig, Postfach 2923, Burgplatz 1, West Germany

First edition 1966

Second edition 1976

The first edition of this book was originally published as
Physical Instrumentation in Medicine and Biology

Library of Congress Cataloging in Publication Data

Dewhurst, D J
An introduction to biomedical instrumentation

First ed. published in 1966 under title: Physical instrumentation in medicine and biology
Includes index
1. Medical electronics. 2. Physiological apparatus. I. Title. [DNLM: 1. Biomedical engineering—Instrumentation. 2. Electronics, Medical—Instrumentation. QT34 D521p]
R856.D49 1975 610'.28 75-17622
ISBN 0-08-018755-2
ISBN 0-08-018884-2 pbk.

Printed in Great Britain by A. Wheaton & Co. Exeter

FOREWORD

THIS book provides a course of study and practical assignments covering the basic principles of medical and biological instrumentation, and typical features of its design and construction. It is based on experience extending over twelve years in conducting such a course in the University of Melbourne, and is aimed primarily at graduates in medicine or the biological sciences who require modern instrumentation in their work. It has been used successfully in conversion courses for electronics technicians entering the field of medical electronics, and as the basis of a course of private study by many individual students.

It must be emphasised that the aim of the course is to provide the student with a fund of knowledge sufficient to allow him to use equipment confidently and competently, and to communicate with electronics engineers and technicians. Unless supplemented by a great deal of further training and experience, the course will not convert him to a technician, or still less into an engineer. Suggestions for further reading are provided at the end of each chapter.

The theory and practical work in the order set out form a logical development of the subject, falling into three parts. The first part consists of a review of electrical principles and components, including transistors, with particular emphasis on the background required in the biomedical field. The second is a review of the basic circuits used in the construction of modern equipment, and the third covers a selection of measuring techniques and concepts. The reader may ask "Why learn to build instruments, such as oscilloscopes, which are much better purchased complete?" My answer is that only by personally assembling and testing currently used circuits can the learner acquire the framework of ideas needed to discuss or employ complete instruments with neither scepticism nor blind faith. I know of no substitute for "hands-on" experience.

Provided that some assistance is given to students by prefabricating the metal work of each unit to be assembled, the allocation of an hour to the theoretical material and 3 hours to the practical work in each chapter has proved satisfactory. Experience has shown that a course of this nature cannot be shortened appreciably without losing contact with the average graduate student, who has little or no background in electronics. None of the commercial "teach-yourself-electronics" kits at present available is at all satisfactory for the purpose of this course. The practice in Melbourne is for students to work in pairs, each member of the course having his or her own multimeter, soldering iron, and hand tools.

The emphasis in this book is on current techniques, and a considerable amount of the material included is not readily available in standard references. Solid state circuitry is used throughout; it is considered that although vacuum tubes are still found in some equipment used for research or in clinical practice, they will soon disappear entirely.

I would like to thank my wife for her whole-hearted assistance in the preparation of this book, including the drawings of nearly all the line diagrams; Professor R. D. Wright, former head of the Department of Physiology, for his continued help and encouragement in the development of biomedical engineering in Australia; and Mr. Lindsay Dally and Mr. Stuart Mackereth, for many years my technical officers and friends.

CHAPTER 1

INSTRUMENT CONSTRUCTION

1.1 Human engineering in instrument design

In both the clinical and biological research fields nearly all measuring instruments in current use are of relatively intricate design, and incorporate electronics to a considerable extent. They are, however, intended for use by persons almost completely ignorant of the techniques involved, though quite familiar with the parameters they wish to measure. The situation places a great responsibility on the purchaser of such an instrument to consider every aspect of the relation between the instrument and its potential user—the *human engineering* or *ergonomic* factors involved.

A good instrument is *functional*. It has a minimum of controls accessible to the operator, and these are clearly labelled, foolproof, easy to operate, and logically laid out. It is *reliable*, requires little or no routine maintenance or recalibration, operates when needed with a minimum of delay, and gives the required information reproducibly under all conditions of use. It is *compact*, so that it does not impede the user in carrying out his normal duties. The case is *solidly constructed*, without sharp edges, and finished with a surface that is easily cleaned, and does not take finger marks. The instrument must function *independently of environmental conditions* likely to be encountered, such as heat, humidity, dust or sometimes an explosive atmosphere. It is *safe* to use, both for the user and for any patient on whom it may be used; a number of quite subtle hazards can exist, particularly when more than one piece of equipment is used on a patient. (This is fully discussed in Chapter 23, and comprehensive safety standards exist in most countries.) Finally, it is *reasonably priced, can be tried* by the potential user *before a purchase order is placed*, can be *delivered within a reasonable time*, is *supplied with an adequate instruction manual*, including a complete circuit, and its sale is *supported by adequate service facilities*, including immediate availability of replacement parts.

1.2 Physical construction

The physical construction of a modern instrument is dictated by two major factors; these are panel layout and compactness of wiring. Given a required number of controls and an arrangement of them on the front panel which is logical *to the user*, the panel dimensions are practically determined. There has been a tendency to reduce the size of switches and knobs as electronic components have progressively diminished in size; unfortunately human hands have remained the same size throughout this process.

The electronic circuitry carrying out the desired functions of the instrument can be made very compact, even where highly complex operations are involved. This circuitry is always assembled on printed circuit boards, consisting of small sheets of epoxy fibre which carry the components and the printed wiring. No other type of construction is suited to modern components; their pin spacing is too small to permit of individual attachment of connecting wires, and their frequency response is so high that individual connecting wires inevitably lead to instability in operation.

Printed wiring may be carried out by several processes. The epoxy fibre board as supplied by the manufacturer is covered on one or both sides with a thin sheet of copper, which is firmly bonded to it. To prepare a circuit, the desired layout of the wiring is drawn up, and is then transferred on to the copper in the form of lines of a material (the *resist*), which is impervious to the etching solution to be used. The board is then immersed in the etching solution (ferric chloride) which removes the copper except where the resist has been placed. The resist is washed off with a solvent, and the desired circuit remains. Components are mounted by drilling fine holes through the board as required, passing the leads of the components through, and soldering them to the copper strips on the board.

A number of such boards can be installed side by side; interconnections are made either between sockets into which individual boards plug, or directly between terminal pins on each board. The former method is less neat, but allows for easy replacement of whole boards should it ever become necessary (Fig. 1.1).

Interconnection can be carried out using plastic-covered stranded wire, or bare single-strand tinned wire; the latter may be insulated by plastic or cambric sleeving slipped over it.

1.3 Heat sinks

Most transistors and integrated circuit units produce very little heat; occasionally however, it is necessary to handle quite large currents. In this case larger transistors are used, and the heat they produce must be carried away and dissipated into the surrounding air. Thick aluminium extrusions, treated by an anodising process to give a matt black finish, are available for use as heat sinks, and the transistors concerned are bolted into their centres, using a film of thermally conducting silicone paste to exclude air between transistor and sink. It is often necessary to insulate the transistor electrically from the sink, while providing intimate thermal contact. This is achieved by the use of a thin mica washer, and insulating washers under the heads of the retaining bolts. Figure 1.2 shows a power transistor mounted on a heat sink. Manufacturers of heat sink materials provide tables to allow the size of sink required for various purposes to be estimated.

1.4 Interconnection of instruments

It is frequently necessary to connect one instrument to another, providing one or many wires between them. Where slowly varying signals only are concerned, multi-cored cable may be employed, terminated at each end in a suitable plug. For reasons of safety, power *sources* are always supplied from a shrouded socket or plug (a *female* socket or plug). There is a vast range of types of multi-pin connectors in common use; most of those designed for domestic television or gramophone reproduction are far too fragile and unreliable for medical use Suitable types have some type of lock to prevent accidental

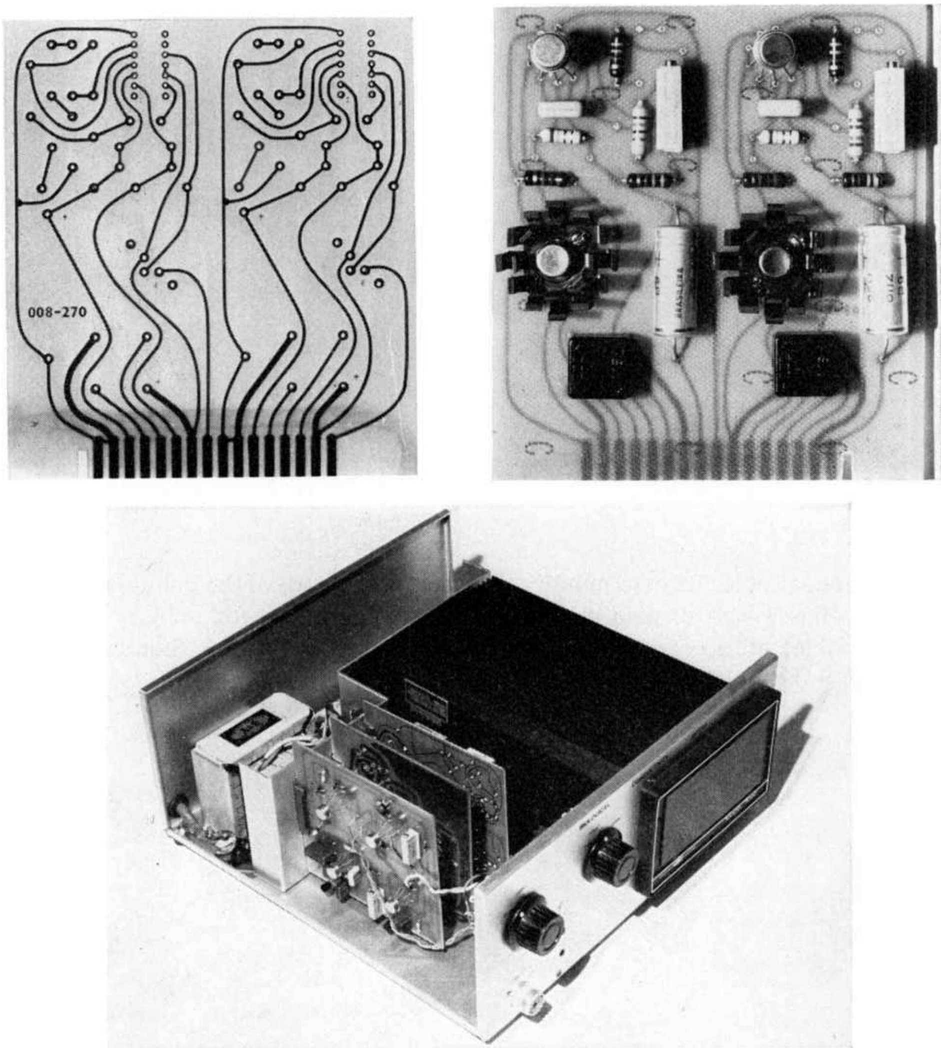


FIG. 1.1. (a) Etched circuit board. (b) Fully assembled board. (c) Boards in complete instrument.

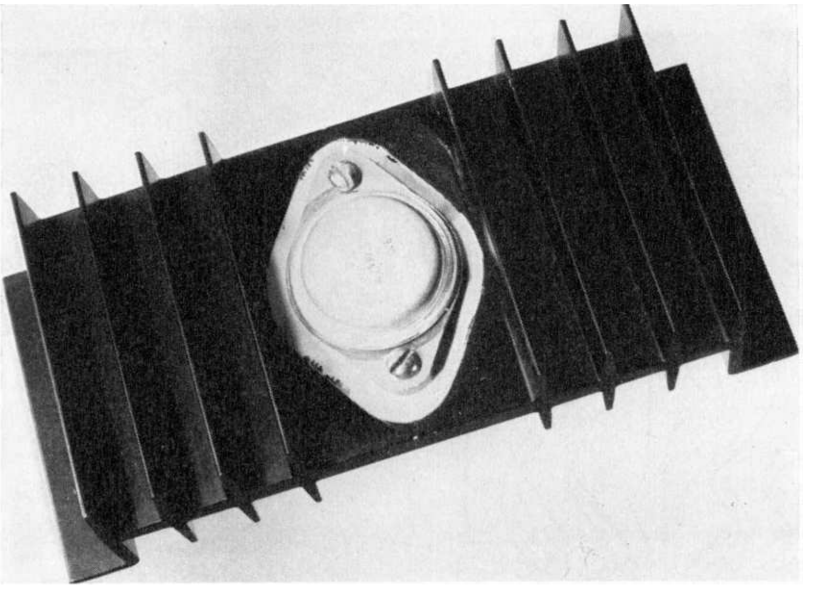


FIG. 1.2. Power transistor on heat sink.

disconnection, a cable clamp to minimise fracture of the leads at the point of entry to the plug, and robust silver- or gold-plated pins. Connectors made for military or aviation use are suitable; some commonly used types are shown in Fig. 1.3. Standardisation of plugs, at least throughout an institution, is highly desirable.

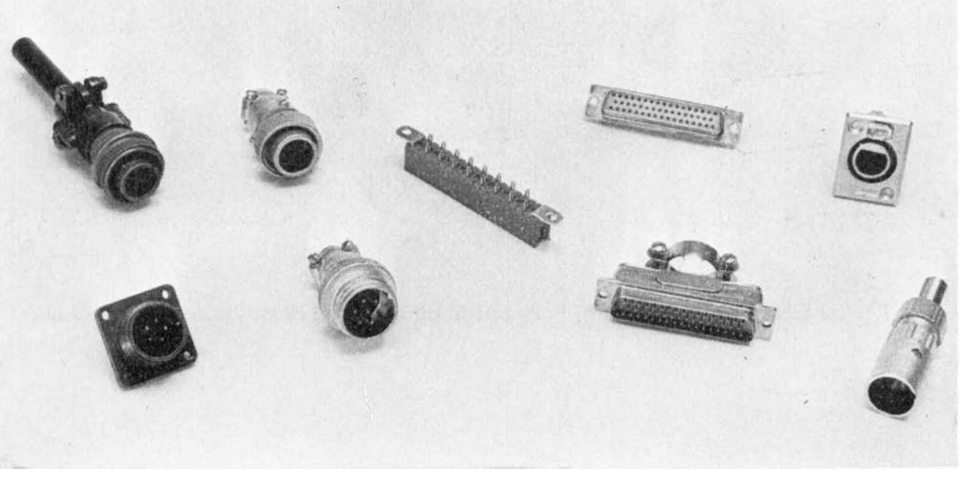


FIG. 1.3. Multi-pin connectors.

Where signals must be carried for some distance, or are varying rapidly, each is carried in a separate *coaxial* cable, consisting of a central conductor, an insulating sheath, a copper braid shield, and an outer protecting sheath. These are terminated in special connectors; the internationally accepted standard connector is the BNC, as shown in Fig. 1.4.

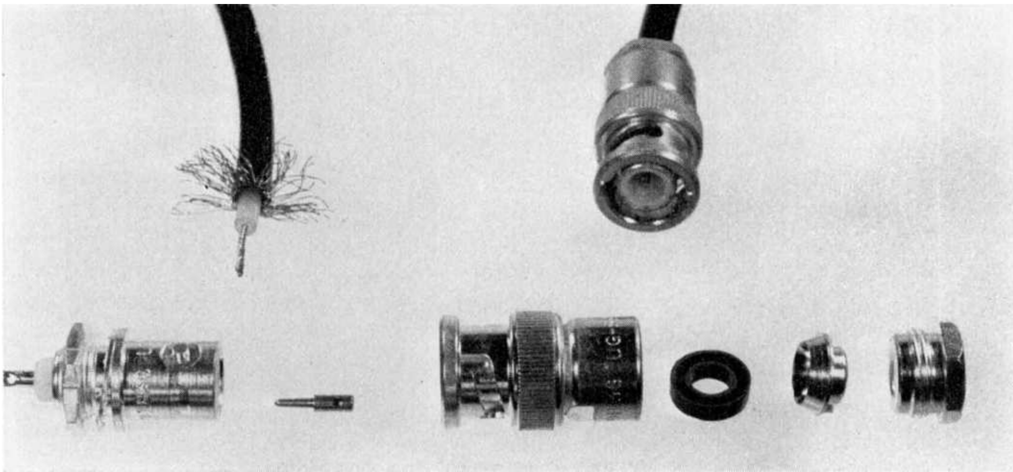


FIG. 1.4. Coaxial cable, BNC connector and socket.

1.5 CONSTRUCTION OF SPECIAL-PURPOSE INSTRUMENTS

The need to construct special-purpose instruments frequently arises in research institutions. The electronics workshop for a research institution, even if quite modestly equipped, can turn out single instruments of professional standard, and there is no reason to tolerate any less.

Steel or aluminium instrument cases are available in a great variety of sizes and shapes, and can form the basis of a new instrument; in many cases fabrication of sheet metal can be avoided altogether. Larger workshops however, normally manufacture their own instrument cases as required, and to produce work of reasonable appearance need a metal-cutting guillotine and a sheet-metal folding machine with a divided head (often called a "pan-brake"). Pieces of a case may be assembled by spot welding (which needs special equipment to be satisfactory for sheet aluminium), or by bolting or riveting (Fig. 1.5). Sheet aluminium is best purchased already hard-anodised with a matt finish, and should be handled carefully to avoid the appearance of scratches in the finished work.

Small holes are made by drilling, and cleaning off any resulting rough edges. A hand-held drill, preferably electric, is essential; an electric pillar drill greatly facilitates many tasks. A small electric grinding wheel to sharpen drills frequently is an absolute necessity; it should not be used for grinding metals other than steel. Larger holes can be produced by a range of punches, or, if a pillar drill is available, by a range of hole saws (Fig. 1.6).

A hole of an odd size can always be produced by cutting a smaller hole to provide entry of a file or saw blade, and then enlarging as required.

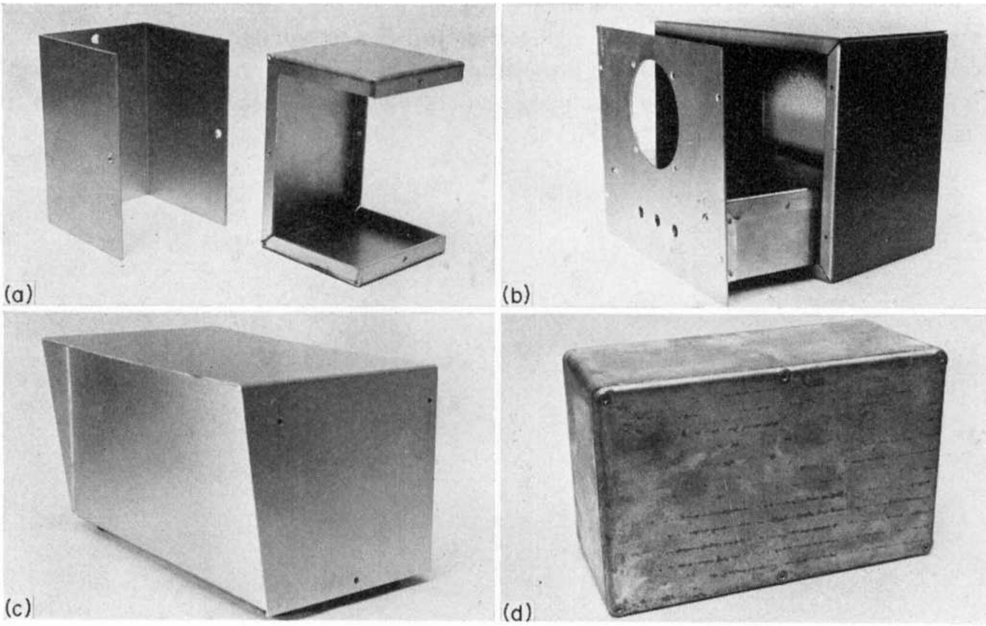


FIG. 1.5. Methods of instrument case construction.

FIG. 1.6. Punches and hole saws.

The lettering of panels for experimental instruments is best done by the use of transfers (Decal or Letraset), applied to a clean anodised aluminium surface; this is subsequently coated by the clear lacquer supplied by the transfer manufacturer.

In the development of a new instrument, the best procedure is first to "mock up" the circuit on the bench, using a piece of insulating board as a support for the work, as shown in Fig. 1.7.

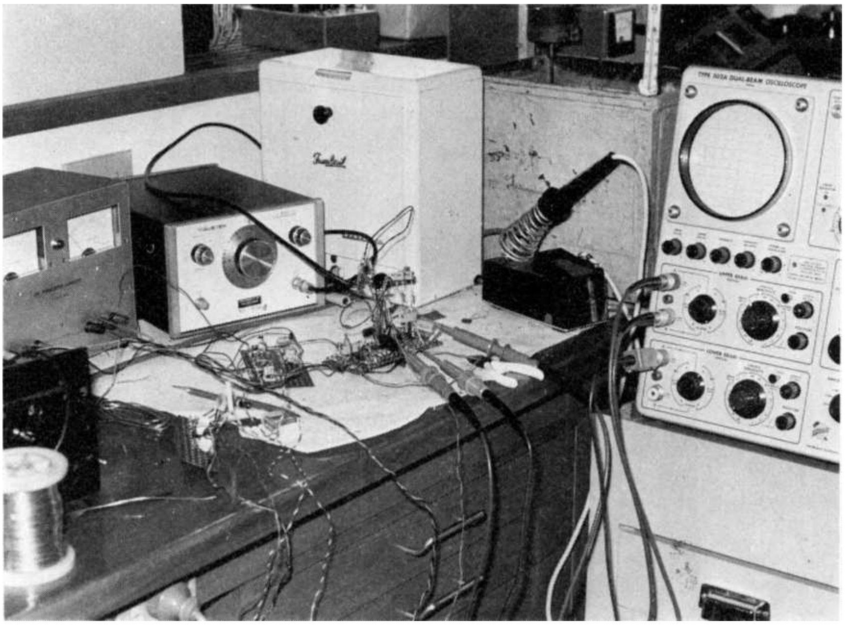


FIG. 1.7. Mock-up of circuit.

When a satisfactory circuit has been developed, it will be necessary to produce a printed board assembly. If only one copy of the instrument is required, the board can be completely manufactured in the workshop. A physical layout is first determined, and is set out accurately on graph paper, with the locations of the component leads marked. This layout is then transferred to a suitable piece of copper-clad board by pricking through the points where leads are to go. The connections are then drawn in on the board by painting bituminous resist in the appropriate places with a fine brush. Etching is carried out in warm ferric chloride solution (36 gm ferric chloride in 100 ml water) for about 10 minutes, using a plastic dish, and rocking the solution gently. When etching is complete, the resist is removed by washing with industrial solvent (lighter fluid or similar), the pricked holes are drilled with a No. 60 (about 1 mm) drill, and the board is ready for assembly (Fig. 1.8).

Where more than one board is required, it is best to carry out the art-work oneself, and then to send it out to a commercial manufacturer. The art-work is usually done on a Mylar sheet, to a scale four times that of the final board. A layout is first sketched, and

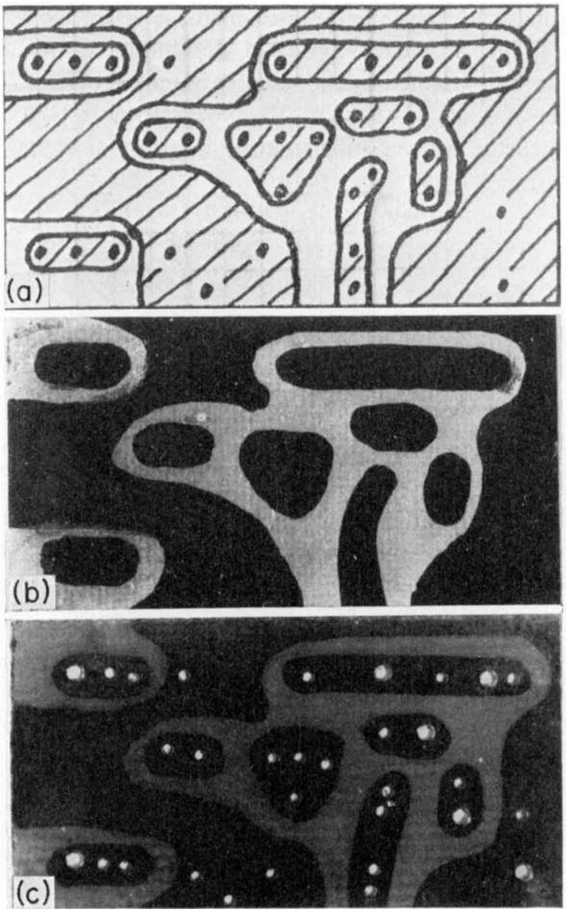


FIG. 1.8. Stages in producing a circuit board.

is then transferred to the Mylar sheet by placing the sheet over a large sheet of detail paper (Fig. 1.9) and executing the connections required in black masking tape.

When completed and carefully checked, the Mylar sheet is sent to the manufacturer, who will reduce it photographically to the correct size. Such a Mylar master may readily be altered if subsequent improvements are made to the circuit.

1.6 SOLDERING

A correctly made soldered joint forms a permanent, reliable electrical and mechanical connection. To obtain such a connection the solder, which is usually an alloy of equal parts of tin and lead (known as 50:50), must be caused to *wet* every part of the joint. This can occur only if the parts are free from a coating of oxide, and are raised to a sufficient temperature during soldering. The removal of oxide is brought about by the application of a *flux*, a substance in which the oxides concerned will dissolve at a temper-

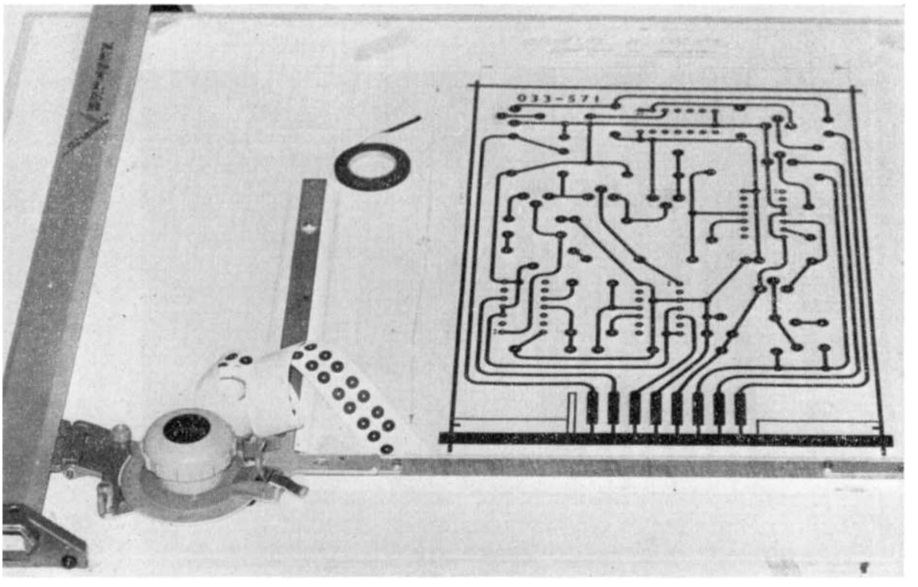


FIG. 1.9. Artwork on the drawing board.

ature lower than that required to melt the solder. Resin compounds are used for electrical soldering, and are incorporated into the solder, which is supplied in the form of thin wire with one or more resin cores. No other flux than this resin should be used under any circumstances; anything else will inevitably lead to slow corrosion in the joint. Since the resin is quite volatile, the solder must be applied to the joint at the same time as the soldering iron, as shown in Fig. 1.10. This also shows an enlarged view of a good joint and a "dry" joint. It is suggested that the beginner inspect his joints with a magnifying glass until he is confident that he has mastered the technique; dry joints are a very common cause of failure in equipment.

A small electric soldering iron of good but conventional construction is recommended; "instant-heating" types should be avoided. The tip must be kept bright and clean during use, and this may be done by frequent wiping on a piece of rag to remove old solder and burnt resin. Since the tip of the iron is usually made of copper, it will slowly dissolve away in the solder, and must be periodically reshaped with a file, and eventually replaced. At the same time, the dissolved copper in the layer of solder on the tip makes an alloy unsuitable for forming a satisfactory joint. Various means of avoiding this erosion have been tried, such as the use of cadmium-plated or iron tips, and solder containing a small percentage of copper, but all of these are harder to use than the simple copper bit and plain solder.

1.7 Failures in equipment

If a record is kept of the occurrence of failures in a number of specimens of one instrument type, it invariably yields a "bathtub" curve, as shown in Fig. 1.11.

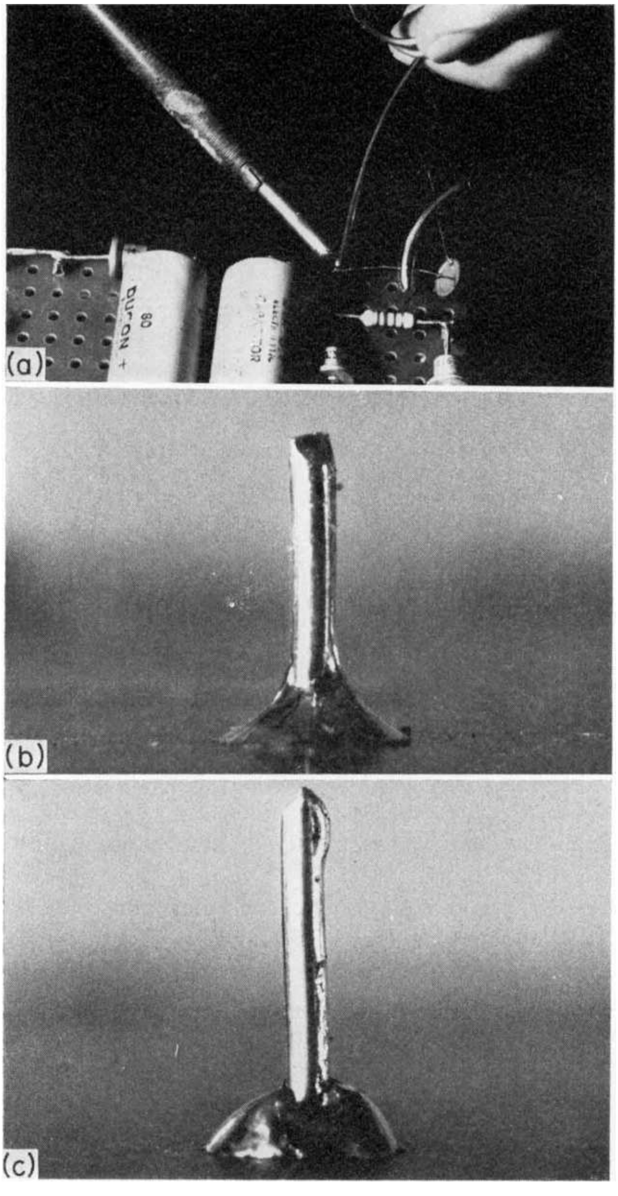


FIG. 1.10. Making a soldered joint.

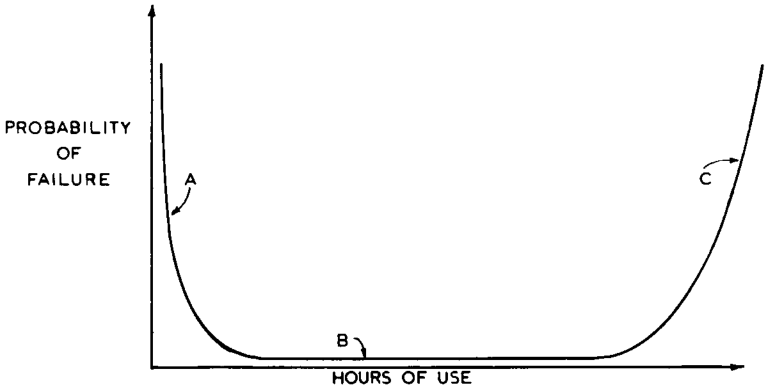


FIG. 1.11. Probability of failure in an instrument.

In region *A*, failures are due to defective components, errors in assembly, or even dry joints. Most of these failures can clearly be kept in the manufacturer's hands and out of sight of the customer by adequate test and inspection procedures; delivery is simply not made until region *A* is almost passed. In region *B*, the working life of the instrument, the failure rate should be very low indeed. As the instrument ages, the failure rate will commence to rise again (region *C*), and although repairs can still be made each time it fails, they will be needed more and more frequently, until the probability of failure in use is unacceptably high, or the cost of repairs uneconomic. Further, with rapid developments in technology, the instrument may well by this time have been rendered obsolete by a more accurate, more reliable, safer, or more convenient successor.

Equipment failures in a properly designed instrument are rare, random and unpredictable events, and no amount of routine maintenance inspection can eliminate or even reduce them; indeed routine maintenance is much more likely to *cause* failures. It should be confined to periodic inspection for superficial faults in knobs, external cables, and the like, and such recalibration procedures as are called for by the manufacturer or user.

FURTHER READING

ARRL Radio Amateur's Handbook, American Radio Relay League, Massachusetts, 1975.
RSGB Radio Communications Handbook, Radio Society of Great Britain.

PRACTICAL

1.1 Assemble two flying leads 50 cm long, and four flying leads 20 cm long, with an alligator clip at each end (Fig. 1.12).
1.2 Wire up the components shown in Fig. 1.13 on a piece of insulating board. This forms a test pulse generator, which will be used later in the course. Carefully inspect the soldered joints with a lens.
1.3 Assemble the amplifier unit of Fig. 1.14 on the printed board supplied, and carefully inspect the soldered joints with a lens.
1.4 If access to sheet-metal working equipment can be arranged, the processes involved in fabricating an instrument case should be observed.

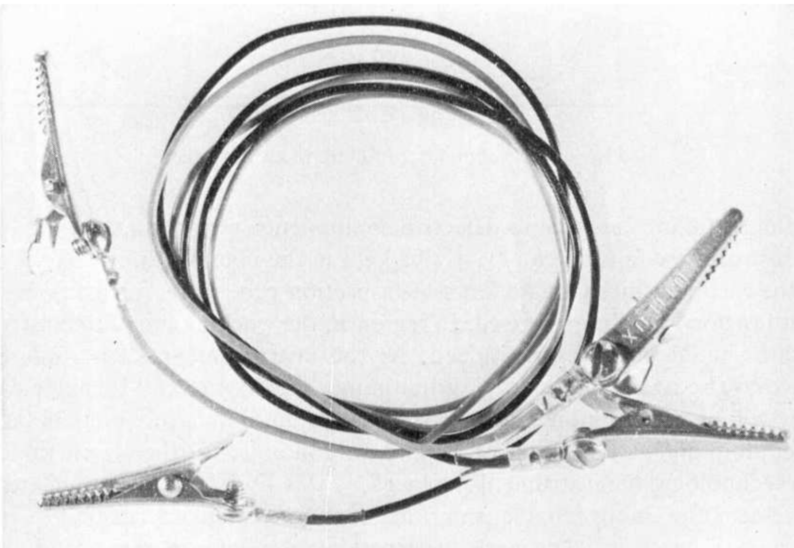
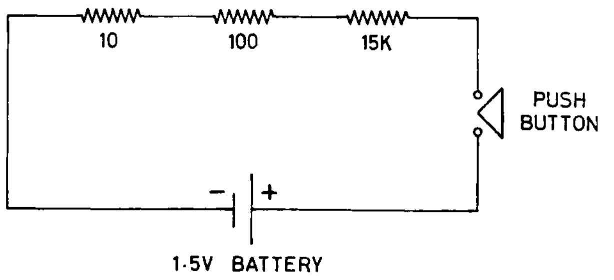
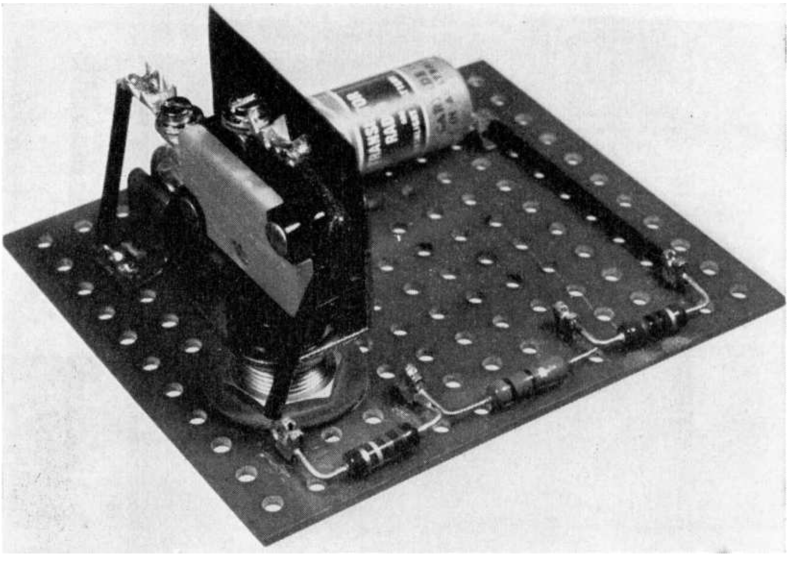


FIG. 1.12. Assembly of flying leads.



(a)



(b)

FIG. 1.13. Test pulse generator.

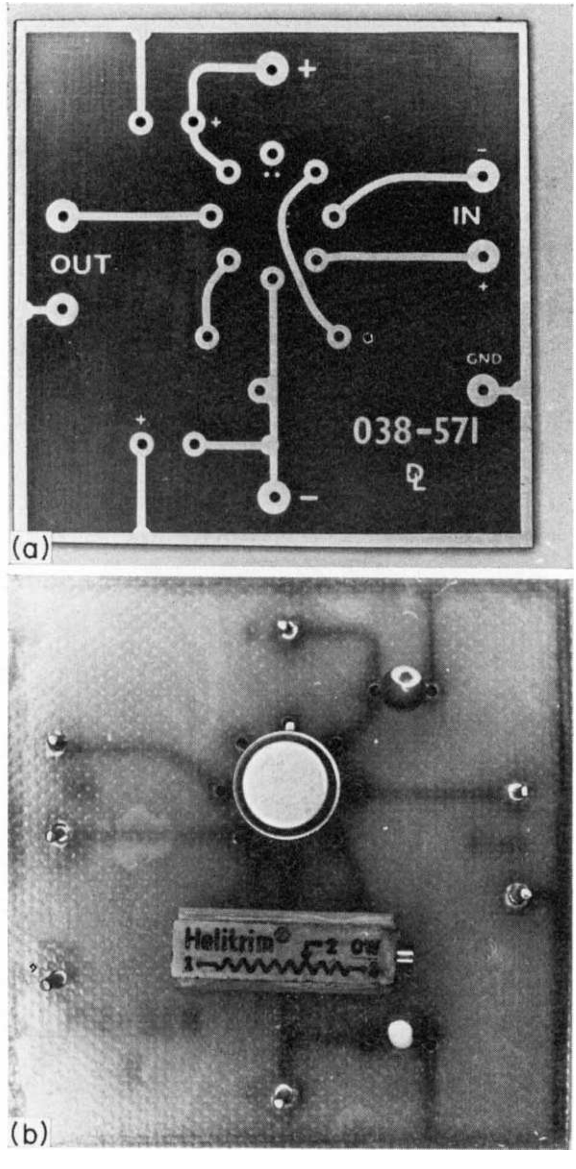


FIG. 1.14. Printed board amplifier unit.

CHAPTER 2

CURRENT, VOLTAGE, AND RESISTANCE

2.1 Definitions

Any solid capable of conducting electricity consists of an array of atoms, usually in a crystalline form. Each atom consists of a central nucleus bearing a positive charge, and a number of orbiting negatively charged electrons, moving in several concentric shells about the nucleus. The electrons in the outermost shell of each atom are capable of moving from one atom to another, and at ordinary temperatures are constantly doing so in great quantities, but in a random fashion.

If a source of electricity such as a battery is applied to the ends of such a conductor, it will inject additional electrons into one end (from its *negative* terminal), and absorb electrons from the other end (into its *positive* terminal), at the same rate. This results in a steady drift of electrons along the conductor, superimposed on the much greater random motion; this drift constitutes an electric current. The practical unit of quantity, or *charge*, of electricity is the coulomb; this represents approximately $6{\cdot}2 \times 10^{18}$ electrons. A flow of 1 *coulomb per second* past a given point is described as a current of 1 *ampere*. If q represents charge in coulombs, and i the current in amperes past a given point, then at any instant

$$i = \text{rate of passage of } q,$$

Using the notation invented by Newton,

$$i = \dot{q}, \tag{2.1}$$

where $\dot{q}$ is the rate of passage of q, in coulombs per second. For a steady rate of passage

$$i = \frac{q}{t}. \tag{2.2}$$

When a source of electricity moves electrons through a conductor, the movement takes place effectively against internal friction; work is done in the conductor, and heat is produced. To produce this motion, it is clear that electricity must be supplied to the ends of the conductor at a definite pressure, usually described as *electromotive force*, or e.m.f. If the e.m.f. is such that for each coulomb that passes through the conductor, 1 *joule* of work is done, and the equivalent amount of heat produced, then the e.m.f. is said to be 1 *volt*. (Frequently the word "voltage" is used as a substitute for e.m.f.) If now the conductor is such that 1 coulomb per second (1 ampere) flows when 1 volt is applied, work is done at a rate of 1 joule per second. Rate of doing work is *power*, and 1 joule per second is 1 watt of power. If W is the power in watts, and e the voltage

$$W = ei. \tag{2.3}$$

The power is also a measure of the rate of heat production in the conductor, and consequently a measure of the rate at which it must *dissipate* heat to its surroundings if its temperature is to remain steady.

Finally, if, as in the last example, the conductor is such that 1 ampere flows when an e.m.f. of 1 volt is applied, the conductor is said to have a *resistance* of 1 *ohm*. If R is the resistance in ohms,

$$R = \frac{e}{i}. \tag{2.4}$$

This last statement is known as Ohm's law. The conductor can also be said to have a conductance (the reciprocal of resistance) of one *siemens*. If G is the conductance in siemens,

$$G = \frac{1}{R} = \frac{i}{e}. \tag{2.5}$$

In routine work, charge is very seldom used, and conductance is less commonly used than resistance. Equations (2.3) and (2.4), however, are the very basis of all design and all testing of equipment. If two practical points are noted, practically all calculations with equations (2.3) and (2.4) can be done mentally with adequate accuracy. First, it is generally quicker to work in terms of volts, *milliamperes*, *milliwatts* and *kilohms*, which form a consistent set of derived units, than in volts, amperes, watts and ohms. Secondly, the best way to remember Ohm's law is by the rule "If volts are known, divide into them; if volts are not known, multiply the given quantities to obtain them". Thus 20 V applied to 5 K (as kilohms are often written) gives 4 mA.

Equations (2.3) and (2.4) are frequently used in combination, to give

$$W = i^2R. \tag{2.6}$$

2.2 Resistors

In practical electronic circuits it is frequently required to produce a known current from a given voltage, or vice versa, and circuit elements consisting of conductors made to have a known amount of resistance are very common. These are known as *resistors*. Resistors are specified in terms of (a) their nominal resistance, (b) the percentage variation in resistance permitted by the manufacturer (the *tolerance*) and (c) the maximum power they can dissipate without an excessive rise in temperature. Thus a typical specification might be 100 K $\pm$ 5% $\frac{1}{4}$ W. (In this book it will be assumed that all resistors have $\frac{1}{4}$ watt dissipation unless otherwise specified.) For dissipations up to 1 W (few circuit requirements exceed this) resistors are commonly made of a carbon composition, typically in sizes of 0·1, 0·25, 0·5 and 1·0 W. Above this they are usually of a metal oxide. Modern first grade resistors should be used throughout any piece of electronic equipment; a small economy here will almost inevitably lead to trouble. 0·1 W resistors should be used only if size is extremely important; they are markedly less stable than the larger sizes.

Commercially available resistors appear at first sight to have rather odd values of resistance; in fact they are arranged in a logarithmic series of *preferred values*, with their tolerances overlapping slightly. Thus in the decade from 10 to 100 ohms for 10% tolerance resistors, the values found are 10, 12, 15, 18, 22, 27, 33, 39, 47, 56, 68, 82 and 100 ohm, and this series is repeated in each decade. Even when 5%, 2% or 1% tolerance resistors are used, it is most unusual to find values other than those in the 10% series, although they are available on order. Resistors are colour coded, as shown in Fig. 2.1, using four bands of colour starting from one end. With a little practice, common values can be identified at a glance. The colours used are shown in Table 2.1.

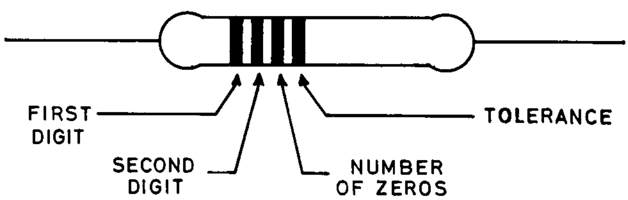


FIG. 2.1. Colour coding of resistors.

TABLE 2.1. COLOUR CODING OF RESISTORS

Digit	Colour	Percentage tolerance	Colour
0	black	±20	none
1	brown		
2	red	±10	silver
3	orange		
4	yellow	± 5	gold
5	green		
6	blue	± 2	red
7	violet		
8	grey	± 1	brown
9	white		

The same colour code is often used to identify conductors in a multi-cored cable, or on components other than resistors.

An electronics workshop will require reasonable stocks of all values of resistors from 1 ohm to 10 megohm in 10% steps; $\frac{1}{4}$ watt resistors are recommended for routine use. Some $\frac{1}{2}$ watt and 1 watt resistors will be required from time to time.

2.3 ELECTRICAL NOISE IN RESISTORS

In highly sensitive equipment, such as is frequently used to record the very small voltages occurring in living tissues, the factor which determines the smallest signals that can be detected is the random movement of electrons in the equipment itself. This is referred to in general terms as *noise*. Although transistors make their own contribution to this noise, every resistor also contributes, in two ways. The first of these is due to the random motion of its outer shell electrons, as discussed in § 2.2 above. This is *thermal agitation noise*, which appears as a very small fluctuating voltage across the ends of the resistor, and whose magnitude depends on the value of the resistance and the absolute temperature. The second is due to imperfections in the structure of the resistor, which appear as slight fluctuations in the value of its resistance when a current flows through it. These imperfections are somewhat greater in standard carbon resistors than in special low-noise ones, but modern standard resistors are suitable for nearly all purposes. Usually amplified noise from the resistors near the input of a biological amplifier swamps that due to any other source.

2.4 Combinations of resistors

Resistors may be connected in *series*, as shown in Fig. 2.2(a), or in *parallel*, as in Fig. 2.2(b). In the first case the total resistance is the sum of the two individual resistances:

$$R_{\text{total}} = R_1 + R_2. \tag{2.7}$$

In the second case, the total conductance is the sum of the two individual conductances:

$$G_{\text{total}} = G_1 + G_2. \tag{2.8}$$

Since it is more convenient to work in terms of resistance, this becomes

$$\frac{1}{R_{\text{total}}} = \frac{1}{R_1} = \frac{1}{R_2} \tag{2.9}$$

For ease of calculation, equation (2.9) may be rearranged as

$$R_{\text{total}} = \frac{R_1 R_2}{R_1 + R_2} \tag{2.9a}$$

the "product-over-sum" rule.

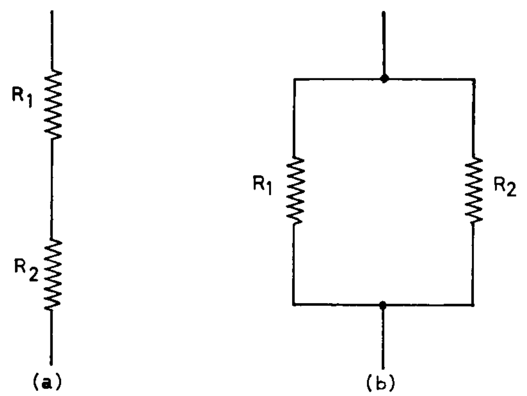


FIG. 2.2. Resistors, (a) in series, (b) in parallel.

In either the series or the parallel combination, the total power dissipated is the sum of that dissipated in each resistor. This fact is sometimes used to replace one resistor of large dissipation by two smaller ones; these may each have twice the desired resistance and be placed in parallel, or each have half the desired resistance and be placed in series.

A circuit configuration often found is the *voltage divider*, as shown in Fig. 2.3. If two resistors in series are placed across a source of e.m.f., a fraction of the source voltage will appear across each resistor. This fraction is directly proportional to the fraction of the total resistance represented by that resistor: in Fig. 2.3

$$V_{\text{out}} = \frac{R_2}{R_1 + R_2} V_{\text{in}} \tag{2.10}$$

This relationship only holds provided that no current is taken from the terminals marked V_{out}.

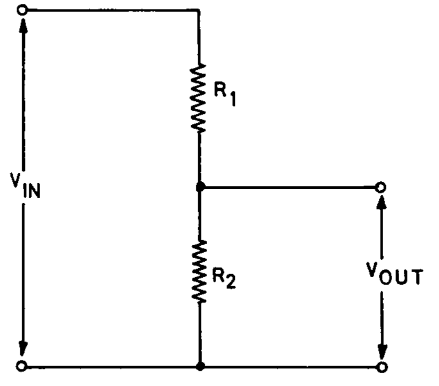


FIG. 2.3. The voltage divider.

Variable voltage dividers, universally (and wrongly!) called *potentiometers*, are often encountered. These constitute the main control devices operated by the knobs in most equipment, and are also incorporated at many points within equipment to permit its initial calibration; in this case they are intended to be set with a screwdriver. They may have a carbon, metal glaze, or wire-wound track, over which a wiping contact is moved by means of a control shaft; a selection of types suitable for use in biomedical equipment is shown in Fig. 2.4. Wire-wound potentiometers are not usually available in values above 50 K, and have a linear relation between angle of rotation and output voltage. Carbon potentiometers, unless specifically ordered with a linear relationship, usually have an approximately logarithmic one, since they are often used to control apparent loudness of a radio set or record player, and this is a logarithmic function of output voltage.

The type of potentiometer ordinarily sold for use in radio and television equipment is quite inadequate for use in biomedical equipment, and will give an excessive failure rate; only types approved for military or aviation use should be considered, despite their relatively high cost.

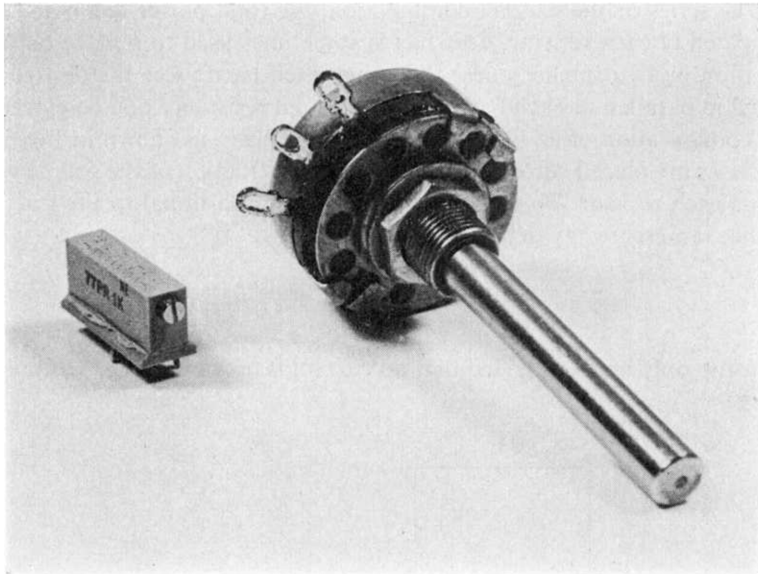


FIG. 2.4. Potentiometers, (a) for front panel use, (b) for screwdriver adjustment.

2.5 POWER SOURCE FOR BIOMEDICAL EQUIPMENT

Since reliability of operation, very often under conditions of intermittent use and no skilled maintenance, is an essential for biomedical equipment, the use of the power lines as a source of supply is strongly recommended. There has been considerable discussion about the enhanced safety of battery-operated equipment, but *with adequate design* of instruments, and also of the wiring of critical areas such as operating theatres and intensive care wards, no risk attaches to the use of line-operated equipment. On the other hand, a battery-operated instrument will almost certainly be inoperative when required in an emergency.

If batteries have to be used for reasons of portability or use in the absence of line power, rechargeable hermetically sealed nickel-cadmium or lead-acid batteries are recommended. These consist of a number of cells arranged in series to give the required voltage. The cell size is rated by the number of *milliampere-hours* it can deliver, usually based on a 10 hour discharge time. Provision for recharging from line power should be built into the instrument, unless size or weight precludes it, and some form of indication of the state of charge of the battery should be provided.

FURTHER READING

ARRL Radio Amateur's Handbook, American Radio Relay League, Massachusetts, 1975.
HUGHES, *Electrical Technology* (4th ed.), Longmans, London, 1969.
Manufacturers' literature on resistors.
Manufacturers' literature on sealed nickel-cadmium and lead-acid batteries.

PRACTICAL

2.1 Connect up the circuit of Fig. 2.5, using four 33 K resistors in series for R. Carefully measure and tabulate the current for $E = 0$ V, 1·5 V, 3 V, 4·5 V and 6 V. Plot a graph showing current as a function of voltage, and draw the straight line of best fit. How much do individual points depart from this line, and why? Using Ohm's law, deduce the value of R from this line.

2.2 From the results obtained above, tabulate the power dissipated in R as a function of the current flowing (eqn. 2.6), and plot a graph to show it. Calculate the maximum current R could carry without exceeding the rated maximum dissipation of the resistors.

2.3 Using a 1·5 V battery for E, carefully measure and tabulate the current for R = 132 K, 99 K, 66 K and 33 K (but *not* zero!). Plot a graph showing current as a function of resistance, and a graph showing current as a function of conductance. Why should this latter graph be a straight line? Using Ohm's law, deduce the true value of E from it.

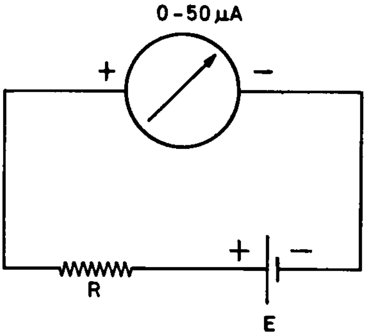


FIG. 2.5. Application of Ohm's law.

CHAPTER 3

METERS

3.1 Basic moving-coil meter

By far the commonest type of meter used in modern electronic instruments is the moving coil or d'Arsonval. This instrument is in effect a miniature electric motor, in which rotation is opposed by a spring. The greater the current that passes through it, the further it turns before the opposition of the spring brings the pointer to rest.

In a modern meter, a uniform magnetic field is supplied by a carefully designed and constructed permanent magnet of almost circular form, as shown in Fig. 3.1; its poles are arranged to embrace a cylindrical iron core in such a way as to form an accurately radial magnetic field.

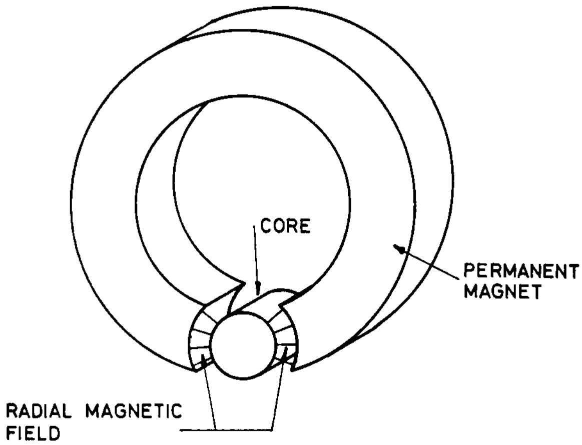


Fig. 3.1. Magnet of d'Arsonval meter.

In this field is suspended a rectangular coil of fine wire, wound on a light aluminium bobbin, as shown in Fig. 3.2. The suspension may consist either of a hard steel pivot at either end, which runs in a sapphire cup, or of a tightly stretched flat bronze band at either end (a *taut band* suspension). In the first case, a coil spring at either end provides the opposing force for the coil when it is deflected, and also leads the current in and out; in the second case the taut bands carry out these functions.

A typical modern meter gives a full scale deflection for 50 μA of current through the coil, which is wound of very fine copper wire. It will usually have a resistance between 2 K and 4 K. Meters with sensitivities down to 5 μA for full scale deflection are obtainable.

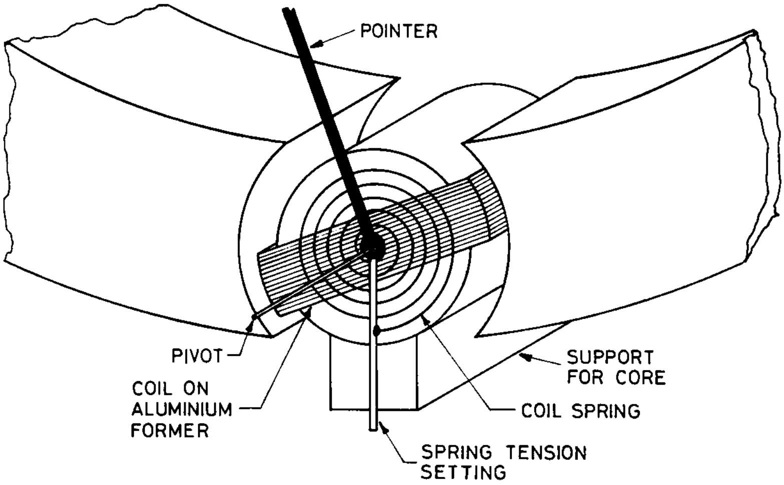


FIG. 3.2. d'Arsonval movement.

3.2 SHUNTING TO READ LARGER CURRENTS

Meters with basic movements of up to 10 mA are available, but it is more usual to start with a basic 50 μA movement, no matter what current range is required, and to *shunt* the meter with a low resistance, so that all the desired full scale current except for 50 μA passes through the shunt (Fig. 3.3). Starting with a 50 μA 2 K movement, if a 5 mA

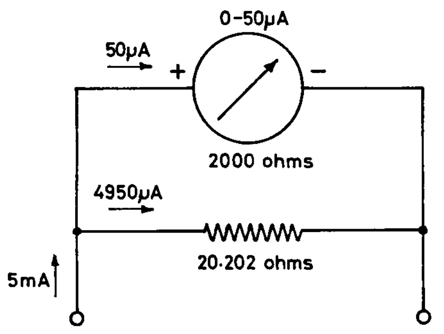


FIG. 3.3. Meter shunt.

full scale deflection is required, the meter scale is relettered to read 0–5 mA. When 5 mA is flowing, 4950 μA must pass through the shunt, leaving 50 μA through the meter. Accordingly, the shunt path must be 4950/50 as easy for the current as the 2 K path through the meter; it must therefore have a resistance of

$$2000 \times \frac{50}{4950} \text{ ohms}$$

or 20·202 ohms. In the general case,

$$R_S = \frac{R_m}{M-1} \tag{3.1}$$

where R_S is the shunt resistance required, R_m is the meter resistance, and M is the factor by which the basic range of the meter is to be multiplied.

In practice, a shunt is first made of copper wire to a resistance somewhat higher than the calculated value, and is then tested by passing a known current through the combination of meter and shunt. By successively switching off, shortening the shunt, allowing the newly soldered joint to cool, and retesting, the meter is finally made to read correctly.

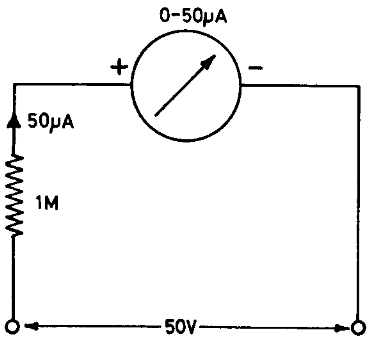


FIG. 3.4. Microammeter as voltmeter.

3.3 USE OF A MICROAMMETER AS A VOLTMETER

By setting up a microammeter with a series resistor, it may be used as a voltmeter (Fig. 3.4). For example, if a range of 0–50 V is required, the application of Ohm's law will show that 50 V will drive 50 μA through 1 megohm (1 M for short). If the total resistance is 1 M, the series resistor must be (1000−2) K, or 998 K. Since an error of 2 parts in 1000 would not be significant a 1 M 1% resistor would be used. The meter scale would be recalibrated to read 0–50 V. In the same way, any other voltage range may be set up; notice that without any series resistor the microammeter reads full scale for $0{\cdot}05 \times 2$, or 100 mV.

3.4 USE OF MICROAMMETER AS OHMMETER

To measure a resistance, a known e.m.f. is used in series with the microammeter and the resistance; the current is observed, and the value of the resistance obtained by Ohm's law. To avoid a tedious calculation every time it is desired to measure a resistor, it is usual to draw up a graph of resistance against meter current, and from it to produce a special meter scale calibrated directly in ohms; this scale will be non-linear, and will increase from right to left.

This simple scheme has a major disadvantage; if the ohmmeter terminals were accidently short-circuited, the meter would be destroyed. This may be overcome by inserting a resistor permanently in series with the battery and microammeter, such that a short-circuit externally (zero ohms) just produces a full-scale reading. By making part of this resistor variable, it is also possible to correct partly for changes in battery voltage as the battery ages (Fig. 3.5 and Table 3.1). To use the instrument the terminals are first

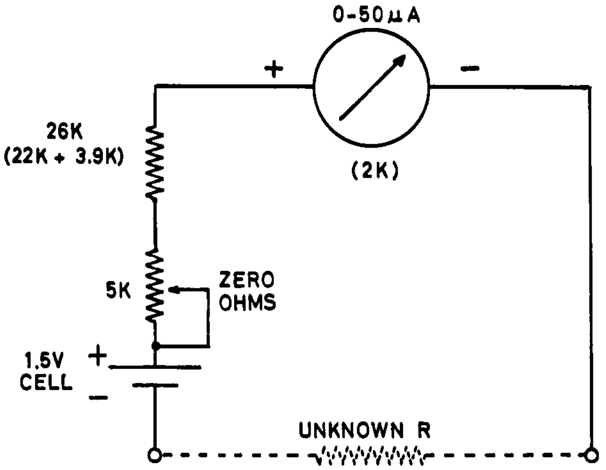


FIG. 3.5. Microammeter as ohmeter.

shorted, and the ZERO OHMS control adjusted to give a full scale deflection on the meter. The unknown resistor is then placed between the terminals, and the resistance read off from the precalibrated scale.

TABLE 3.1. CALIBRATION OF OHMMETER

Meter current (μA)	0	10	20	30	40	50
Total R for battery voltage of 1·50 (by Ohm's law)	infinity	150 K	75 K	50 K	37·5 K	30 K
Unknown R for battery voltage of 1·50 (total R − 30 K)	infinity	120 K	45 K	20 K	7·5 K	0

When the battery voltage falls below 1·4 V, it will no longer be possible to set full scale deflection on shorting the terminals, and the battery must be replaced. At this stage all scale readings will be 6·6% high, e.g., a 42 K resistor would read 45 K on the scale. Since there are very few situations in a piece of equipment where this much error would affect normal operation, this degree of accuracy is quite acceptable for most purposes. If high accuracy is required, measurement should be carried out on a bridge, as discussed in Chapter 4.

3.5 MULTIMETERS

For practical use in electrical and electronic testing, a single basic meter is normally converted into a *multimeter* by the addition of a variety of shunts, series resistors, and sources of e.m.f.; the appropriate function can be selected at will by a rotary multi-position switch. When purchasing a commercial instrument, it is the calibration rather than just an assembly of components which is being acquired. There is little advantage in attempting to build a multimeter oneself, except to learn what can and cannot be expected of such an instrument in terms of accuracy and method of use. The instrument which is to be constructed in the practical session has only a limited number of ranges, but it does clearly show the principles of construction, and has a wide variety of uses.

3.6 Insertion errors

An ideal measuring instrument does not in any way affect the quantity which is to be measured; a practical instrument does not affect it by a significant amount. A moment's thought will show that an ideal current meter should have no resistance, and require no voltage to operate it, while an ideal voltmeter should have an infinite resistance, and draw no current from points to which it is applied. A multimeter based on a 50 μA 2 K meter will satisfy these requirements in most circumstances. As an ammeter, the 100 mV it requires is not significant in circuits where the applied e.m.f.s are several volts or greater; as a voltmeter, 50 μA is usually small compared with the currents actually flowing past the points where voltage measurements are being taken. However, circumstances do arise where the resulting errors are serious, and the user must continually be aware that errors can arise when measuring currents from low voltage sources, or voltages in high resistance circuits. An easy test for error is to measure the quantity concerned with two meters (in series for current, in parallel for voltage), and then to remove one of them, and observe to what extent the reading of the other alters.

In this situation it may be possible to deduce that magnitude of the desired quantity by measurements elsewhere in the circuit, followed by the application of Ohm's law; otherwise a more sensitive meter will be required.

3.7 Meter accuracy

A typical 50 μA meter with a scale diameter of about 10 cm will have an accuracy as quoted by the manufacturer of $\pm 2\%$ of full scale deflection; that is, there may be an error of $\pm$ 1 μA at any point on the scale. When new, the full scale reading will have been set to be correct, but since in ordinary meters the scales are mass-produced, slight non-linearities in reading may occur elsewhere. More expensive meters, in which the scales are individually calibrated, will have accuracies of $\pm$ 1% or even better, but even these will require periodic recalibration to maintain this figure.

For higher accuracy there is an increasing tendency to turn to the use of *digital* meters, in which the scale is replaced by a direct read-out, to three or more decimal places, with numerical indicators. The mechanism of these instruments is based on the principles to be discussed in Chapter 4. It should be noted that these instruments are *no more immune from insertion errors* than are moving coil meters, though they are usually of higher sensitivity. Further, they should be subjected to periodic calibration checks, since they can lose accuracy like any other meter.

FURTHER READING

Terman and Pettit, *Electronic Measurements*, McGraw-Hill, New York, 1952.
Scroggie, *Radio and Electronic Laboratory Handbook* (8th ed.) Iliffe, London, 1971.

PRACTICAL

3.1 The circuit of the multimeter to be constructed is shown in Fig. 3.6, and photographs of the front and rear of the completed instrument in Figs. 3.7 and 3.8. As shown, it is assembled in a diecast box, with an aluminium overlay marked with the switch positions. The only difficulty likely to arise is in the assembly of the switch, and this should be clear from a comparison between the circuit of Fig. 3.6 and the detail of Fig. 3.9. The switch should be assembled as far as possible before mounting it in the instrument.

3.2 When assembly is complete, carefully check the circuit by visual inspection. Test the current ranges with suitable batteries and resistors, using a standard meter for comparison. Test the voltage ranges with suitable batteries or other supplies.

3.3 Using Ohm's law, draw a calibration graph for the instrument as an ohmmeter. Use this calibration to measure the 33 K resistors from the experiment in Chapter 2.

3.4 Produce several series, parallel, and series parallel combinations of the 33 K resistors. Calculate the resistance of each combination, and then measure it and compare the results.

3.5 Set up a voltage divider of two 100 K resistors in series across a 6 V battery. What is the voltage across each resistor? Try to measure this voltage with your multimeter.

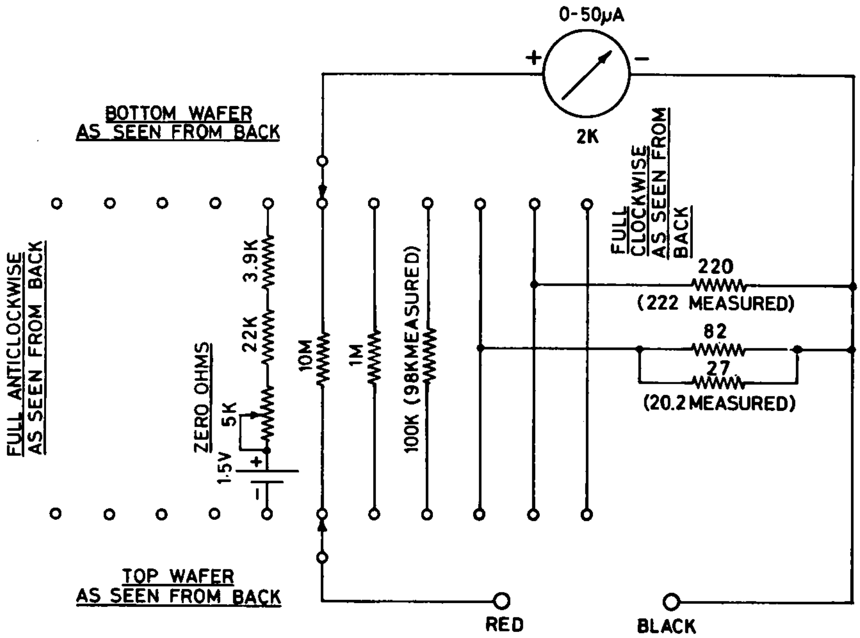


FIG. 3.6. Circuit of multimeter.

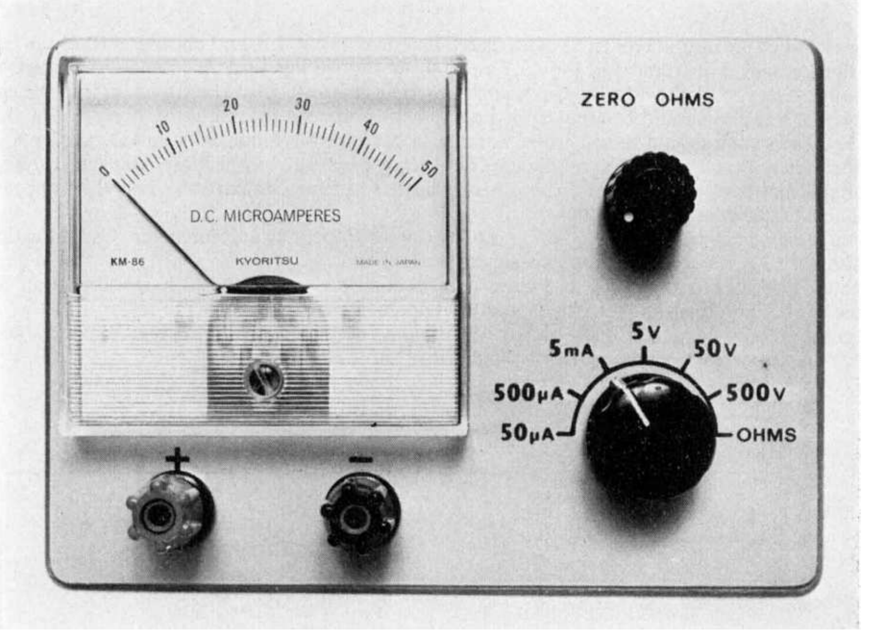


FIG. 3.7. Front of completed multimeter.

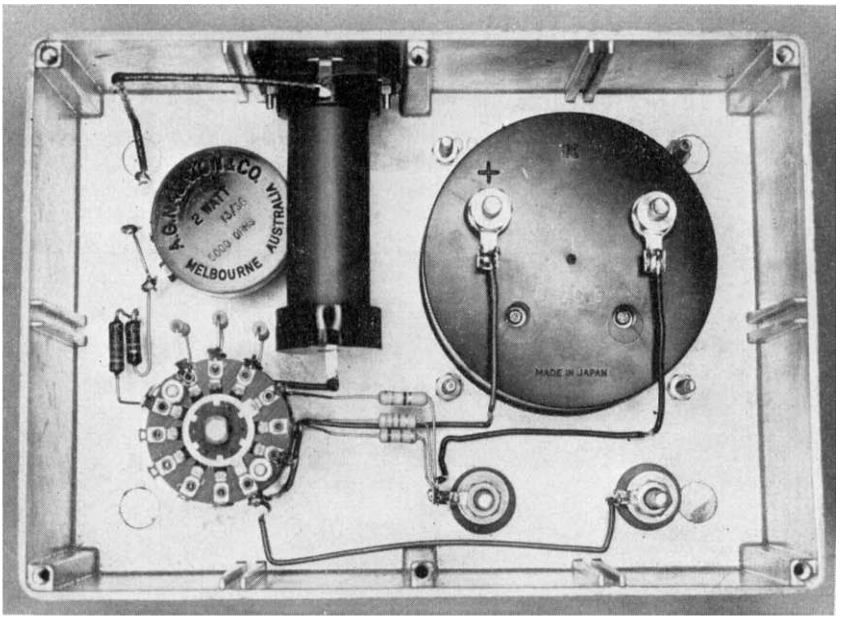


FIG. 3.8. Rear of completed multimeter.

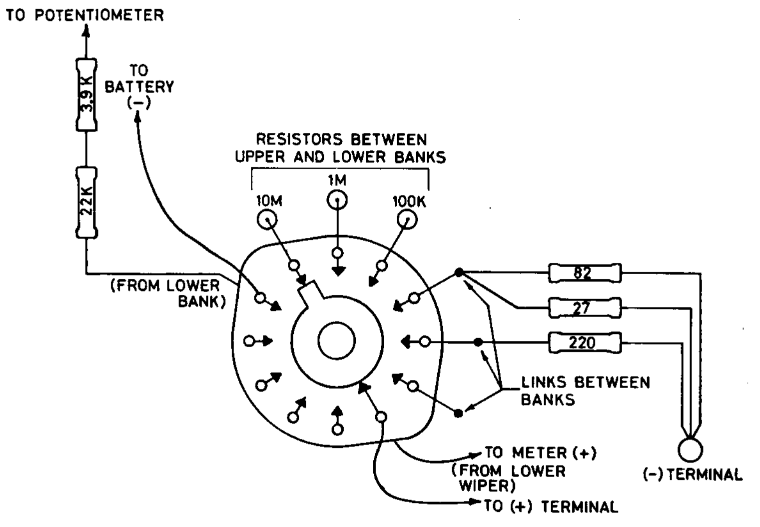


FIG. 3.9. Switch detail.

CHAPTER 4

THE POTENTIOMETER AND WHEATSTONE BRIDGE

4.1 Use of potentiometer and Wheatstone bridge

These two instruments are basically intended for high precision laboratory measurements of voltage and resistance. The potentiometer, developed by Kohlrausch in Germany, is a device for comparing an unknown with a standard voltage. The Wheatstone bridge, developed by Sir Charles Wheatstone in Britain, is a device for comparing an unknown with a standard resistance. Each is capable of an accuracy of six significant figures if properly constructed and used. Adaptations of these instruments are very commonly used for less precise measurements in the workshop and for use in thermometry.

4.2 The potentiometer

The principle of the potentiometer is shown in Fig. 4.1, and it will be seen that it is essentially a calibrated voltage divider. A long uniform wire is set up beside a scale calibrated in volts (in this case 0 to 1·5) and its ends are connected to a 2 V battery A in series with a variable calibrating resistor R. Any voltage V from 0 to 1·5 V can be picked off the slide wire between the origin and the moving point. This voltage is connected in series with a sensitive meter and a source of e.m.f. E less than 1·5 V, which is to be measured. Since E is connected in opposition to V, the meter will read zero when they are exactly equal.

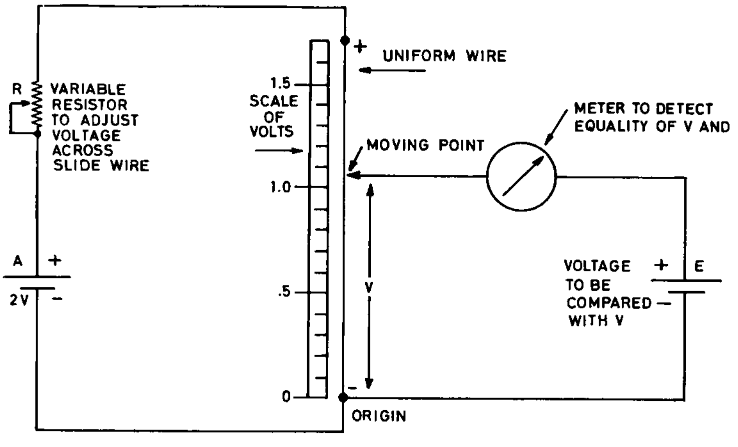


FIG. 4.1. Potentiometer circuit.

The potentiometer is first calibrated by putting an accurately known standard source of voltage at E (suitable *standard cells* are available for this purpose), setting the moving point to the value of the standard cell voltage on the scale, and adjusting R until the meter reads zero. The scale is now reading correctly for its whole length. An unknown source of e.m.f. may now be substituted at E, and the moving point adjusted until the meter again reads zero. The unkhown e.m.f is then read off the scale. It should be noticed that when the reading is taken, *no current* is being drawn from the slide wire, and it functions as a true voltage divider; also no current is being taken from the standard cell, and it gives its true voltage.

One of the most useful commercial forms of the potentiometer is the self-balancing recording instrument. In this device the meter is replaced by a very sensitive electronic *amplifier*, whose output drives a small electric motor, and this slides the moving point along the wire. If the potential V and the e.m.f. E are not equal, the motor will run until they become equal, and this is the desired condition of balance. The moving point also carries a pen, which moves across a chart, and this is driven forward by a clockwork movement. In this way a continuous accurate record of the e.m.f. E is kept. Such instruments are widely used in research and industry; they are discussed further in Chapter 27. In a typical modern recorder the chart has a width of 250 mm, 1 second is required to traverse it completely and it has an accuracy of 0·1%.

4.3 Thermocouple thermometry

The potentiometer (particularly in the form of a recording instrument) is widely used for *thermocouple thermometry*. If two dissimilar metals are connected as shown in Fig. 4.2, and the junctions 1 and 2 are held at different temperatures, a small e.m.f. will be generated, which will be proportional to the difference in temperature, and of a magnitude depending on the two metals. For a temperature difference of 100°C, the e.m.f. is typically about 5 mV. For accurate measurements, one junction is held at a constant temperature (often at 0°C, in melting ice) and the other junction is used as a temperature-measuring probe. It is inconvenient to maintain a melting ice reference, and this may be dispensed with if the reference junction is maintained at room temperature, and the potentiometer fitted with a special compensating unit.

For temperatures up to about 500°C, metal A may be copper, and metal B the alloy *constantan*, which is frequently used in the manufacture of wire-wound resistors. For higher temperatures A may be platinum, and B an alloy of platinum and iridium.

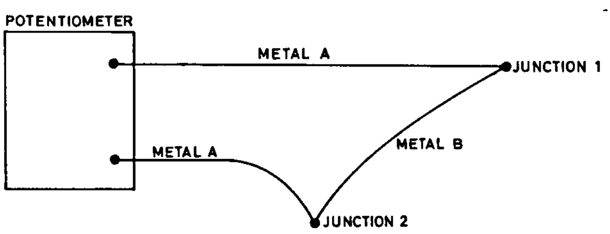


FIG. 4.2. Thermocouple thermometer.

4.4 Thermoelectric cooling elements

If current from a battery is passed through a thermocouple, one junction is warmed slightly, and the other cooled slightly. This *Peltier effect* is superimposed on the overall resistive heating due to the passage of the current, and until recently was a scientific curiosity only. It has now become possible to make thermocouples of materials which show a very marked Peltier effect, and at least on a small scale provide a means of obtaining refrigeration directly by passage of an electric current. A battery of thermocouples in series is usual; the typical unit shown in Fig. 4.3 requires 25 A at 2 V to operate, and can maintain a temperature difference of 90°C between its faces. If the heated junctions are cooled by a large heat sink or circulating water, temperatures below 0°C at the cooled junctions are easily obtained. These devices are in common use in such applications as freezing stages for microtomes; their expense at present precludes their use for domestic refrigeration.

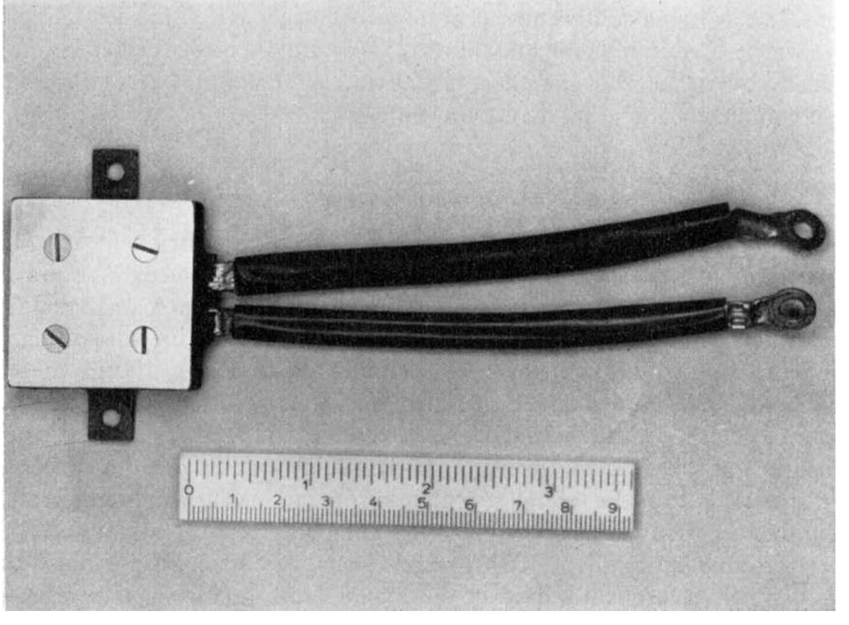


Fig. 4.3. Thermoelectric cooling element.

4.5 The Wheatstone bridge

The Wheatstone bridge is used for the measurement of unknown resistances, by comparing them with a standard resistance. The basic circuit is shown in Fig. 4.4. Omitting the meter G for the moment, this circuit may be regarded as two voltage dividers, ABD and ACD, connected in parallel across the source of voltage E.

The potential at B will be $R_2/(R_1 + R_2)$ of E, and the potential at C will be $R_4/(R_3 + R_4)$ of E.

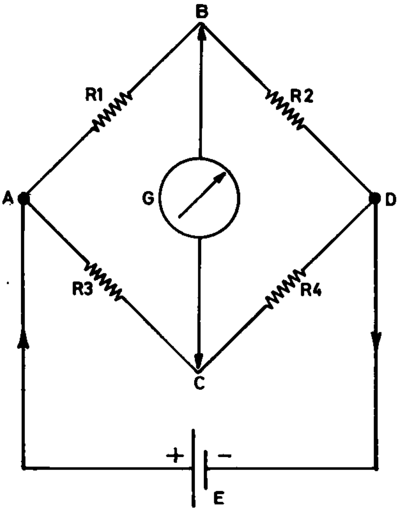


FIG. 4.4. Wheatstone bridge.

Let us find the condition for these two potentials to be equal, so that inserting the meter G between B and C will produce no current from B to C through the meter. This will occur when

$$\frac{R_2}{R_1 + R_2} E = \frac{R_4}{R_3 + R_4} E, \tag{4.1}$$

i.e., when

$$\frac{R_2}{R_1 + R_2} = \frac{R_4}{R_3 + R_4}. \tag{4.2}$$

This reduces to

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}. \tag{4.3}$$

In this condition, where the meter reads zero, the bridge is said to be *balanced.*

If R_3 and R_4 are equal resistors, balance occurs when $R_1 = R_2$; if R_1 is an unknown variable resistor, R_1 may be measured by setting R_2 for balance.

Whatever the ratio of R_3 to R_4, the same ratio of R_1 to R_2 will give balance. For example, if R_3 is ten times R_4, and the bridge is balanced, then R_1 is ten times R_2. In this way resistors very much larger or smaller than the standard R_2 can be measured.

Although the condition of balance is independent of E, the deflection of the meter G for a small imbalance is directly proportional to E, and is greatest when the four arms are of equal resistance.

As a laboratory instrument, the Wheatstone bridge is highly accurate, and although quite suitable for a biomedical electronics workshop, is rather slow to use. Less precise bridges, which give an accuracy of between 0·1 and 1% are available for use in the workshop, but have been largely superseded by *digital ohmmeters* of the same order of accuracy.

4.6 Resistance thermometry

The Wheatstone bridge is widely used for *resistance thermometry*. All metallic conductors increase in resistance as the temperature is raised; the two most suitable for resistance thermometry are nickel and platinum. Both of these increase in resistance by about 0·4% for each degree Celsius rise.

In addition to metals, a large range of semiconductor materials shows considerable resistance variations with temperature, usually giving a fall in resistance as the temperature rises. This property is used in the construction of *thermistors*, which are made in a variety of physical shapes, some extremely small. A typical thermistor consists of a bead of semiconductor material between two fine platinum wires, and mounted in a small glass envelope (Fig. 4.5). The resistance of a thermistor is usually quoted at 20°C; typically it will halve for every 25°C rise in temperature. Notice that the resistance is not linear, but exponential, as a function of temperature.

The Wheatstone bridge circuit is often used as a pseudo-bridge. In this form three arms are fixed in resistance, and the fourth, which may be a thermistor, is allowed to vary. The meter current is used as a measure of resistance of the fourth arm. Although the meter current depends on the voltage of the supply used, and is not directly proportional to the resistance of the fourth arm, a useful instrument results. If the fourth arm varies only over a small range near balance, the meter current is reasonably close to being directly proportional to the resistance of the fourth arm; this condition obtains in many applications. In the case of thermistor thermometry, the pseudo-bridge is particularly useful, since the curvature of the thermistor characteristic almost compensates for the

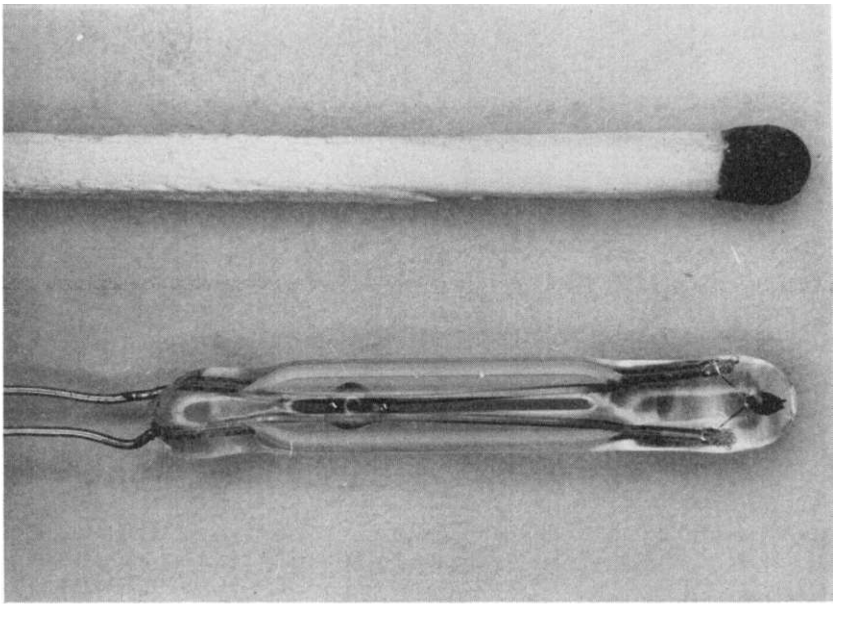


Fig. 4.5. Bead thermistor. A match is shown for comparison beside the thermistor.

curvature of the bridge characteristic over quite a wide range. For further details, reference should be made to the paper by Cole listed at the end of the chapter.

The circuit of a simple thermistor thermometer is shown in Fig. 4.6, and the front and rear views of such an instrument in Fig. 4.7. The meter used has a 50 μA movement, and is calibrated for the range 20 to 45°C. To calibrate the instrument, the thermistor is placed in water at exactly 20·0°C, and potentiometer 1 is adjusted to balance the bridge, giving a meter reading of 20°C. (This balance will be independent of the setting of potentiometer 2 or the battery voltage.) Then the thermistor is transferred to water at a temperature in the range 40–45°C, and potentiometer 2 adjusted to give the correct reading. The instrument is then calibrated over the whole range.

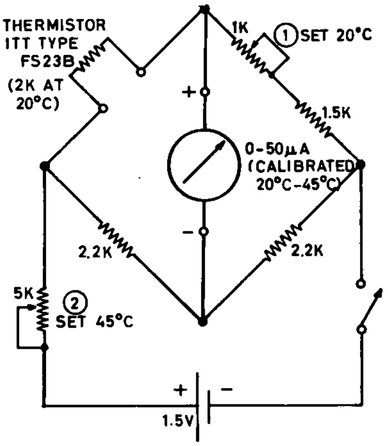


FIG. 4.6. Circuit of thermistor thermometer.

This simple instrument has two serious defects. First, since the battery voltage changes with time, the instrument must be periodically recalibrated in water at 40–45°C. Second, since most thermistor types have a wide tolerance in resistance, the instrument must be completely recalibrated if the thermistor requires replacement. The first defect may be overcome by providing an additional potentiometer, which can be substituted for the thermistor by suitable switching, and which is set to have the same resistance as the thermistor would have at 45°C. By periodically substituting this potentiometer, and then resetting potentiometer 2 for a 45°C reading on the meter, battery changes are completely compensated. The second defect is overcome by using thermistors of close tolerance, so that they are interchangeable. Such thermistors may be obtained from the Yellow Springs Instrument Corp., of Yellow Springs, Ohio, USA. This company makes up thermistors for rectal, oesophageal, skin and other temperature measurements, and also a range of complete instruments for thermistor thermometry.

FURTHER READING

HUGHES, *Electrical Technology* (4th ed.) Longmans, London, 1969.
TERMAN and PETTIT, *Electronic Measurements*, McGraw-Hill, New York, 1952.
COLE, A thermistor thermometer bridge, *Rev. Sci. Instr.* **28**, 326 (1957).

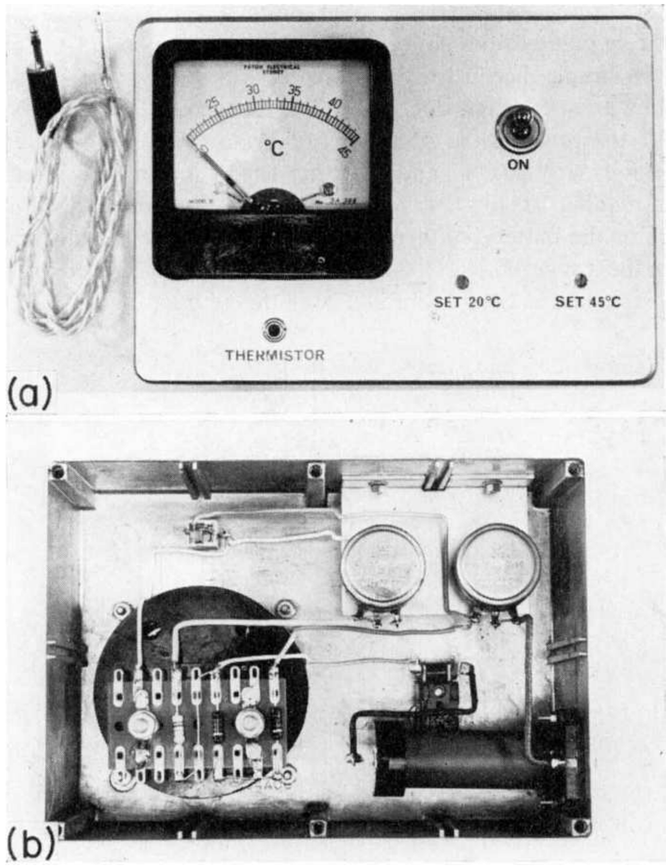


FIG. 4.7. Thermistor thermometer.

PRACTICAL

4.1 If a recording potentiometer with a range of 0–20 or 0–25 mV is available, set up a copper–constantan thermocouple, using wires of about 22 gauge SWG (20 gauge B & S). Immerse one junction in a mixture of ice and water; immerse the other in boiling water, and note the deflection produced. Now use the hot junction to estimate the temperature of molten solder, and the temperature produced at distances of 5 cm, 2 cm, 1 cm and 0·5 cm along a copper wire as the result of a normal soldering operation. What is a satisfactory length of pigtail to leave on a delicate component when soldering it into a circuit? (If available, an electronic millivoltmeter, or a special low-resistance moving coil meter intended for thermocouple thermometry may be used instead of the recording potentiometer. Such a moving coil meter would typically be 0–1 mA, and have a resistance of about 20 ohms instead of the usual 100 ohms.)
4.2 Construct and calibrate the simple thermistor thermometer of Fig. 4.5 and Fig. 4.6. Test at several intermediate temperatures, and by measuring body temperature.

CHAPTER 5

ALTERNATING CURRENTS

5.1 Nature of an alternating e.m.f.

If a conductor is caused to move across a magnetic field, a small e.m.f. appears across its ends; this e.m.f. is proportional to the length of the conductor, the strength of the magnetic field, and the rate at which the field is traversed; its polarity depends on the direction of movement. To utilise this principle in a practical machine to serve as a source of e.m.f., it is necessary to keep the conductor in motion in relation to the field, and the simplest way of achieving this is to attach it along the surface of an insulating cylinder, which is rotated on its axis in a magnetic field, as shown in Fig. 5.1. Starting with the conductor at the top of the drum, the angle M shown in Fig. 5.1 will be 0°. As the drum is rotating at a uniform rate, the conductor will at this stage be moving along the field, not cutting it at all, and the e.m.f. generated will be zero. (Fig. 5.2, point A.) As

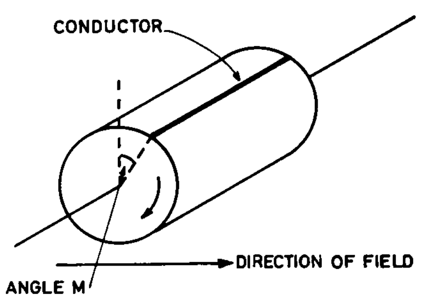


Fig. 5.1. Conductor rotated in magnetic field.

the drum continues to rotate, the conductor commences to cut across the field, at first obliquely, and then more and more sharply, until at $M = 90°$, it is cutting at right angles. The e.m.f. generated thus rises, reaching a maximum at $M = 90°$, at point B. The conductor now commences to cut the field obliquely again, and the induced e.m.f. falls, until at 180° (point C) it is again moving along the field, and no e.m.f. is produced. Past 180°, it commences to move across the field, upward instead of downward. The picture from A to C is reproduced, but with the opposite polarity. Over one revolution the e.m.f. will follow a *sinusoidal* pattern, which may be expressed by the equation

$$e = E_m \sin M \tag{5.1}$$

where e is the e.m.f. at any instant (the *instantaneous* e.m.f.) and E_m is the maximum e.m.f. generated (the *peak* e.m.f.). Each rotation produces one complete cycle of *alter-*

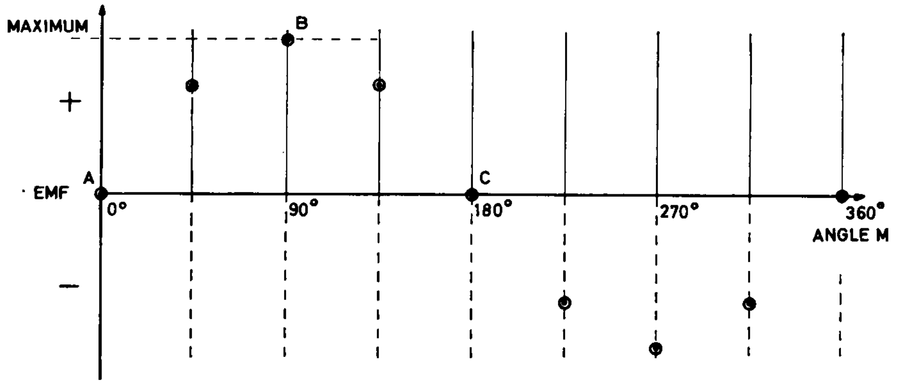


FIG. 5.2. E.m.f. as angle M increases.

nating e.m.f. If the cylinder is rotating at f revolutions per second, the e.m.f. is said to have a *frequency* of f cycles per second, or f hertz (Hz). We speak of any point in a cycle as a *phase*; for example, $e = E_m$ at the 90° phase. Two other methods of expressing phase are in common use. The angle M is often expressed in *radians* rather than degrees; there are 2π radians in 360°. Thus we could say that $e = E_m$ at the $\pi/2$ phase. Secondly, if we know the time of one cycle (the *period*), the phase can be expressed in time rather than angle. For a 50 Hz frequency, $e = E_m$ at the 5 msec phase.

The e.m.f. from a single conductor is too small to be of practical use; if however two conductors are placed diametrically opposite each other on the drum, the e.m.f.s induced in them will always be equal and opposite. If these two conductors are joined across the rear of the drum, their e.m.f.s will be put in series, as shown in Fig. 5.3. This arrangement may be extended by the use of many conductors in series to give a usable e.m.f.; such simple generators appear in old-style manual telephones and bicycle lighting sets.

To make this e.m.f. available externally, some arrangement of sliding contacts must be used. A very simple example is seen in the telephone or bicycle generator, where one contact is made through the bearings of the rotating drum, (the *rotor*), and the other through an insulated spring contact. In larger machines a pair of *slip rings* is used, as shown in Fig. 5.4.

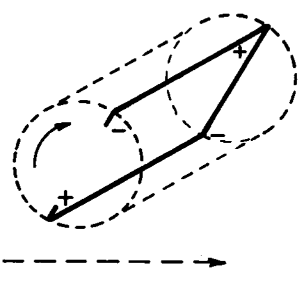


FIG. 5.3. Two rotating conductors in series.

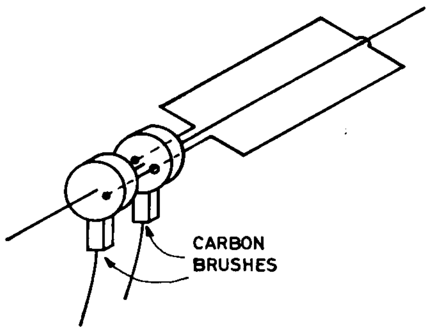


FIG. 5.4. Rotor and slip rings.

If a resistive load is connected to the output of such an *alternator*, a current will flow through the circuit. At any instant this current will be directly proportional to the e.m.f. at that instant (Ohm's law is valid at any instant). The current through the load will also be sinusoidal in form; it will be an *alternating current*, as in Fig. 5.5.

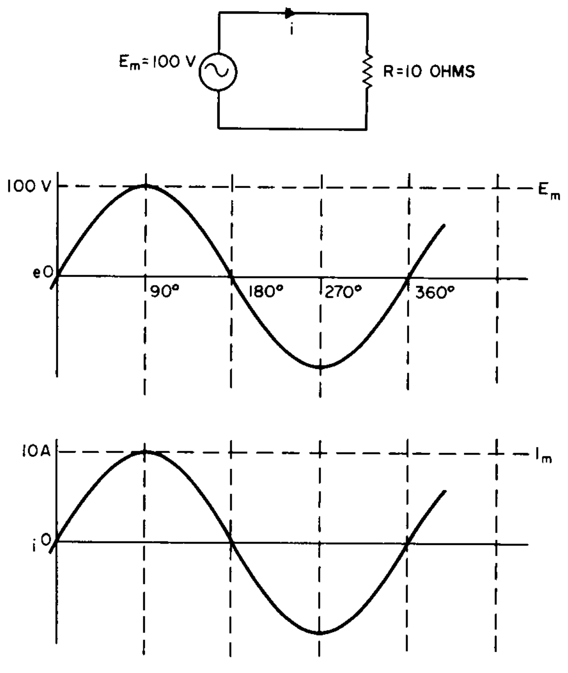


FIG. 5.5. Alternating current in resistive load.

5.2 POWER IN A RESISTOR CARRYING A.C.

Since power dissipated in a resistor is given at any instant by the product of voltage and current, it is possible to plot out the instantaneous power dissipated when the resistor is carrying an alternating current. Tabulating power for the example of Fig. 5.5,

TABLE 5.1. INSTANTANEOUS POWER CALCULATION

Phase	0°	45°	90°	135°	180°	225°	270°	315°	360°
e	0	70·7	100	70·7	0	−70·7	−100	−70·7	0
i	0	7·07	10	7·07	0	−7·07	−10	−7·07	0
w	0	500	1000	500	0	500	1000	500	0

we obtain Table 5.1. If the instantaneous power is graphed, Fig. 5.6 results. The power curve is a sine wave, always positive, and twice the frequency of the voltage or current in the load. Since the curve is symmetrical, it is clear that the *average power* delivered to the load is half the peak power W_m:

$$W_{av} = \tfrac{1}{2}W_m \tag{5.2}$$

The average power is clearly the figure of commercial importance, and the value governing the rise in temperature of the load.

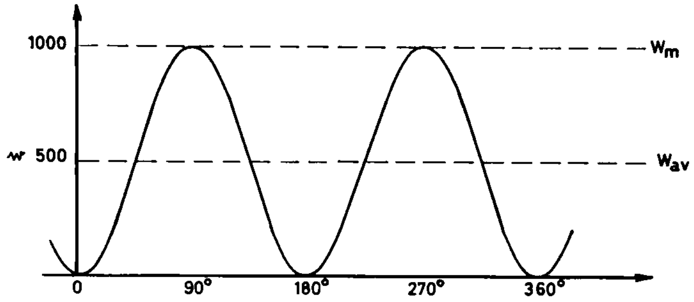


FIG. 5.6. Instantaneous power curve.

5.3 EFFECTIVE VOLTAGE AND CURRENT

To allow the direct calculation of average power, it is usual to specify an alternating voltage or current not in terms of its peak value, but in terms of its *effective* value: this is defined so that effective voltage multiplied by effective current gives average power. The required effective value E turns out to be $1/\sqrt{2}\,E_m$ or $1/\sqrt{2}\,I_m$:

$$\begin{aligned} 1/\sqrt{2}\,E_m \times 1/\sqrt{2}\,I_m &= \tfrac{1}{2}\,E_m\,I_m \\ &= \tfrac{1}{2}\,W_m \\ &= W_{av} \end{aligned} \tag{5.3}$$

The term "effective value" is used in almost every language but English; in English the term *Root Mean Square value* (r.m.s.) is generally employed, from the way in which the effective value is derived. $1/\sqrt{2}$ is approximately 0·707; from Table 5.1 and Fig. 5.6 it will be seen that instantaneous voltage and current pass through their r.m.s. values at the 45° phase.

$$E = 0{\cdot}707\, E_m \simeq \tfrac{2}{3} E_m \tag{5.4}$$

$$E_m = 1{\cdot}414\, E \simeq 1\tfrac{1}{2}\, E \tag{5.5}$$

If an alternating supply line is stated to have a voltage of 240, this is the r.m.s. value; the corresponding peak value is 340 V. All wiring must of course be insulated adequately to withstand the *peak* voltage.

5.4 MEASUREMENT OF ALTERNATING VOLTAGES

Alternating voltages are in modern practice measured by the use of moving-coil meters, in conjunction with one or more *semiconductor diodes*. This device, which will be discussed fully in Chapter 8, consists of a junction between two dissimilar materials, selected to allow a current to pass freely in one direction, but not at all in the other. Such a diode is also known as a *rectifier*; it converts alternating into uni-directional current.

The simplest circuit arrangement is the *half-wave series* rectifier. It is shown in Fig. 5.7 as an alternating voltmeter.

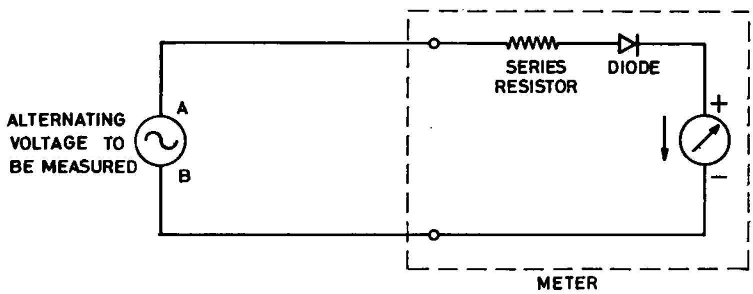


FIG. 5.7. Half-wave series rectifier.

When the e.m.f. from the source is at a phase such that A is positive to B, the diode conducts, behaving like a short circuit, and current flows through the meter in the direction of the arrow. When B is positive to A, the diode does not conduct, behaving like an open circuit, and no current flows through the meter. For example, if the source to be measured has a peak voltage of 100, and a 1 mA meter and a 100 K series resistor are used, the situation will be as shown in Fig. 5.8. The average value of an alternating current is zero, and unless the frequency is extremely low, a d.c. meter simply fails to read it at all, due to the inertia of its moving parts. The average value of a half-rectified wave, however, is clearly a positive value, and this average is what a d.c. meter will read. Calculation shows it to be $1/\pi$, or 0·318, of its peak value; the meter of Fig. 5.7 will read 0·318 mA.

A more commonly used and efficient meter circuit is the full-wave bridge rectifier, as shown in Fig. 5.9. When A is positive to B, current flows through the series resistor, the rectifier P, the meter, the rectifier S, and back to B. When B is positive to A, current flows through the rectifier R, the meter (in the *same direction* as before), the rectifier Q, the series resistor, and back to A. The meter will thus receive two pulses of current in

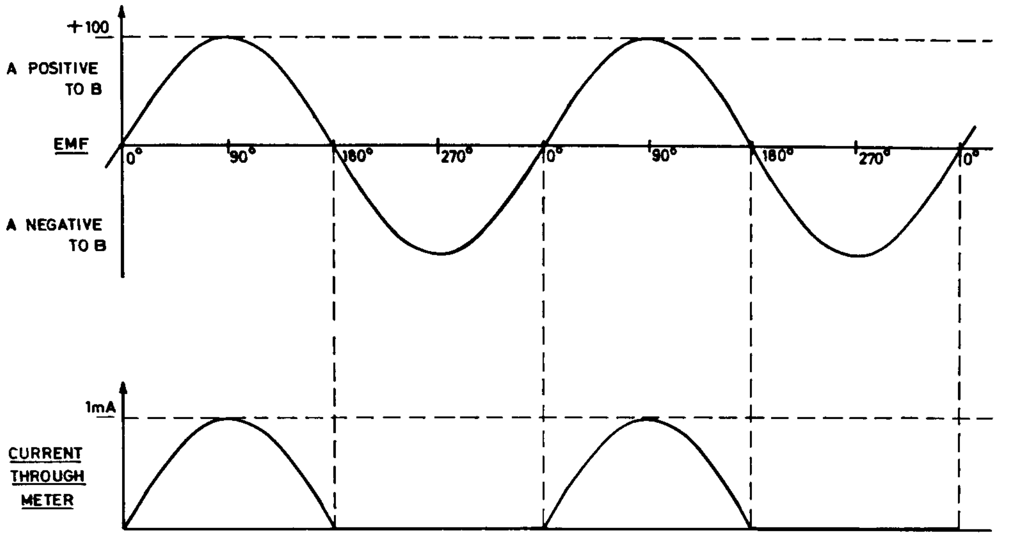


FIG. 5.8. Half-wave rectified current.

each cycle, giving an average current of $2/\pi$, or 0·637, of the peak current. The situation for a 1 mA meter and a 100 K series resistor connected to a source of 100 V peak is shown in Fig. 5.10. In this case the meter will read 0·637 mA.

Since the r.m.s. value of the voltage (0·707) is most commonly required, it is usual to calibrate the meter directly in r.m.s. volts, and to adjust the series resistor to obtain the correct calibration. In the example given the series resistor should be 100 K $\times$ 0·637/0·707, or 90·1 K.

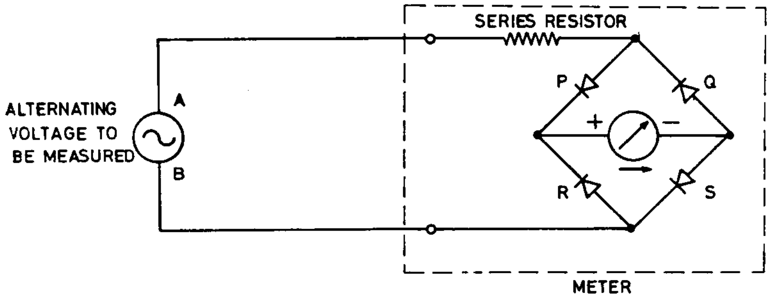


FIG. 5.9. Full-wave bridge rectifier.

FURTHER READING

HUGHES, *Electrical Technology* (4th ed.) Longmans, London, 1969.

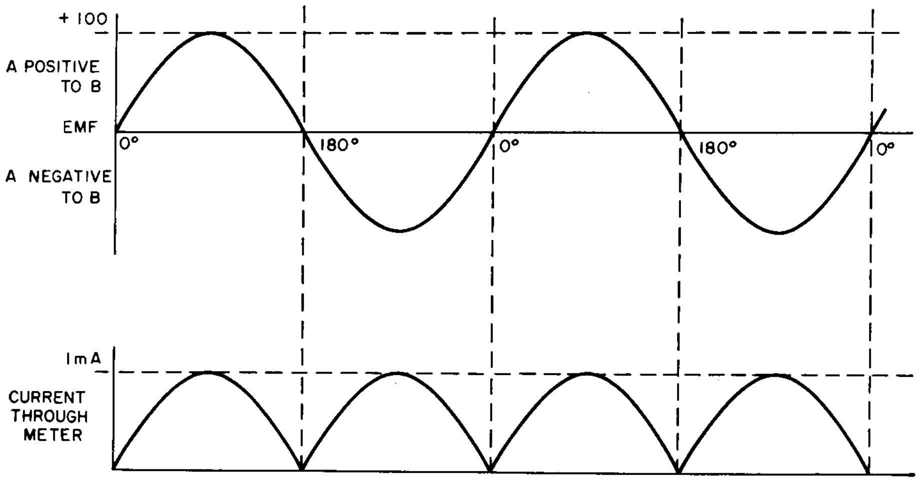


FIG. 5.10. Full wave rectified current.

PRACTICAL

Generators (often called magnetos) from old-style or military telephone equipment are still readily available, and form an excellent source of low-frequency a.c. for experimental purposes. It is best to remove the gear wheels and to attach the driving handle direct to the rotor. An alternative source of a.c., of very similar construction, is the generator from a bicycle lighting set.

5.1 Examine the construction of the generator to be used, and note the features described in this chapter. Using a 0–50 μA meter in series with a 1 M resistor, examine the output of the generator as the rotor is turned, slowly at first, and then faster and faster.

5.2 Connect a diode in series with the circuit of experiment 5.1, to form a half-wave series rectifier. Notice the reading now as the rotor is turned, slowly at first, and then faster and faster. Note the reading obtained when the generator is turned as fast as possible.

5.3 Set up a 50 μA meter for full-wave bridge rectification, again using a 1 M series resistor. Observe the number of output pulses per revolution when the generator rotor is turned slowly. Turn the rotor as fast as possible, note the reading obtained, and compare it with that obtained in experiment 5.2.

5.4 Set up a properly enclosed and safe source of about 6 V r.m.s. at supply-line frequency. Using a 50 μA meter, half-wave rectifier, and 200 K series resistor (two 100 K resistors in series), observe the meter reading when the 6 V supply is connected. From this reading calculate the peak voltage, and hence the r.m.s. voltage. Compare this calculated r.m.s. voltage with that measured directly on a commercial a.c. voltmeter.

Note that your measured values are somewhat smaller than those from a good commercial meter. This is due to the fact that real diodes require a fraction of a volt in the forward direction before they conduct at all. This is discussed further in Chapter 8.

5.5 Repeat experiment 5.4, using a bridge rectifier.

CHAPTER 6

CAPACITANCE

6.1 Constant voltage and constant current sources

Any real source of electrical energy, such as a battery, is an imperfect device. When current is drawn from a battery, the voltage measured at its terminals falls; we may represent such a battery as a *perfect constant voltage source with a resistance in series with it*, as shown for example in Fig. 6.1. In Fig. 6.1, a load current of 1 A causes the terminal

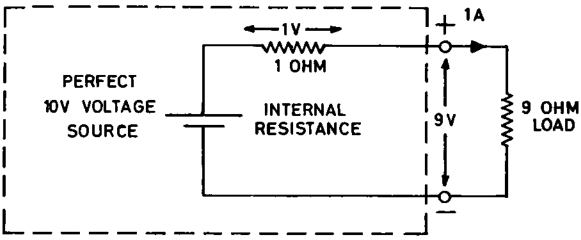


Fig. 6.1. Constant voltage equivalent of a real battery.

voltage of the battery to fall from 10 V to 9 V. If the current drawn were 0·1 A, the terminal voltage would be 9·9 V; provided that the current drawn is always less than 0·1 A, we may regard the real battery as a good approximation to a perfect constant voltage source, as shown in Table 6.1. A less obvious but equally valid representation of the battery of Fig. 6.1 would be in terms of a *perfect constant current source* of 10 A, with the same internal resistance of 1 ohm in parallel with it, as shown in Fig. 6.2. In Fig. 6.2

Table 6.1. Behaviour of circuit of Fig. 6.1

Load current (amp.)	Terminal voltage (volt)
10	0·00
3	7·00
1	9·00
0·3	9·70
0·1	9·90
0·03	9·97
0·01	9·99
0·003	10·00
0·001	10·00

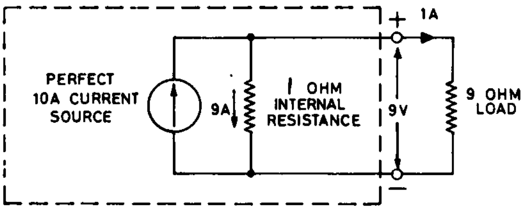


FIG. 6.2. Constant current equivalent of a real battery.

a load current of 1 A leaves 9 A flowing through the internal resistance, and gives a terminal voltage of 9 V. A little calculation will show that this circuit yields exactly the same terminal voltage values as those given in Table 6.1 for any external load current.

An *ideal constant voltage source* has zero internal resistance. Irrespective of the current drawn by a load, its voltage remains unchanged; if short-circuited, it will deliver infinite current. It is turned off by opening a switch in series with its output, thereby open-circuiting it.

An *ideal constant current source* has infinite internal resistance. Irrespective of the voltage required to maintain it, its current remains unchanged; if open circuited, it will deliver infinite voltage. It is turned off by closing a switch in parallel with its output, thereby short-circuiting it.

Practical energy sources are nearly always approximations to constant voltage devices, although sources which approximate constant current devices do occur. Either type of energy source may readily be approximated to any required degree of accuracy by suitable electronic circuitry. Such circuits are usually *protected*; a constant voltage device is caused to decrease in voltage if a current above a preset maximum is drawn, and a constant current device to decrease in current if a voltage above a preset maximum should occur across its terminals.

6.2 STORAGE OF ENERGY IN A CAPACITOR

Suppose that two metal plates are set up, parallel and in close proximity to each other, but not actually in contact. Let them be connected as in Fig. 6.3, to a constant current supply which is protected so that it can be open-circuited without damage to itself. Let a very high-resistance voltmeter be connected between the plates, and suppose that we start with the switch open and the voltmeter reading zero (region *A*, Fig. 6.4). On closing the switch, a constant current flows in the circuit, and the voltmeter reading is found to rise steadily (region *B*, Fig. 6.4). If allowed to continue, the pointer would eventually reach and exceed its full scale value. If the switch is opened before this occurs, the reading will at once stop increasing and will retain its last value indefinitely (region *C*, Fig. 6.4). If finally the constant current supply is reversed and the switch again closed, the voltmeter reading will fall steadily (region *D*, Fig. 6.4). If allowed to continue, the pointer will pass zero, and commence to show a reading in the reverse direction.

This experiment implies that the arrangement formed by the parallel plates and the space between them is capable of storing energy; such an arrangement is known as a *capacitor*. If instead of withdrawing the stored energy slowly, as in Fig. 6.4*D*, the constant

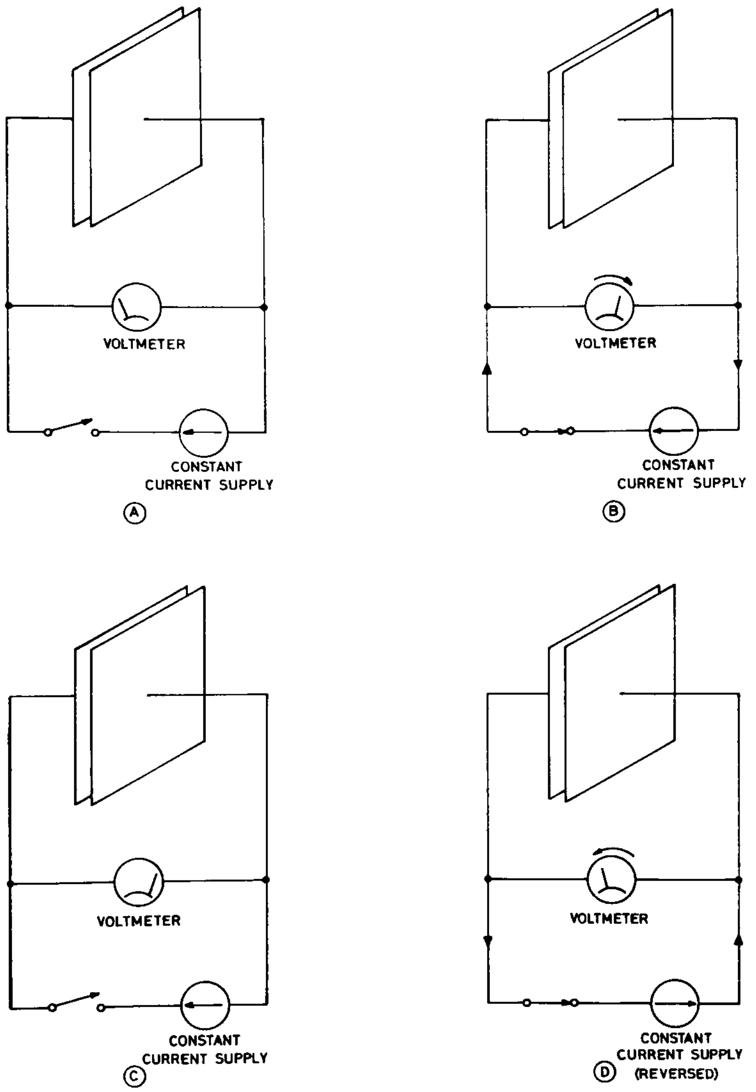


FIG. 6.3. Charge and discharge of capacitor.

current supply is replaced by a short circuit, the voltmeter reading will fall at once to zero, and a spark may be observed at the point where the circuit is closed; the stored energy is abruptly dissipated in the form of heat.

6.3 CAPACITANCE

If the magnitude of the applied constant current is varied the rate of rise of voltage will be found to be always directly proportional to it:

$$i = C\dot{e}, \tag{6.1}$$

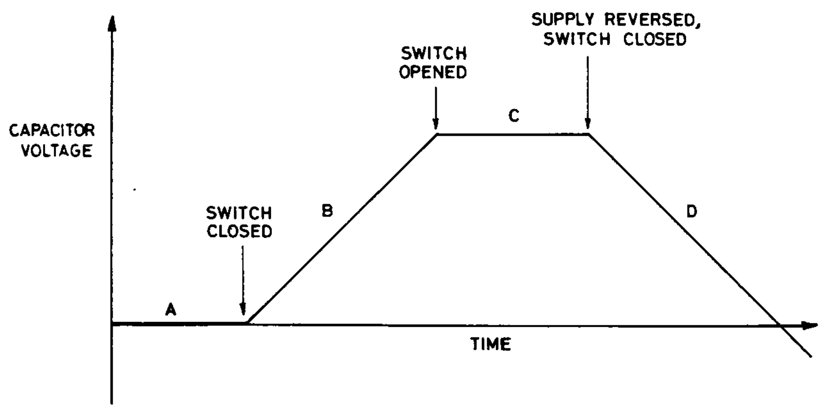


FIG. 6.4. Voltage on capacitor.

where i is the current in amperes, $\dot{e}$ the rate of rise of voltage in volts per second, and the constant C is the *capacitance* of the capacitor. If a current of 1 ampere causes a rate of rise of 1 volt per second in a particular capacitor, that capacitor is said to have a capacitance of 1 *farad*. The farad is in practice a very large amount of capacitance; the usual units employed are the microfarad (μF, 10^{-6} farad) and the picofarad (pF, 10^{-12} farad). The nanofarad (nF, 10^{-9} farad) is also found.

Practical measurements show that capacitance is directly proportional to the area of the plates of a capacitor, and inversely to their distance apart; the actual capacitance in a vacuum is found to be

$$C = \frac{8{\cdot}86 \times 10^{-12} \times a}{d}, \tag{6.2}$$

where a is the area of each plate in square metres, d is their distance apart in metres, and C is the capacitance in farad. Equation 6.2 is for practical purposes more usefully expressed as

$$c = \frac{0{\cdot}0886a}{d}, \tag{6.3}$$

where a is the area of each plate in square cm, d is their distance apart in cm, and c is the capacitance in pF.

The figure $8{\cdot}86 \times 10^{-12}$ is a universal constant, known as the *permittivity of free space*, and is usually written ϵ_0 (epsilon nought). It is measured in farad per metre. In charging the capacitor we must imagine that energy is being stored in the space between the plates (the *dielectric*) as a form of *electric strain*. The nature of this strain is not easily imagined, but it is certain that it exists. ϵ_0 is a measure of the ease with which space is strained by a potential difference between the two plates. The substitution of air for vacuum makes only a slight difference to the measured capacitance of a capacitor. The strained region of space is said to contain an *electric field*.

6.4 Dielectric constant and dielectric strength

The capacitance of a capacitor can be considerably increased by the insertion of a sheet of insulating material between its plates. The *dielectric constant* or *relative permittivity* of a substance is the factor by which the capacitance is increased when that substance is substituted for a vacuum between the plates. Typical insulating materials used have dielectric constants in the range from 4 to 8, although it is possible to produce synthetic ceramic materials with a value of 100 or more.

A second property which must be considered in selecting a dielectric material is *dielectric strength*. This is usually expressed as the number of volts required to rupture a sheet of the material one millimetre thick. Since a reduction in thickness of the dielectric will give high capacitance in a small physical size, a high dielectric strength is clearly desirable. In modern practice, thin films of polystyrene or polycarbonate are used, the plates consisting of metallic layers deposited on each side of the dielectric film.

6.5 Stored energy

The energy stored in a charged capacitor is given by

$$J = \tfrac{1}{2}Ce^2, \tag{6.4}$$

where J is the energy in joules, C the capacitance in farad, and e the potential difference between the plates in volts. As a rough working rule, a stored energy of 1 J or more is considered potentially lethal. In circumstances where this amount of energy is stored (e.g., 200 V or more on 50 μF, 1000 V or more on 2 μF) great care must be taken to *discharge* capacitors with an insulated tool *after* switching off the equipment and *before* they are handled.

6.6 Capacitors as used in electronic equipment

An enormous variety of capacitors is manufactured, the physical form depending on the purpose for which the capacitor is required. A capacitor is specified in terms of its capacitance, its tolerance, the peak voltage it must withstand, and the type of dielectric required. Preferred values, as for resistors, are commonly used. The type of dielectric is governed by the permissible energy loss in the dielectric as the capacitor is taken through a cycle of charge and discharge; some slight loss always occurs. For details the manufacturers' literature should be consulted, but Table 6.2 is typical of modern practice. Only the highest quality of modern capacitor should be used in medical electronic equipment. Electrolytic capacitors are made by the deposition from a suitable solution of a very thin insulating dielectric on a metal plate. The earliest capacitors of this type consisted of two sheets of aluminium hung in a solution of ammonium borate. When an e.m.f. was applied between them a film of aluminium borate was formed on the positive plate, and the device then constituted a capacitor; the solution formed one plate, the film the dielectric, and the positive plate the second plate of the capacitor. Capacitors in the range 50 μF to about 1 F are still made by a modification of this principle; the solution is held in a sheet of absorbent paper in contact with the aluminium plate, and

TABLE 6.2. SELECTION OF CAPACITOR TYPES

Capacitance	Nature of dielectric: Peak volts up to 50	Peak volts up to 500	Peak volts above 500
1 pF			
10			mica
100	ceramic	ceramic	
0·001 μF			
0·01			
0·1	polyester, polycarbonate	polyester, polycarbonate	oil or oil-impregnated paper
1			
10	tantalum electrolytic		
100		electrolytic	
1000	electrolytic		
10,000			
100,000			

the whole capacitor is hermetically sealed to prevent drying. A variant of this principle is to use tantalum instead of aluminium; by partially fusing together fine beads of tantalum a very large surface area may be obtained in a small volume. Tantalum capacitors are very common in the range from about 2 to 50 μF, especially for low working voltages. A selection of electrolytic capacitors is shown in Fig. 6.5. Electrolytic capacitors cannot readily be made in small tolerances, and are unsuitable for applications where precise values of capacitance are required.

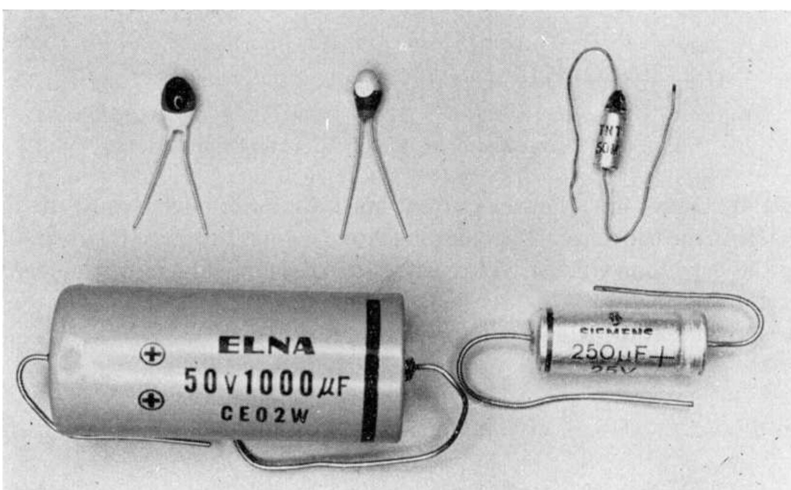


FIG. 6.5. Electrolytic capacitors.

It should be noted that electrolytic capacitors are *polarised*; unless specially designed they cannot be used with alternating voltages, and in a direct voltage circuit must be arranged so that the end marked as positive connects to the positive side of the supply. They are almost instantly destroyed if this is overlooked.

Variable capacitors in the range 1–1000 pF are also available in a wide variety of forms, with either ceramic or air dielectric. Above this range, variability can be obtained only by the selection of a range of fixed capacitors with a switch.

6.7 Charge and discharge of capacitor–resistor combination

In the circuit of Fig. 6.6, suppose that the capacitor is initially discharged, with $e_c = 0$. On pressing button A, the capacitor commences to charge through the resistor. At any instant the sum of the voltage across the capacitor and the voltage across the resistor must equal E:

$$E = e_c + e_r. \tag{6.5}$$

Since the capacitor voltage initially is zero, the resistor voltage at this instant must equal E. Thus the initial current will be, by Ohm's law, 100 μA. From eq. (6.1) ($i = C\dot{e}$), e_c must commence to rise at 10 V/sec. However as e_c rises, e_r must diminish correspondingly.

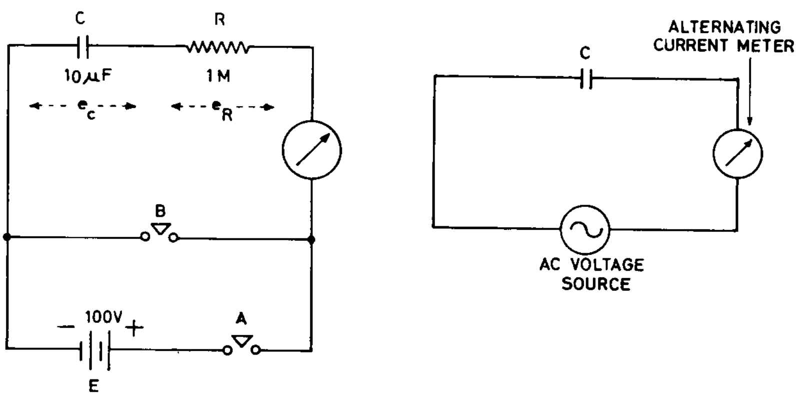


FIG. 6.6. Charge and discharge of RC series combination.

This implies, by Ohm's law, that the current must diminish, and so must the rate of rise of voltage across the capacitor. The charging curve is found to have the form of Fig. 6.7. Eventually the capacitor voltage will reach E, the resistor voltage will be zero, and no further current will flow.

It can be shown that the equation to an RC charging curve is

$$e_c = E(1-\epsilon^{-t/RC}), \tag{6.6}$$

where ϵ has the value 2·718. . . ϵ is a mathematical constant known as the *base of natural logarithms*.

Values of $\epsilon^{-t/RC}$ and $(1-\epsilon^{-t/RC})$ are frequently required in calculating charging curves. Table 6.3 gives some representative points. Once the values of R and C (megohm

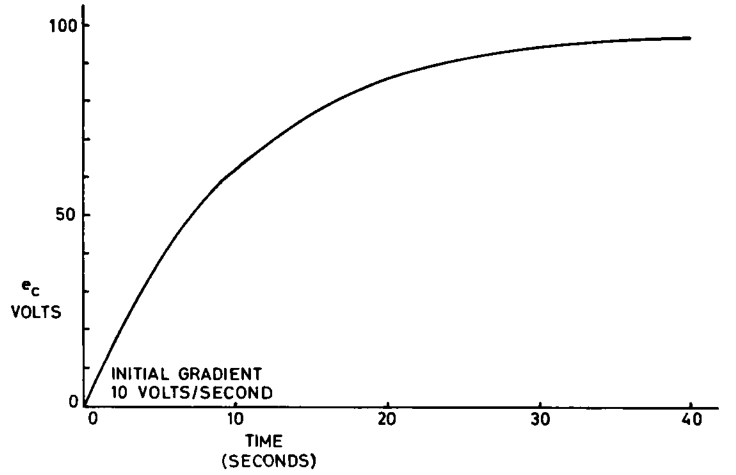


FIG. 6.7. Charging curve for RC series combination.

and μF) are determined in a particular circuit, the value of t/RC is calculated for a number of points in time after charging has started. Table 6.3 and eqn. (6.6) are then used to find the value for e_c at each of these points.

The term RC is described as the *time constant* of the circuit; it alone governs the rate at which the capacitor voltage approaches its final value. When $t = RC$ (after one time constant) the voltage will always have covered 63% (approximately 2/3) of its total rise; when $t = 5RC$ (after five time constants) the voltage is within 1% of its final value, and for practical purposes the capacitor is considered to be fully charged.

TABLE 6.3. VALUES REQUIRED FOR EQN. (6.6)

t/RC	$\epsilon^{-t/RC}$	$1-\epsilon^{-t/RC}$
0	1·00	0
0·2	0·82	0·18
0·4	0·67	0·33
0·6	0·55	0·45
0·8	0·45	0·55
1·0	0·37	0·63
2·0	0·13	0·87
3·0	0·05	0·95
4·0	0·02	0·98

If, having charged the capacitor fully, button A is released and button B pressed, current will flow out of the capacitor through the resistor. The flow will be rapid at first, and then slower and slower as the capacitor empties. The discharge curve will be as shown in Fig. 6.8 for the circuit of Fig. 6.6. The equation to a discharge curve is

$$e_c = E\epsilon^{-t/RC}, \tag{6.7}$$

so that again 63% of the total voltage change will occur in one time constant, and 99% in five.

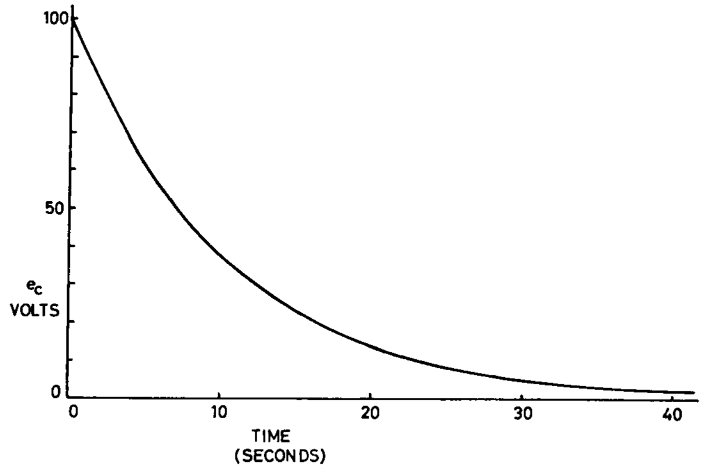


FIG. 6.8. Discharge curve for an RC series combination.

The curves of Figs. 6.7 and 6.8 are said to be *exponential* in form. An exponential curve is as definite a shape as is a circle; it is completely determined if we know any three points on it, or two points and its final value (its *asymptote*).

The importance of this circuit is that it is the basis of most practical time-delay and timing circuits. It will be noticed that the initial portion of both charging and discharging curves is very nearly a straight line; this fact is used when it is desired to approximate a voltage rising or falling linearly with time. If the voltage continued to change at its initial rate, complete charging or discharging would take exactly one time constant.

6.8 Capacitor leakage testers

Capacitors suspected of having developed an internal leakage between their plates may be tested by a modification of the ohmmeter circuit; a voltage equal to the rated maximum working voltage is applied, and the leakage current observed. A series resistor must be included to protect the meter in the event of complete failure of the capacitor. Capacitors other than electrolytic types should show *no* measurable current on a 50 μA meter under these conditions. Aluminium electrolytic types may show several mA of leakage when first connected, but this should diminish to a value of less than about 0·2 CV μA after 1 or 2 minutes (C μF, V test voltage). Miniature tantalum types should show less than about 0·02 CV μA at any time.

6.9 Alternating voltage applied to a capacitor

If a capacitor is connected to a source of alternating voltage, as in Fig. 6.9, at any instant that capacitor voltage must be the same as that of the source. As the source voltage rises, current must flow in the circuit as the capacitor charges. When the source voltage is at its peak, the capacitor is fully charged, so no current is flowing (Fig. 6.10, phase A). As the voltage falls to zero, (phase B) and reverses, current flows out of the

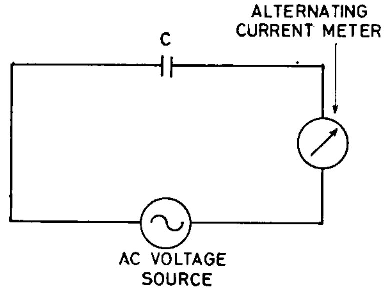


FIG. 6.9. Alternating voltage applied to a capacitor.

capacitor, and is actually flowing at its maximum rate as the voltage passes through zero. The capacitor then charges up in the reverse direction, the current again dying away to zero as the voltage reaches its maximum reverse value (phase *C*). This process is repeated for each reversal of the applied voltage, so that the meter in Fig. 6.9 registers the flow of an alternating current through it; it will show a steady reading, although no electrons ever cross between the plates of the capacitor. It will be seen from Fig. 6.10 that the peaks of current occur 90° earlier than the peaks of voltage; we say that the current leads the voltage, and is *in quadrature* with it.

The magnitude of the current which flows is proportional to the magnitude of the applied voltage, although the peak values of voltage and current do not occur simultaneously. If we ignore this *phase difference* in the occurrence of the peaks, and consider the

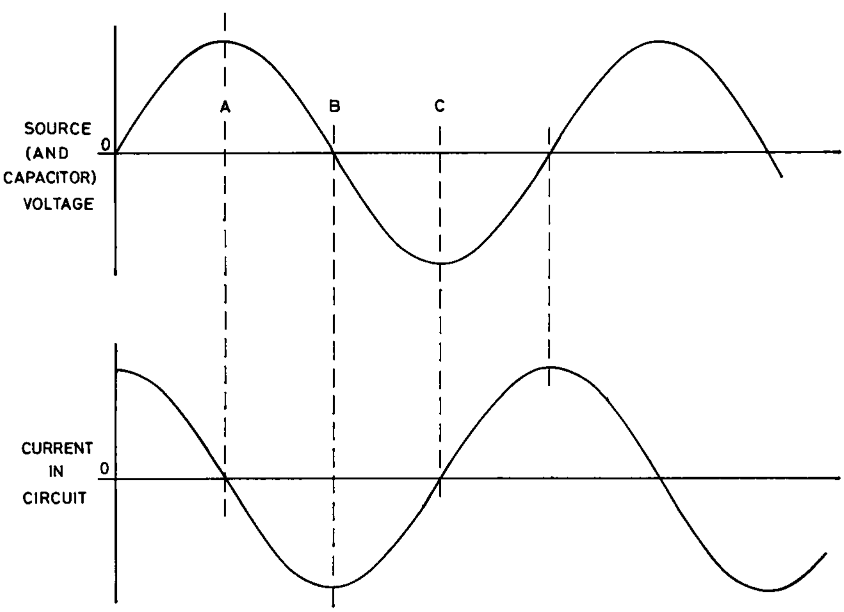


FIG. 6.10. Alternating voltage applied to a capacitor.

magnitudes only, we may write, by analogy with Ohm's law

$$i = \frac{e}{Z}, \tag{6.8}$$

where i is the current flowing, e is the voltage of the source, and Z is the *impedance* of the capacitor to current flow. i and e may both be peak values, or, more usually, both r.m.s. values. Z is measured in ohms, as is resistance, but it should be noted at once that ohms of capacitive impedance cannot be combined additively with ohms of resistance.

To make use of eqn. (6.8), it is necessary to know the value of Z in any particular case; it may be shown that

$$Z = \frac{1}{2\pi fC}, \tag{6.9}$$

where Z is in ohms, f in hertz, and C in farad. $2\pi f$ is known as the *angular frequency*. If the frequency f is in hertz (cycles per second) then $2\pi f$ is in radians per second, since there are 2π radians in a circle (or cycle). It is usual to write

$$\omega = 2\pi f, \tag{6.10}$$

so eqn. (6.9) may be rewritten as

$$Z = \frac{1}{\omega C}. \tag{6.11}$$

Notice that the impedance is inversely proportional to the capacitance; the larger the capacitance, the lower the impedance, and the greater the current that will flow for a given applied voltage. The impedance is also inversely proportional to the frequency; it is meaningless to state an impedance unless the relevant frequency is stated or implied.

A useful practical value to memorise is that a capacitor of 1 μf has an impedance of 1591 ohms (say 1600 ohm) at 100 Hz. The impedance of any other capacitor at any other frequency can then be obtained by inverse proportion.

It is interesting to notice the instantaneous power in the circuit of Fig. 6.9; this can be obtained by multiplying instantaneous voltage by instantaneous current in Fig. 6.10. In the first quarter cycle, power is positive; energy is being accepted by the capacitor. In the second quarter cycle, power is negative; energy is being returned to the circuit from the capacitor. In the third quarter cycle energy is again being accepted; in the fourth it is being returned, and so on. *No energy is ever lost from the circuit, and no heat is produced.*

6.10 Alternating voltage applied to a capacitor and resistor in series

Eq. (6.8) may be regarded as a more general form of Ohm's law, which is applicable both to a.c. and to d.c. circuits; it may be applied to any composite circuit in which the value of Z in ohms can be determined. (Where resistance alone exists in an a.c. circuit, $Z = R$, and eqn. (6.8) reduces to Ohm's law.) In the circuit of Fig. 6.11, Z must be determined by suitably combining the individual impedances of C and R. As explained

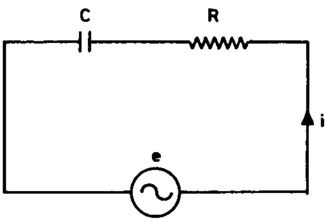


FIG. 6.11. Capacitor and resistor in series.

earlier these are *not additive*; it may be shown that

$$Z = \sqrt{(Z_c^2 + R^2)}, \tag{6.12}$$

where $Z_c = 1/2\pi fC$.

To determine the *phase* of the current flowing in terms of the source voltage, we can argue that if R alone were present, the current would be *in phase* with the voltage; if C alone were present, the current would be *in quadrature* with the voltage with its peaks occurring 90° earlier than those of the voltage. For the combination of R and C the phase should be somewhere between these two limits. In fact, the angle ϕ between voltage and current can be shown to be

$$\phi = \text{artan}\,\frac{Z_c}{R}, \tag{6.13}$$

with the peaks of current always earlier than the peaks of voltage. For example, consider the circuit of Fig. 6.11 when $e = 240$ V r.m.s., $f = 50$ Hz, $C = 10\ \mu$F, and $R = 300$ ohm. Then

$$\begin{aligned} Z_c &= \frac{10^6}{2\pi \times 50 \times 10} \\ &= 318 \text{ ohms,} \\ Z &= \sqrt{(318^2 + 300^2)} \\ &= 437 \text{ ohms,} \\ i &= \frac{240}{437} \\ &= 0{\cdot}548 \text{ amp. r.m.s.,} \\ \phi &= \text{artan}\,\frac{318}{300} \\ &= 46{\cdot}7^\circ\text{; current leading voltage.} \end{aligned}$$

6.11 MEASUREMENT OF CAPACITANCE

Capacitance is most commonly measured by the use of an *a.c. bridge circuit*; a typical arrangement is shown in Fig. 6.12. Balance will be obtained when

$$\frac{C_1}{C_2} = \frac{R_2}{R_1} \tag{6.14}$$

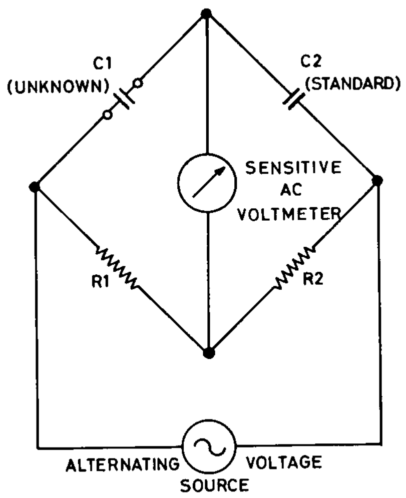


FIG. 6.12. Capacitance bridge.

It is usual to employ a cathode ray oscilloscope as the sensitive a.c. voltmeter in the circuit of Fig. 6.12.

Bridge methods cannot easily be used to measure the capacitance of electrolytic capacitors; these are tested by observing their time constants of discharge, in conjunction with known resistors.

FURTHER READING

HUGHES, *Electrical Technology* (4th ed.) Longmans, London, 1969.
Manufacturers' literature on capacitors.

PRACTICAL

6.1 Set up the charging circuit of Fig. 6.13, using for C a combination of polyester or oil-impregnated capacitors of at least 10 μF. When ready, close the switch and plot the charging current as a function of time. (The experiment may be repeated as often as required by switching off, discharging the capacitor by a short circuit across it, and starting again.) From your curve, and using Ohm's law, deduce and plot the curve of capacitor voltage as a function of time. Read from this curve the time constant of the combination; does it agree with the calculated value? Is the capacitor potentially lethal when fully charged? Comment on the validity of the "1 joule rule" given in § 6.5. Repeat the experiment with a different value of C. What would be the effect of varying R instead?

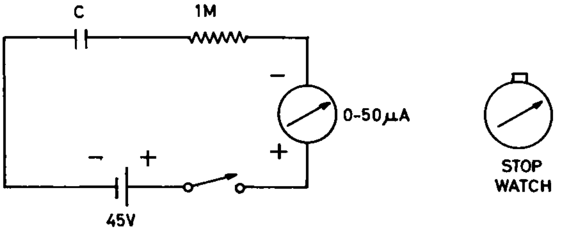


FIG. 6.13. Charging curve experiment.

6.2 Set up the circuit of Fig. 6.14, using the same value of *A* as in experiment 6.1. With the switch open, charge *C* to 45 V from your battery (observe the polarity shown in Fig. 6.14), and remove the battery. When ready, close the switch, and plot the discharge current as a function of time. From this curve, deduce and plot the curve of capacitor voltage as a function of time. Now replot the last graph on two decade semi-logarithmic graph paper, using the logarithmic scale for voltage and the linear scale for time. Draw the straight line of best fit through the points. From either of these last two graphs, deduce the time constant of the circuit.

6.3 Place a 0–50 μA d.c. meter in series with a 1 M resistor and connect them across the output terminals of a telephone or bicycle generator. Turn the rotor slowly *and smoothly*, and observe the positions of the handle at which the peaks of current occur. Now substitute a 0·047 μF capacitor for the 1 M resistor, again turn the rotor slowly *and smoothly*, and observe that an a.c. flows through the capacitor. At what positions of the handle do the peaks occur now?

6.4 Set up a capacitance bridge, as shown in Fig. 6.15; the d.c. milliammeter will indicate any a.c. caused by the bridge being out of balance. The ratio potentiometer should be provided with a pointer knob and blank card scale. Calibrate the bridge by the use of a number of known capacitors placed in the *Cx* position.

Now connect up several parallel capacitor combinations and measure their capacitances. How may capacitance of a parallel bank of capacitors be calculated? Connect two 0·22 μF capacitors in series and measure the capacitance of the combination. How can the observed result be obtained by calculation? Test your rule by measuring other series combinations.

6.5 If a cathode ray oscilloscope with a very slow time base and a long persistent screen is available, observe directly the charge and discharge curves of experiments 6.1 and 6.2, and read off the time constant. (Note: a resistor of at least 100 M must be placed in series with the oscilloscope input, to avoid excessive discharge of the capacitor through the oscilloscope itself.)

6.6 Examine any commercial capacitance bridges and capacitor leakage testers which may be available.

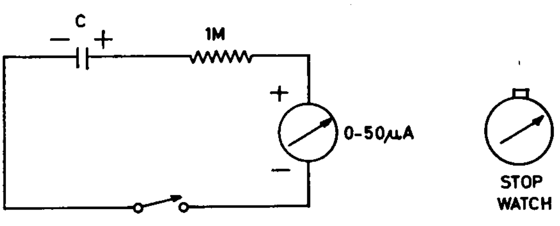


FIG. 6.14. Discharge curve experiment.

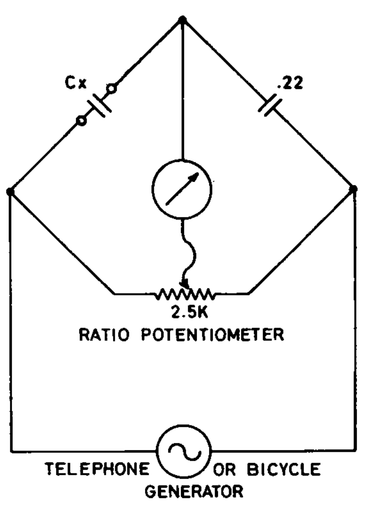


FIG. 6.15. Capacitance bridge experiment.

CHAPTER 7

INDUCTANCE

7.1 Storage of energy in an inductor

Suppose that a hollow ring is wound uniformly with a layer of insulated wire, the two ends being brought out to allow a current to be passed. We speak of such a device as a toroidal coil, or simply as a toroid. Let the ends of the coil be connected as in Fig. 7.1, to a constant voltage supply which is protected so that it can be short-circuited without

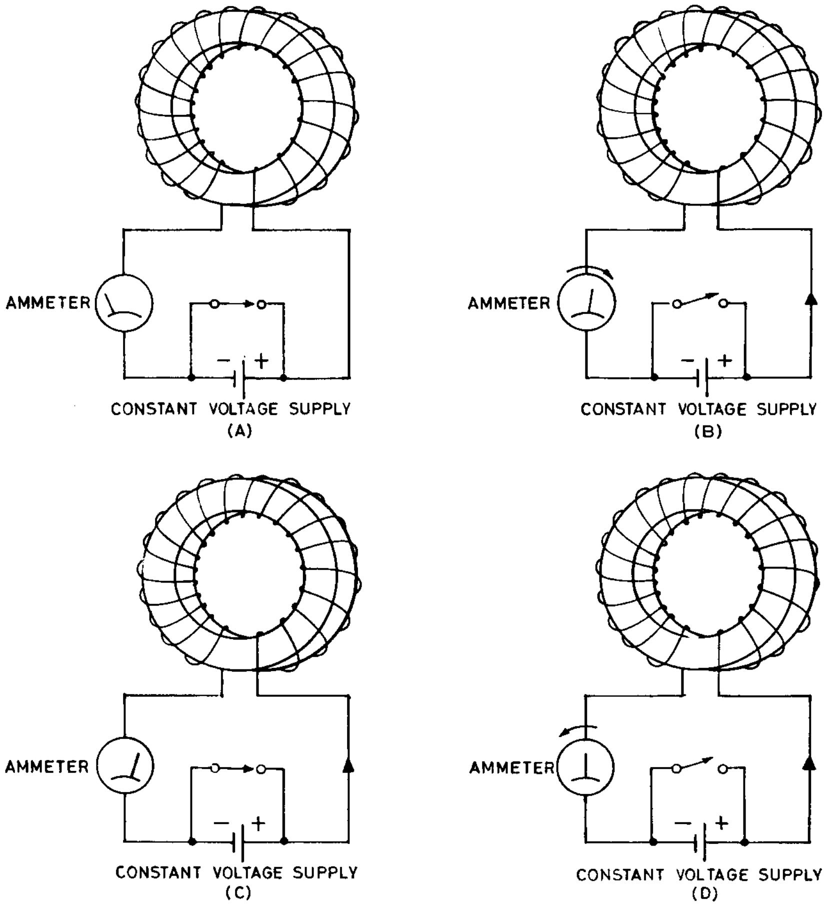


Fig. 7.1. Charge and discharge of inductor.

damage to itself. Let a very low-resistance ammeter be connected in series with the circuit, and suppose that we start with the switch closed and the ammeter reading zero (region A, Fig. 7.2). On opening the switch, a constant voltage is applied to the circuit and the ammeter reading is found to rise steadily (region B, Fig. 7.2). If allowed to continue, it would eventually reach and exceed its full scale reading. If the switch is closed before this occurs, the ammeter reading will at once cease to rise, and will retain its last reading indefinitely (region C, Fig. 7.2). (It should be noted that this experiment is in fact impossible for any real coil at room temperature; the energy losses in its resistance are always excessive. The experiment can however be done at temperatures approaching absolute zero, when many metals become *super-conductors*.) If finally the constant voltage supply is reversed and the switch again opened, the ammeter reading will fall steadily (region D, Fig. 7.2). If allowed to continue, it will pass zero, and commence to show a reading in the reverse direction.

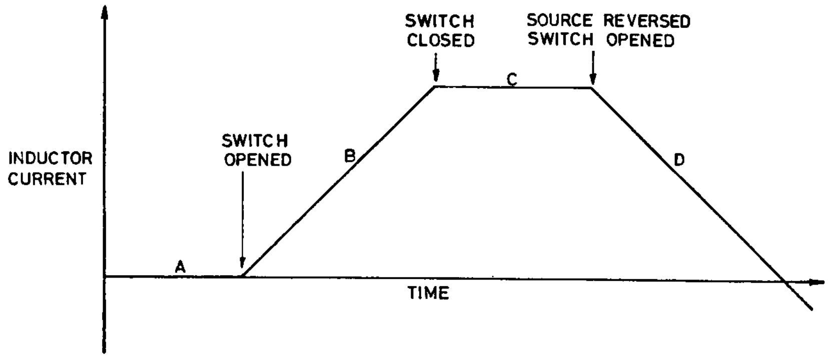


FIG. 7.2. Current in inductor.

This experiment implies that the arrangement formed by the coil and the toroidal space enclosed by it is capable of storing energy; such an arrangement is known as an *inductor*. If instead of withdrawing the stored energy slowly, as in Fig. 7.2D, the constant voltage supply is replaced by an open circuit, the ammeter reading will fall at once to zero, and a spark may be observed at the point where the circuit is opened; the stored energy is abruptly dissipated in the form of heat.

7.2 INDUCTANCE

If the magnitude of the applied constant voltage is varied, the rate of rise of current will be found to be always directly proportional to it:

$$e = L\dot{i}, \tag{7.1}$$

where e is the applied e.m.f. in volts, $\dot{i}$ is the rate of rise of current in amperes per second, and the constant L is the *inductance* of the inductor. If an e.m.f. of 1 volt causes a rate of rise of 1 ampere per second in a particular inductor, that inductor is said to have an inductance of 1 *henry*. The henry is in practice quite a large amount of inductance;

smaller units commonly employed are the millihenry (mH, 10^{-3} henry) and the microhenry (μH, 10^{-6} henry).

Practical measurements show that the inductance of a toroid is directly proportional to its area of cross-section, and to the square of the number of turns of wire used; it is inversely proportional to the mean circumference of the toroid. The actual inductance in a vacuum is found to be

$$L = \frac{12{\cdot}57 \times 10^{-7} \times a \times N^2}{l}, \tag{7.2}$$

where a is the area of cross-section of the toroid in square metres, N is the number of turns, l is the mean circumference in metres, and L is the inductance in henry.

The figure $12{\cdot}57 \times 10^{-7}$ (actually $4\pi \times 10^{-7}$) is a universal constant, known as the *permeability of free space*, and is usually written μ_0 (mu nought). In building up current in an inductor, we must imagine that energy is being stored in the space inside the coil (the *core*) as a form of *magnetic strain.* The nature of this strain, like that of electric strain, is not easily imagined, but it is certain that it exists. μ_0 is a measure of the ease with which space is strained by a current in a coil. The substitution of air or non-magnetic solids for vacuum makes only a very slight difference to the measured inductance of an inductor. The strained region of space is said to contain a *magnetic field.*

7.3 Relative permeability and saturation

The inductance of an inductor can be considerably increased by manufacturing its core of a magnetic material, such as iron or an iron alloy. The *relative permeability* (usually referred to simply as the *permeability*) of a magnetic material is represented by the symbol μ; it is the factor by which the inductance is increased when that material is substituted for a vacuum in the core. Typical magnetic materials have permeabilities in the range from 500 to 5000, although it is possible to produce materials with permeabilities up to 100,000.

A second property which must be considered in selecting a magnetic material is the ease with which it *saturates.* If a constant voltage is applied to a coil wound on a core of any magnetic material, the current rises relatively linearly at first, but presently it will be found that it starts to rise faster and faster; this implies that the apparent inductance, and hence the permeability, is falling. The core is said to be approaching saturation; if a magnetic material is used as a core, the current in the coil about it must be limited to a point below that where saturation occurs.

7.4 Stored energy

The energy stored in an inductor carrying a current is given by

$$J = \tfrac{1}{2}Li^2, \tag{7.3}$$

where J is the energy in joules, L the inductance in henry, and i the current in amperes through the inductor. A large inductor carrying a current can store quite enough energy to be lethal.

7.5 Inductors as used in electronic equipment

Although large iron-cored inductors were at one time used extensively in electronic equipment, their weight, bulk, and expense have caused them to be discarded for most purposes. Their principal use is now in high-frequency equipment, such as radio transmitters and receivers, and even in the latter they are now used less and less. Cores in modern inductors are often composed of powdered iron pressed into a suitable shape, such as a toroid. Many inductors are constructed in the form of a cylindrical coil rather than a toroid; in this case the *magnetic circuit* lies partly inside and partly outside the coil. A selection of modern inductors is shown in Fig. 7.3.

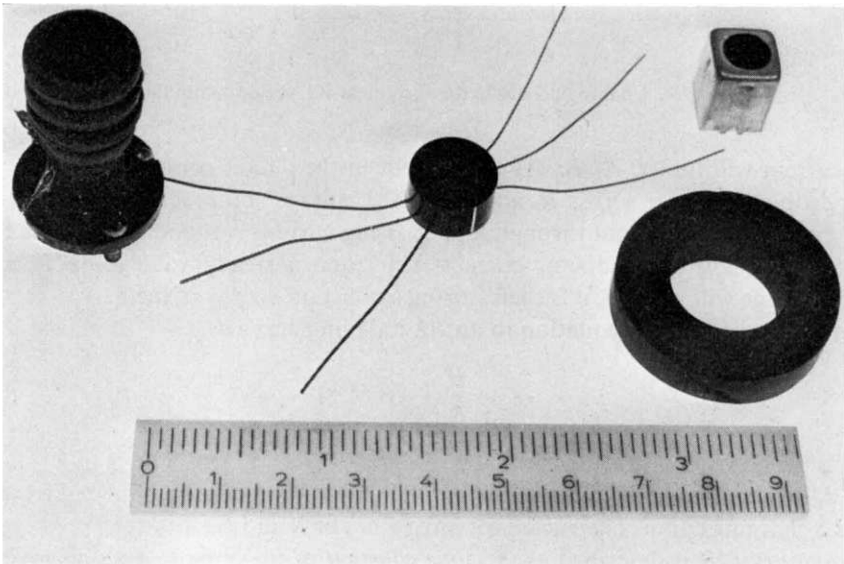


FIG. 7.3. Typical modern inductors.

7.6 Charge and discharge of inductor–resistor combination

A real inductor may be represented as a perfect inductor without resistance, in series with a resistor equal to the measured resistance of the real inductor. Whereas it is relatively easy to obtain an almost perfect capacitor, all real inductors have quite significant amounts of resistance, and it is important to understand the properties of RL combinations for this reason, as well as in the cases where the resistor is deliberately placed in series with an inductor. In the circuit of Fig. 7.4, E is a *protected* 100 V battery. Suppose that initially there is no current in the inductor, and button A is pressed. At any instant the sum of the voltage across the inductor and the voltage across the resistor must equal E

$$E = e_L + e_R. \tag{7.4}$$

Since the inductor current initially is zero, the resistor current at this instant must also be zero, and hence e_R must be zero. So e_L must equal E at this instant, and the initial rate of

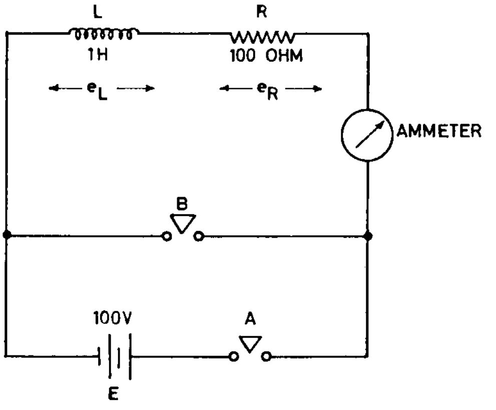


FIG. 7.4. Charge and discharge of typical RL series combination.

rise of current will be 100 A/sec. As the current in the circuit commences to increase, a voltage drop will appear across R, and hence the voltage across L will diminish, and so will the rate of rise of current through it. The rising current is found to have the form of Fig. 7.5. Eventually the inductor current will reach a steady value of E/R, and the inductor voltage will be zero; it is then storing a constant energy of 0·5 J.

It can be shown that the equation to an RL charging curve is

$$i_L = \frac{E}{R}(1-\epsilon^{-tR/L}), \tag{7.5}$$

where ϵ is the base of natural logarithms.

If tR/L is calculated, the values of $\epsilon^{-tR/L}$ and $(1-\epsilon^{-tR/L})$ may be obtained from Table 6.3, and a charging curve constructed for any values of R and L.

The quantity L/R is described as the *time constant* of the circuit; it alone governs the rate at which the inductor current approaches its final value. When $t = L/R$ (after one

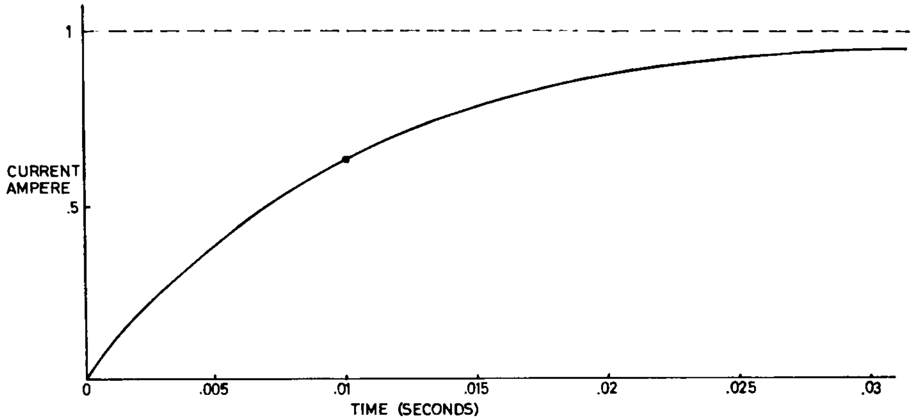


FIG. 7.5. Charging curve for typical RL series combination.

time constant) the current will always have covered 63% (approximately $\frac{2}{3}$) of its total rise; when $t = 5\,L/R$ (after five time constants) the current is within 1% of its final value, and for practical purposes is considered to have reached a steady state.

It is quite difficult to manufacture an inductor with a resistance low enough to give a time constant of 1 second; a typical 10 H inductor would have a resistance of about 100 ohms. For this reason inductors are never used to store energy for any length of time, while capacitors are quite suitable.

If, having allowed the current to reach its maximum value, button *B* in Fig. 7.4 is pressed while *A* is still pressed (remember that the source of e.m.f. must be protected against short circuits), the current in the inductor will decay exponentially, with the same time constant as previously.

$$i_L = \frac{E}{R}\epsilon^{-tR/L}. \tag{7.6}$$

The graph of current against time will be similar to that of Fig. 6.8.

7.7 Alternating current passed through an inductor

If an ideal inductor is connected to an ideal source of alternating current, as in Fig. 7.6, a voltage must appear across the inductor which at any instant will be proportional to the rate of change of current:

$$e = L\dot{\imath}. \tag{7.1}$$

When the current is at its maximum, it is neither rising nor falling, and the voltage is zero (Fig. 7.7, phase *A*). As the current falls to zero (phase *B*) and reverses, it is changing negatively at its maximum rate. The inductor voltage will then have its maximum negative value; this will die away to zero as the current reaches its maximum reverse value (phase *C*). This process repeats for each reversal of the current, so that the meter of Fig. 7.6 registers an alternating voltage; it will show a steady reading proportional to the applied current. The peak values of current and voltage do not occur simultaneously; if we ignore the 90° phase difference in the occurrence of the peaks, and consider the magnitudes only, we may write

$$e = iZ, \tag{7.7}$$

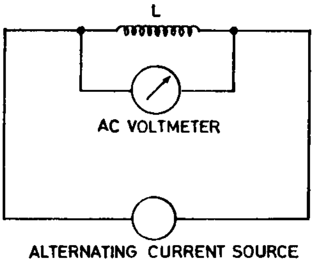


FIG. 7.6. Alternating current passed through an inductor.

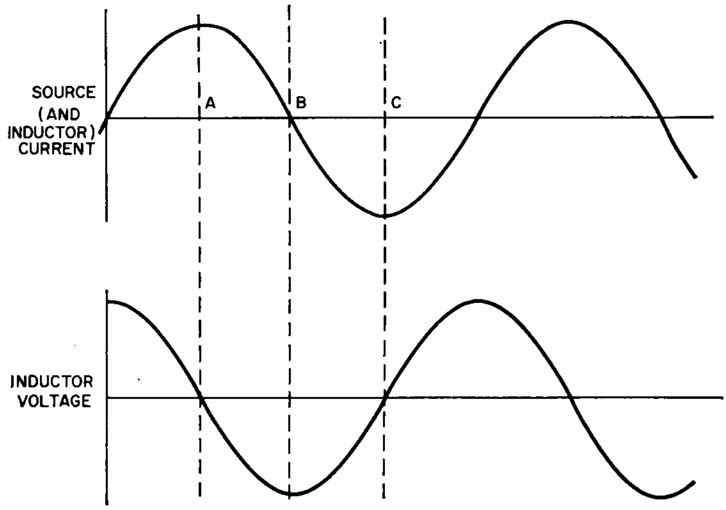


FIG. 7.7. Alternating current passed through an inductor.

where e is the voltage across the inductor, i is the alternating current passed by the source, and Z is the *impedance* of the inductor. e and i may both be peak values, or, more usually, both r.m.s. values. A is measured in ohms; but ohms of inductive impedance cannot be combined additively with either ohms of resistance or ohms of capacitive impedance.

It should be noticed that eqn. (7.7) is identical in form with eqn. (6.8) for the capacitive case, and with Ohm's law for resistance.

To make use of eqn. (7.7), it is necessary to know the value of Z in any particular case; it may be shown that

$$Z = 2\pi fL, \tag{7.8}$$

where Z is in ohms, f is in hertz, and L is in henry.

Notice that the impedance of an inductor is directly proportional to its inductance. The impedance is also directly proportional to the frequency, and it is meaningless to state an impedance unless the relevant frequency is stated or implied.

It is interesting to notice the instantaneous power in the circuit of Fig. 7.6; this can be obtained by multiplying instantaneous voltage by instantaneous current in Fig. 7.7. In the first quarter cycle, power is positive, energy is being accepted by the inductor. In the second quarter cycle, power is negative; energy is being returned to the circuit from the inductor. In the third quarter cycle, energy is again being accepted, in the fourth it is being returned, and so on. *No energy is ever lost from the circuit, and no heat is produced.*

7.8 ALTERNATING CURRENT PASSED THROUGH AN INDUCTOR AND RESISTOR IN SERIES

As in the case of a capacitor and resistor in series, the relationship between current and voltage in the circuit of Fig. 7.8 may be obtained by the use of eqn. (7.7), the more

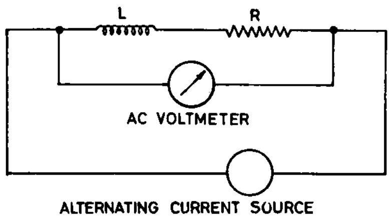


FIG. 7.8. Inductor and resistor in series.

general form of Ohm's law

$$e = iZ. \tag{7.7}$$

To obtain the total impedance in the circuit, it may be shown that, exactly as for the capacitor and resistor in series,

$$Z = \sqrt{(Z_L^2 + R^2)}, \tag{7.9}$$

where $Z_L = 2\pi fL$.

To determine the *phase* of the voltage across the combination, in terms of the current passed through it, we can argue that if R alone were present, the voltage would be *in phase* with the current; if L alone were present, the voltage would be *in quadrature* with the current, with its peaks occurring 90° earlier than those of the current. For the combination of R and L the phase should be somewhere between these two limits. In fact, the angle ϕ between voltage and current can be shown to be

$$\phi = \operatorname{artan} \frac{Z_L}{R} \tag{7.10}$$

with the peaks of voltage always earlier than those of current. For example, in Fig. 7.8, consider an inductor of inductance 10 H and resistance 1000 ohms, with a voltage of 240 V r.m.s. and frequency 50 Hz across it. What current will flow?

$$\begin{aligned} Z_L &= 2\pi \times 50 \times 10 \\ &= 3142 \text{ ohm}, \\ Z &= \sqrt{(3142^2 + 1000^2)} \\ &= 3297 \text{ ohm}, \\ i &= \frac{240}{3297} \\ &= 0{\cdot}0727 \text{ ampere} \\ \phi &= \operatorname{artan} \frac{3142}{1000} \\ &= 72{\cdot}3°, \text{ voltage leading current.} \end{aligned}$$

7.9 Measurement of inductance

Inductance is measured by the use of an *a.c. bridge circuit*. Since practical inductors have a significant resistance as well as their inductance, an inductance bridge must have two balance controls, one for resistance and one for inductance, as shown in Fig. 7.9. Balance will be obtained only when

$$\frac{L_1}{L_2} = \frac{R_3}{R_4} = \frac{R_1}{R_2}, \tag{7.11}$$

and this must be reached by alternately adjusting L_2 and R_4. Since suitable variable inductors are not easily manufactured, other bridge circuits are most commonly used for inductance measurements, but Fig. 7.9 illustrates the basic principle.

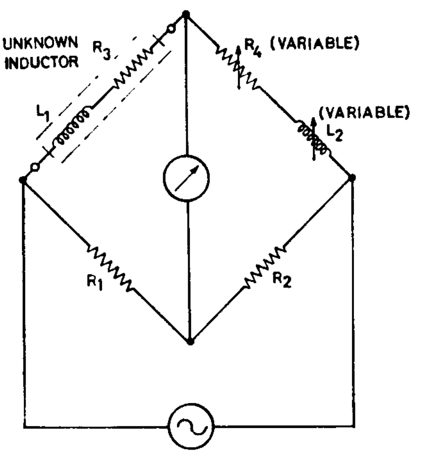


FIG. 7.9. Inductance bridge.

7.10 Mutual inductance: the transformer

If a toroid is wound with two identical coils, and a varying current is passed through one of them, we have already seen that a varying voltage will appear across it, as given by eqn. (7.1):

$$e = L\dot{i}. \tag{7.1}$$

This voltage may be pictured as being due to the varying magnetic field in the toroid. Since however this field is common to *both* coils on the toroid, we would expect to find an identical varying voltage across the second coil, and so indeed we do. We say that there is *mutual inductance* between the two coils. We describe such an arrangement of two coils as a *transformer*, and call the two coils the *primary* and *secondary* windings. If now the secondary winding is given twice the number of turns the primary has, the secondary voltage will be double that of the primary, and so on. This rule holds in particular if the voltages concerned are the peak or r.m.s. values of alternating voltages;

transformers are very commonly used to raise or lower alternating voltages, by a suitable selection of the ratio of primary to secondary turns.

If the secondary of a transformer is connected to a *load*, such as a resistor, alternating current will be caused to flow in the load and through the secondary, and power will be dissipated in the load. This secondary current will itself produce an alternating field in the core, and this field is always in such a direction as to reduce the original field from the primary; the effect is that the primary current rises. Since there is no source of power within the transformer itself, the power dissipated in the load must have been supplied to the primary from the original source of a.c. Neglecting any internal losses in the transformer, we may write

$$W_{SEC} = W_{PRI},$$

and this implies that

$$E_{SEC}\, I_{SEC} = E_{PRI}\, I_{PRI}, \tag{7.12}$$

or

$$\frac{E_{SEC}}{E_{PRI}} = \frac{I_{PRI}}{I_{SEC}}.$$

If for example the secondary voltage is twice that of the primary, the primary current is twice that of the secondary. In the circuit of Fig. 7.10, this effect is shown for a *step-down* transformer with a 10:1 ratio; Fig. 7.10 also shows the conventional fashion of drawing a transformer.

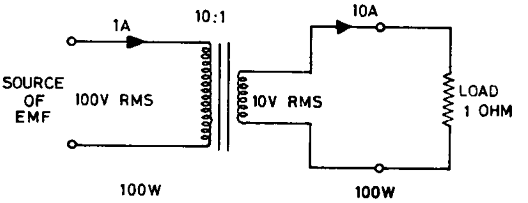


FIG. 7.10. Step-down transformer.

There is an interesting way of looking at Fig. 7.10. The actual load is 1 ohm, but if Ohm's law is applied to the primary, the load appears to be E_{PRI}/I_{PRI}, or 100 ohms. There has been a transformation of the load *impedance* by the transformer; we say that the source of EMF "sees" 100 ohms.

By a similar argument for the general case of a transformer of turns ratio N, the impedance seen will be increased by N^2. Transformers are frequently used in this fashion.

7.11 Magnetising current, transformer losses

If the primary of a transformer is connected to a source of alternating voltage, and no load current is taken from the secondary, a small primary current will flow; this is due to the inductance and resistance of the primary itself. The current is known as the *magnetising current*; in a small transformer intended for operation from the supply line, it is typically about 100 mA r.m.s.

In any real transformer there will be several causes of power loss; each of these will contribute to heating of the transformer. They are divided into *copper losses*, and *iron losses*. Copper losses are due to power dissipation in the resistances of the primary and secondary or secondaries. Iron losses are due to the internal work done on the iron core as it is repetitively magnetised and demagnetised in each cycle (*hysteresis loss*) and to the fact that the iron in the core tends to act as a single turn short-circuited secondary (*eddy current loss*); currents through the resistance of the iron dissipate power in the form of heat. Copper losses are minimised by the use of adequate diameter copper wire in primary and secondary; iron losses are minimised by the use of suitable iron alloys, and by *laminating* the core. A typical core is constructed of many thin sheets of iron, each insulated from its neighbours; for frequencies above a few hundred hertz, it is common to find powdered iron cores.

The maximum current that can be drawn from a transformer is set by its copper and iron losses, and also by the fact that the iron core will saturate if the magnetic field in it exceeds the limit set for the type of iron alloy used.

7.12 Power transformers

The most common use of transformers in modern equipment is for the conversion of supply line voltages to values more suitable to the circuitry to be used; at the same time they serve to *isolate* the circuitry from the supply line, and thus provide for much safer operation. Transformers of this type are known as *power transformers*. Stringent rules are laid down in every country to govern the design and construction of power transformers, particularly those used in medical equipment.

It is usual to include a *Faraday shield* between primary and secondary of a power transformer. This is a thin sheet of copper, insulated so that it does not form a complete one turn short-circuited secondary itself. It is separately earthed to the chassis of the equipment. It forms an additional barrier to accidental contact between primary and secondary windings, and also greatly reduces the *capacitance* between primary and secondary. At supply line frequencies, this capacitance may otherwise be great enough to pass a dangerously large alternating current through a patient. Fig. 7.11 shows typical small power transformers.

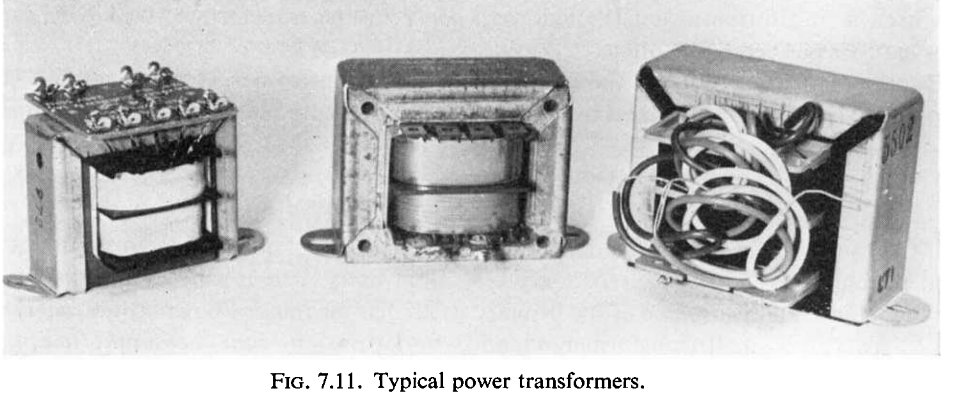


Fig. 7.11. Typical power transformers.

7.13 Power transformer failures

One of the commonest causes of power transformer failures is a breakdown in insulation within either primary or secondary, which causes one or more turns of a winding to be short-circuited. The short-circuited turns behave like a short-circuited secondary winding; an excessive current flows, and the transformer overheats, with a characteristic odour of burning varnish. The presence of short-circuited turns is also indicated by an excessive magnetising current.

The design of power transformers is governed by the frequency of the supply lines for which they are intended; in particular, a transformer designed for 60 Hz operation will almost certainly overheat on a 50 Hz supply, even although the applied voltage is correct.

7.14 Resistive, capacitive and inductive circuit elements

The three types of circuit element so far described (resistive, capacitive, and inductive) are in fact the only possible types of basic element. Of these three, a resistive element alone is capable of dissipating energy in the form of heat; capacitive or inductive elements can store energy in electric or magnetic fields, but this energy is returned to the circuit as soon as the field is allowed to collapse.

The reader will have noticed the similarity of the treatments given to capacitance and inductance in Chapters 6 and 7. These two elements are often described as duals. Table 7.1 shows two features of this dualism; these features have been described above, and there are many others. Note that each equation is converted into its dual by substituting L for C or vice versa, and i for e or vice versa.

Table 7.1. Duals

	Capacitance	Inductance
Definition	$i = C\dot{e}$	$e = L\dot{i}$
Energy stored	$J = \frac{1}{2}Ce^2$	$J = \frac{1}{2}Li^2$

7.15 Inductance, resistance and capacitance in series: resonance

In the circuit of Fig. 7.12 an alternating voltage E is shown applied to a series combination of R, L, and C. As before, to determine the alternating current I which will flow, it is first necessary to determine the total impedance, and then to apply the generalised

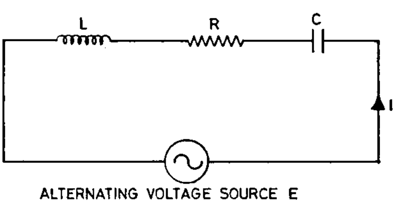


Fig. 7.12. R, L and C in series.

form of Ohm's law. The total impedance may be obtained by first determining the impedances Z_L, Z_R and Z_C separately, at the frequency of the voltage source; these, as we have seen previously, will be

$$Z_L = 2\pi fL,$$
$$Z_R = R,$$
$$Z_C = 1/2\pi fC.$$

Z_L and Z_C are combined first, by taking the *difference* between them. This difference will be an inductive or capacitive impedance depending on whether Z_L or Z_C is the greater:

$$Z_{\text{equiv}} = (Z_C - Z_L) \quad \text{or} \quad (Z_L - Z_C),$$

whichever gives a positive answer.

Z_{equiv} is now combined with R, as before:

$$Z = \sqrt{(Z_{\text{equiv}}^2 + R^2)}.$$

Having obtained Z, Ohm's law gives

$$I = E/Z,$$

and

$$\phi = \text{artan}\,(Z_{\text{equiv}}/R).$$

For example, if E is 200 V r.m.s. at 50 Hz, and $L = 1$ H, $R = 100$ ohm, $C = 20\ \mu F$, we have

$$\begin{aligned} Z_L &= 2\pi \times 50 \times 1 \\ &= 314 \text{ ohm}, \\ Z_R &= 100 \text{ ohm}, \\ Z_C &= 1/(2\pi \times 50 \times 20 \times 10^{-6}) \\ &= 159 \text{ ohm}. \end{aligned}$$

So

$$\begin{aligned} Z_{\text{equiv}} &= (314 - 159) \\ &= 155 \text{ ohm}, \end{aligned}$$

and is inductive.

Then

$$\begin{aligned} Z &= \sqrt{(155^2 + 100^2)} \\ &= 185 \text{ ohm}, \end{aligned}$$

and

$$\begin{aligned} I &= 200/185 \\ &= 1{\cdot}08 \text{ ampere}. \end{aligned}$$

The phase angle will be

$$\begin{aligned} \phi &= \text{artan}\ 155/200 \\ &= 38°, \end{aligned}$$

with the voltage leading the current, since E_{equiv} is inductive.

Suppose however that in the previous example C were reduced to $10{\cdot}1\ \mu F$, while all the other values remained the same. Now

$$Z_C = 1/(2\pi \times 50 \times 10{\cdot}1 \times 10^{-6})$$
$$= 314 \text{ ohm},$$

and $$Z_{\text{equiv}} = 0.$$

Now Z simply becomes R, and

$$I = 200/100$$
$$= 2{\cdot}0 \text{ ampere}.$$

Under the condition where $Z_C = Z_L$, these two impedances simply cancel out, and the current in the circuit is limited by R alone, and is as large as it can ever become. The current is *in phase* with the applied voltage, and the circuit is said to be *resonant*.

This phenomenon of resonance is a very important one; it is the basis of most circuits used to separate signals of different frequencies.

7.16 Radiation of energy

Although the foregoing sections have discussed inductance as a property of a coil, even a single straight wire produces a magnetic field when carrying a current, and so has a small but finite inductance. Similarly even a single straight wire produces an electric field when a voltage difference exists along it, and so has a small but finite capacitance between any two points.

If an alternating current flows in a wire, successive electric and magnetic fields appear about it in each half cycle. As the frequency is raised, an increasing proportion of the energy transferred to these fields never returns to the wire, but travels outward from the wire at the speed of light as linked magnetic and electric fields. This radiation constitutes radio waves.

FURTHER READING

Hughes, *Electrical Technology* (4th ed.) Longmans, London, 1969.
ARRL Radio Amateur's Handbook, American Radio Relay League, Massachusetts, 1975.

PRACTICAL

7.1 Measure the resistance of an iron-cored inductor, using an ohmmeter. (A suitable inductor is the primary of a small power transformer, which will typically have a resistance of about 40 ohms, and an inductance of about 20 H). Calculate the current which will flow if a d.c. voltage of 6 V is applied. Now set up the circuit of Fig. 7.13, using a fast responding milliammeter of suitable range, a *heavy duty* 6 V battery (a car battery is best), the inductor, and a push button whose contacts are clearly visible. Press the button and observe the rise of current in the inductor. The rate of rise will be noticeably sluggish, the final deflection taking about 2 seconds to be attained.

Roughly estimate the time constant of the current rise, and hence estimate the inductance of the inductor. Release the button; the meter reading will fall to zero at once, and a small spark may be observed at the contacts of the push button as the current attempts to continue to flow through the inductor.

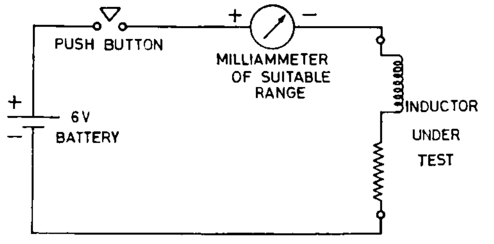


FIG. 7.13. Inductor test circuit.

Now insert a semiconductor diode in the circuit, as shown in Fig. 7.14; carefully observe the polarity of the diode. On pressing the button, the meter rises as before. The diode cannot conduct, since it is connected in such a direction as to behave as an *open circuit*. On releasing the button, however, the current attempts to continue flowing in the inductor; it now flows through the diode, and the energy stored is dissipated in the resistance of the inductor. There is no spark and the meter reading falls sluggishly, at the same rate as it rose originally. A diode is often used in this fashion, to prevent damage to switch contacts or other equipment when a circuit containing inductance is turned off.

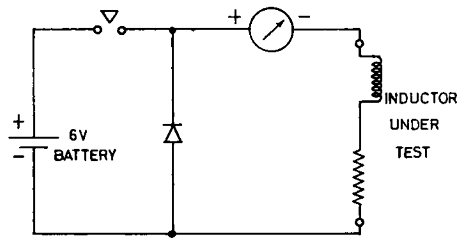


FIG. 7.14. Suppressor diode.

The method above is rough, but serves to illustrate energy storage in an inductor.

If a long-persistence or storage cathode ray oscilloscope is available, its input may be shunted with a 10 ohm resistor, and it may be used in place of the milliammeter; in this way the exponential growth and decay of the current can be clearly demonstrated.

7.2 Using the inductor of experiment 7.1, connect up the circuit of Fig. 7.15. Take great care not to touch any part of the circuit unless the *supply line is unplugged*. Record the voltage and current, and hence calculate the impedance of the inductor at the supply-line frequency. Using this value and the measured resistance, calculate the inductance. Is the resistance significant in determining the total impedance at the supply-line frequency? Calculate the phase angle between voltage and current.

7.3 Set up the circuit of Fig. 7.16, using any small transformer. Raise and lower the primary current slowly, then more and more rapidly, and observe the magnitude and direction of the induced secondary e.m.f.

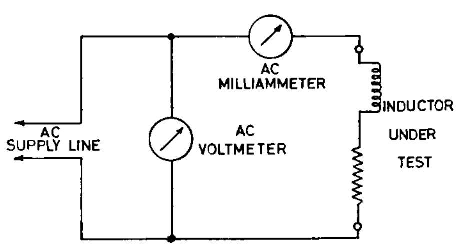


FIG. 7.15. Impedance measurement.

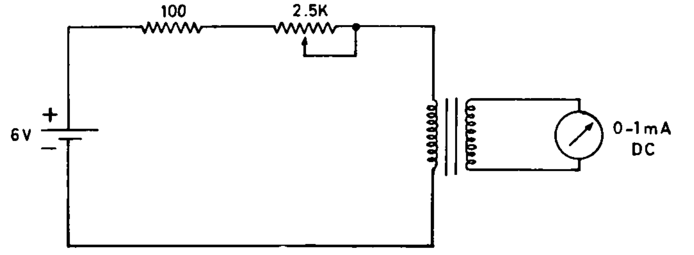


FIG. 7.16. Principle of transformer.

7.4 The properties of a power transformer may be studied by using the circuit of Fig. 7.17. (A suitable transformer is one converting the supply-line input to 110 V output.) For each load used, tabulate the four meter readings, and calculate and tabulate load resistance, power dissipated in load, and the apparent power ($E_{PRI}I_{PRI}$) taken by primary. From the primary readings on no load calculate the primary impedance, and hence, neglecting the primary resistance, calculate the primary inductance.

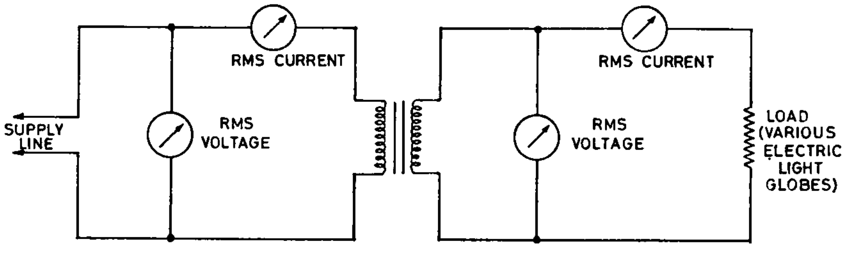


FIG. 7.17. Properties of power transformer.

7.5 If an oscilloscope is available, the use of a series resonant circuit for frequency selection may easily be demonstrated. Using an air-cored transformer made to the approximate dimensions shown in Fig. 7.18, and a suitable variable capacitor, it is possible to resonate L2 and C to any frequency within the medium-wave broadcast band. L1 effectively induces a voltage in *series* with L2 and C, and the oscilloscope is used to observe the voltage across C; this is a measure of the alternating current flowing in the circuit. Retain this tuned circuit for use in the practical work of Chapter 8.

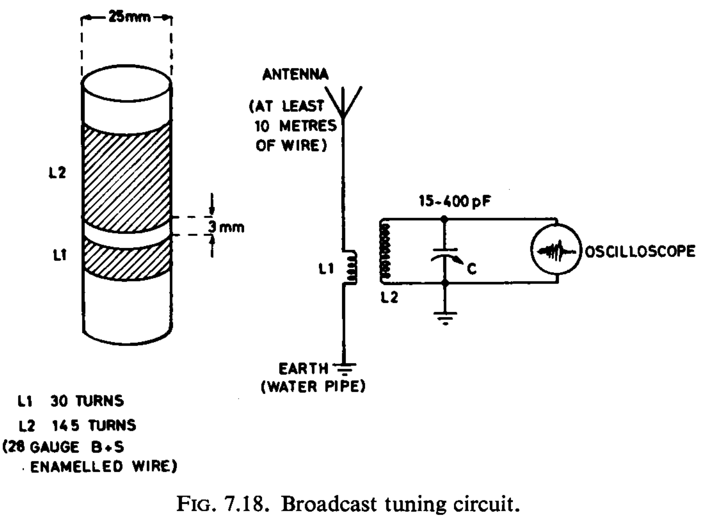


FIG. 7.18. Broadcast tuning circuit.

CHAPTER 8

DIODES AND TRANSISTORS

8.1 Semiconductors

Semiconductors are materials whose electrical properties lie between those of conductors and those of insulators; the semiconductor material most commonly used is silicon. By producing silicon in a very high state of purity, and then "doping" it with a minute trace of a suitable impurity, it can be converted into either of two forms. One is the negative or N form, in which the impurity gives rise to a supply of free electrons, like those in a metallic conductor. The other is the positive or P form, which contains a supply of sites deficient in electrons. These sites are known as *holes*, and can be caused to move through the material by the application of a potential difference. In almost all respects they behave as though they were free positive charges.

8.2 The PN junction

A junction formed between a wafer of the P form and a wafer of the N form constitutes a *semiconductor diode*. If the P layer is made positive to the N layer (*forward bias*), holes and electrons move through their respective layers as shown in Fig. 8.1. At the junction, holes and electrons meet and cancel each other. At the negative terminal, new electrons are continually injected, and at the positive terminal holes are continually generated as electrons are removed. There is thus a steady flow of current through the device.

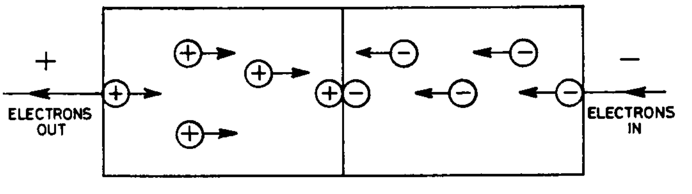


Fig. 8.1. Forward bias at PN junction.

On the other hand, if the N layer is made positive to the P layer (*reverse bias*) holes and electrons move as shown in Fig. 8.2. At the junction, nothing is available to pass a current. At the negative terminal, holes accumulate (some are cancelled by electrons from the negative lead when the potential difference is first applied to the device). At the positive terminal, electrons accumulate (some are removed by the positive lead when the potential difference is first applied).

In an ideal device no steady current could flow; in actual diodes there is always a very small reverse leakage current.

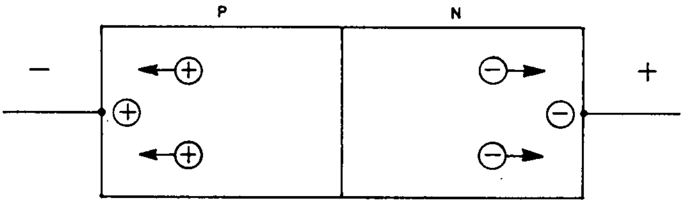


FIG. 8.2. Reverse bias at *PN* junction.

As was mentioned previously in Chapter 5, a small forward bias voltage is required to "turn on" a semiconductor diode. For a silicon diode, current will commence to flow at about 0·4 V, and will be appreciable at about 0·6 V; the latter figure is often used as an estimate of the voltage across a conducting silicon diode.

Commercial diodes are available in two broad classifications, *signal diodes*, and *power diodes*. The latter are physically somewhat larger, since they may be required to dissipate appreciable power when forward biased and passing a current. Some typical signal and power diodes are shown in Fig. 8.3. The larger sizes require the use of a heat sink.

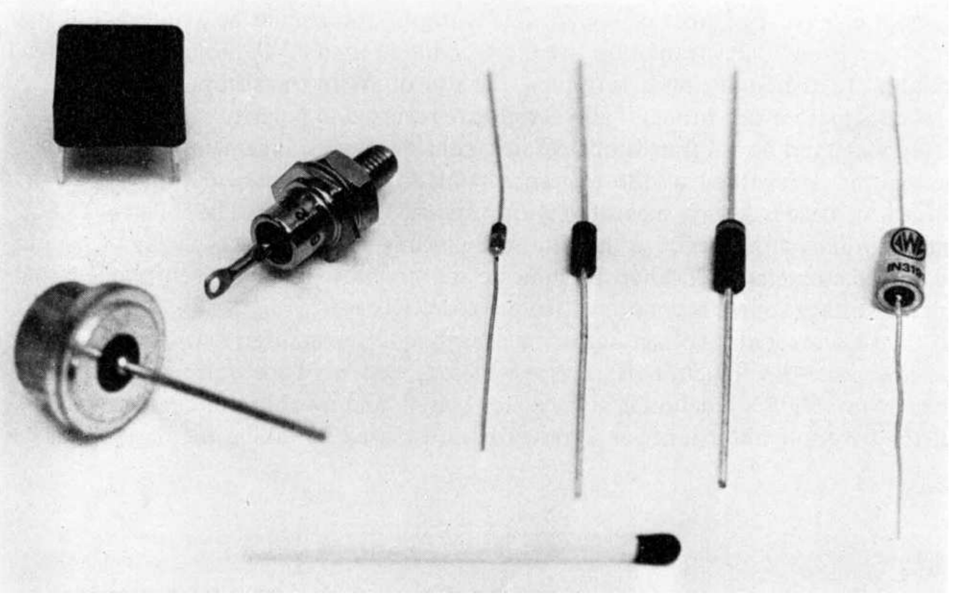


FIG. 8.3. Typical semiconductor diodes.

8.3 THE BIPOLAR TRANSISTOR

A *bipolar transistor* consists of two *PN* junctions back to back, to form either a *PNP* or an *NPN* combination. The three layers are referred to as the *collector*, *base*, and *emitter*; the two possible configurations are shown diagrammatically in Fig. 8.4, together with the conventional drafting symbols used to represent them. In an actual transistor

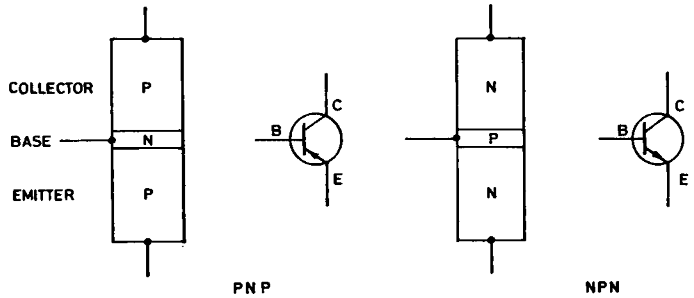


FIG. 8.4. Transistor configurations.

the collector and emitter differ in physical construction, and are not interchangeable, as Fig. 8.4 might suggest. The base layer is extremely thin; this is necessary if the transistor is to function.

All modern transistors are made of silicon, and can operate up to temperatures of about 120°C; many older transistors are of another semiconductor material, germanium. These latter have a number of undesirable features, and should be avoided. For manufacturing reasons, *NPN* transistors are more common than *PNP*, though both are readily available. In the following description, the use of *NPN* transistors is assumed; *PNP* transistors may be substituted if all voltages are reversed in polarity.

To understand how a transistor operates, consider the arrangement shown in Fig. 8.5. The emitter is regarded as the reference point in the circuit, and the voltage of the collector or base is always measured with respect to the emitter. The collector in Fig. 8.5 is held positive with respect to the emitter by means of a battery, typically of 6 to 15 V, and a milliammeter is inserted to indicate current flow from collector to emitter. A variable voltage source is connected from emitter to base.

If the base is set at a voltage somewhat negative to the emitter, it will be seen that the base-to-emitter *PN* junction is reversed biased, and no base current will flow. The collector-to-base *PN* junction is also reverse biased, and no collector current will flow.

If the base-to-emitter junction is now forward biased by taking the base about 0·6 V

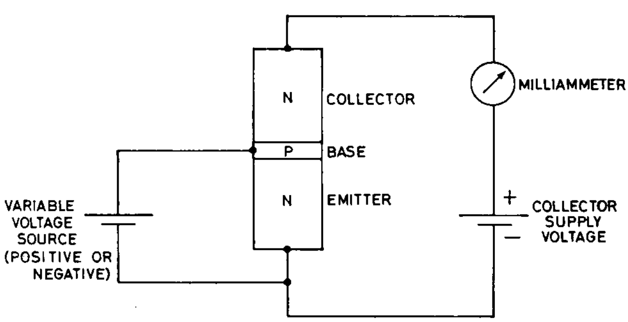


FIG. 8.5. *NPN* transistor connections.

positive to the emitter, a copious stream of electrons from the emitter will cross the junction. Because the base layer is extremely thin, most of these electrons will not be cancelled by holes from the base, but will be attracted across the base-to-collector junction, and will pass to the collector terminal, constituting a current in the collector circuit. (In terms of the convention used for direction of current flow, this is said to be from collector to emitter.) The magnitude of this current is much greater than that of the original base current, and so the transistor can be used in a circuit as a *current-controlling element.*

8.4 Transistor transfer characteristic

For convenience and clarity in drawing transistor circuits, a number of conventions are adopted. Transistors are drawn with the base to the left and the collector up, except on rare occasions. Voltage sources are usually omitted, and connections to them are indicated by an arrow and a note to indicate the source referred to. The other terminal of each voltage source is assumed to be connected to a common reference point, usually the metal chassis of the equipment, and this reference point is indicated by an *earth* symbol. These conventions have been adopted in the circuit of Fig. 8.6, which shows a

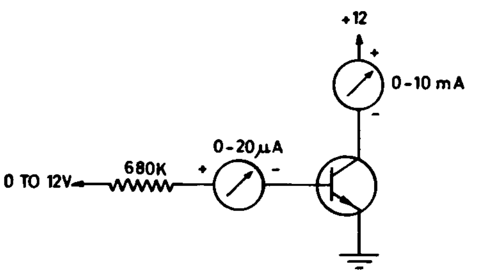


Fig. 8.6. Transistor test circuit.

practical method of obtaining an easily controlled base *current*, by means of a variable voltage source and a resistor. This circuit may be used to study the basic properties of a transistor. By selecting a number of values of base current, and observing the collector current of each, the *transfer characteristic* for the transistor may be drawn. This graph relates collector (or *output*) current to base (or *input*) current; its gradient is the *current gain* of the transistor. The current gain (usually represented by the symbols β or h_{fe}) may be regarded as a figure of merit for a transistor; values of 400 to 800 are typical. A typical transfer characteristic for a small transistor is shown in Fig. 8.7. If the graph is redrawn for different values of collector supply voltage, it is found to vary very little; *the collector current* of a transistor *is almost independent of the collector voltage.*

If the base voltage of a transistor is carried negative with respect to the emitter, the collector current remains cut off; care should be taken to prevent this negative voltage ever exceeding 5 V, or the base-to-emitter diode may break down and suffer permanent damage.

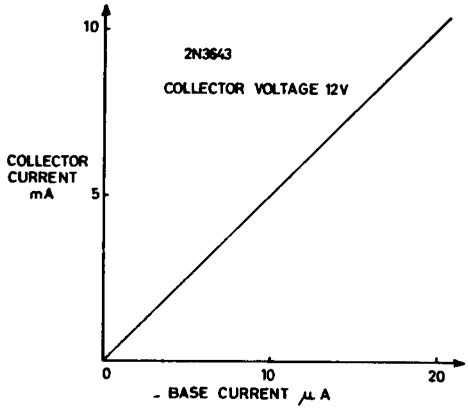


FIG. 8.7. Transistor transfer characteristics.

8.5 COMMERCIAL TRANSISTORS

Commercial diodes and transistors are available in an enormous and bewildering range. It is strongly recommended that any workshop for carrying out experimental work should have its own "preferred list", governed by local availability and price. A very limited number of types will be needed; as a guide it is suggested that one type in each of the classes of Table 8.1 will be quite adequate. This table includes categories

TABLE 8.1. DIODE AND TRANSISTOR TYPES

Application	Class	Rating	Example of type
Amplifier	*NPN*	100 mA	SE4002 (Fairchild)
	PNP	100 mA	2N3645 (Fairchild)
	NPN	250 V max supply	SE7056 (Fairchild)
Power amplifier, voltage regulator	*NPN*	3 W	AY8140 (Fairchild)
	NPN	40 W	2N3055 (Almost any manufacturer)
Switching	*NPN*		2N3643 (Fairchild)
	PNP		2N3645 (Fairchild)
Signal and meter rectifier		50 mA, 60 V peak inverse voltage	AN2001 (Fairchild)
Power rectifier		1 A 1000 V peak inverse	IN4007 (International Rectifier)
		Low voltage bridge	MB1 (ITT)
Zener reference element		6·2 V 1 W	IN4735 (International Rectifier)
Field effect transistor	*N* channel junction		2N3819 (Texas Instruments)

discussed later in this book, and gives the types in current use in the University of Melbourne.

8.6 Transistor voltage amplifier

The arrangement described in § 8.4 is a *current* amplifier; a small current at the base controls a larger current in the collector circuit. This is of value when it is required to operate a meter or other current-sensitive device by means of a very small current; more often it is desired to obtain a *voltage* which is proportional to the input current. This is done by the insertion of a *load resistor* in the collector circuit, as shown in Fig. 8.8. As the current through this resistor varies, so will the voltage across it. The output voltage e_{out} may be obtained by subtracting the voltage across R_L from the supply voltage:

$$e_{\text{out}} = 12 - i_c R_L. \tag{8.1}$$

For example, if i_c is 500 μA, $i_c R_L$ will be 5 V, and e_{out} will be 7 V: as i_c rises, e_{out} falls in proportion.

As we have seen in § 8.4, i_c is proportional to i_b:

$$i_c = \beta i_b. \tag{8.2}$$

So combining equations 8.1 and 8.2

$$e_{\text{out}} = 12 - \beta i_b R_L. \tag{8.3}$$

It is clear that e_{out} is proportional to i_b, and moves in the opposite direction to it; for $R_L = 10$ K and $\beta = 500$, a base current of 1 μA gives a value for e_{out} of 7 V.

e_{out} remains proportional to i_b so long as $\beta i_b R_L$ does not reach 12 V. If this occurs the transistor is said to have *bottomed*; e_{out}, which is also the collector voltage of the transistor, has fallen to zero, and can fall no further. From eqn. (8.3), this occurs if

$$\begin{aligned} e_{\text{out}} &= 0 = 12 - \beta i_b R_L, \\ i_b &= \frac{12}{\beta R_L} \\ &= 2{\cdot}4\,\mu\text{A}. \end{aligned}$$

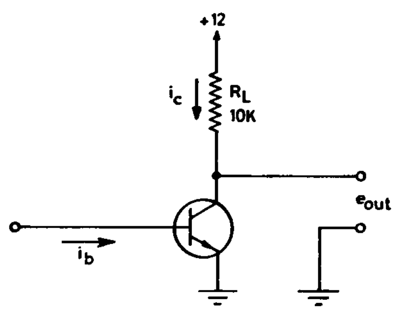


Fig. 8.8. Use of load resistor.

If it is desired to reproduce accurately an alternating input signal, such as speech or music, it is first necessary to set the transistor operating with some fixed value of base current, known as *bias* current. The signal to be amplified can then be combined with this bias current, so that it alternately adds to and subtracts from the bias current. It must never be great enough either to completely override the bias current (thus cutting off the collector current entirely) or to raise the total base current to a value that bottoms the transistor.

The bias and signal currents are most usually added by use of the circuit of Fig. 8.9. The input capacitor is large enough to offer a low impedance to any signal frequency in the range it is desired to amplify, but it prevents the steady bias current from passing back into the signal current source. (The selection of a suitable value is discussed in Chapter 12.)

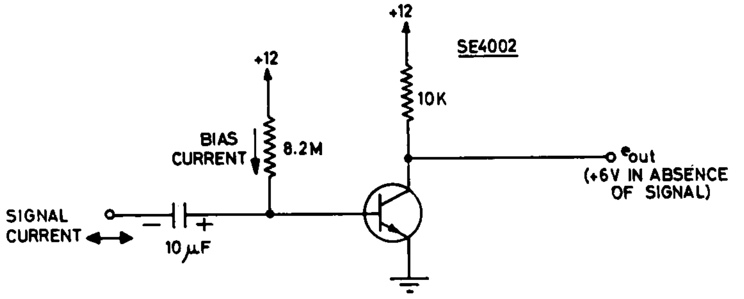


FIG. 8.9. Simple amplifier circuit.

To find the value of bias resistor required, we first assume that in the absence of a signal e_{out} should rest at *half the supply voltage*. It can then rise or fall by the same amount in operation. If e_{out} is to rest at $+6$ V, the collector current must be 6/10, or 600 μA. Assuming a typical β of 400 for the SE4002 transistor, the base current must be

$$i_b = \frac{i_c}{\beta}$$

$$= \frac{600}{400} = 1{\cdot}5\ \mu\text{A}.$$

To obtain the $1{\cdot}5\ \mu$A we can use the existing 12 V supply. Allowing 0·6 V for the base-to-emitter voltage, Ohm's law gives

$$R = \frac{12 - 0{\cdot}6}{1{\cdot}5}$$

$$= 7{\cdot}6\ \text{M}.$$

The nearest preferred value is 8·2 M.

8.7 Reproducibility

It is instructive to assemble the circuit of Fig. 8.9, and to measure its output voltage. It is unlikely that the resting collector voltage will be exactly 6 V; it may be anywhere from almost zero to almost 12 V. This is due to the fact that there is a wide tolerance allowed in the β of nominally identical transistors. Although specimens having unacceptably low β are rejected during manufacture, a range of values from 200 to 600 would not be exceptional in a single batch. Since transistors practically never need to be replaced, it is possible to adjust the bias resistor to suit the specimen used, after setting up the circuit. This technique is clearly unsuitable for commercial production, and more elaborate amplifier circuits are often used, to permit the employment of any specimen of the selected transistor type without "tailoring" of the bias resistor. The references at the end of the chapter may be consulted for further details.

8.8 Interconnection of amplifier stages

Subject to the lower limit set by the thermal noise in the components, and the upper limit set by the maximum undistorted peak-to-peak voltage swing a transistor can deliver, it is possible to obtain as much amplification of a signal as may be desired. This is done by *cascading* several amplifier stages, the output of the first driving the input of the second, and so on. Two coupled stages are shown in Fig. 8.10. Since the impedance of

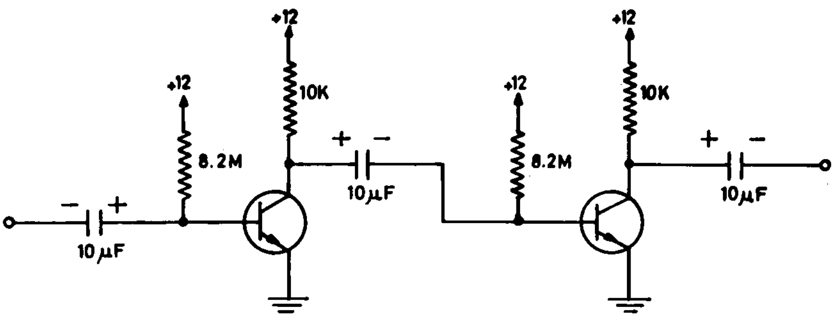


FIG. 8.10. Coupling between stages.

the coupling capacitor is small in the working frequency range of the amplifier, it may be neglected in a simple discussion. The load resistance that the first transistor "sees" is then its own collector load, the bias resistor of the second transistor, and the base-to-emitter resistance of the second transistor, all in parallel. The bias resistor, of 8·2 M, may clearly be neglected. The resistance of the second transistor from base to emitter at the bias current used is about 5 K. The transistor behaves as an almost perfect constant current generator, and its alternating current output when a signal is applied thus divides between the collector load and the base-to-emitter resistance of the next transistor, in proportion to their conductances. Hence the signal base current driving the second stage is

$$i_{b2} = \beta i_{b1} \frac{R_L}{R_L + R_{b2}} \tag{8.4}$$

where R_{b2} is the base-to-emitter resistance of the second transistor. The gain of the second transistor can then be calculated as before.

FURTHER READING

General Electric Transistor Manual (7th ed.) General Electric, New York, 1964.
ARRL Radio Amateur's Handbook, American Radio Relay League, Massachusetts, 1975.
RSGB Radio Communication Handbook, Radio Society of Great Britain.

PRACTICAL

Warnings (i) Always disconnect all power supplies before attempting alterations to a transistor circuit. Transistors are very easily destroyed by short circuits.

(ii) In experimental circuits, keep the wiring as short as possible. Modern transistors can easily generate high frequency oscillations if this point is not watched.

8.1 Connect a suitable resistor in series with a 50 μA moving coil meter, so that it serves as a 0–1 V voltmeter. Set up the circuit of Fig. 8.11 and tabulate the voltage across the diode for a range of currents up to 1 mA, by inserting suitable resistors in the position marked R. (Do not overlook the need to subtract the current being drawn by the voltmeter from the milliammeter reading!)

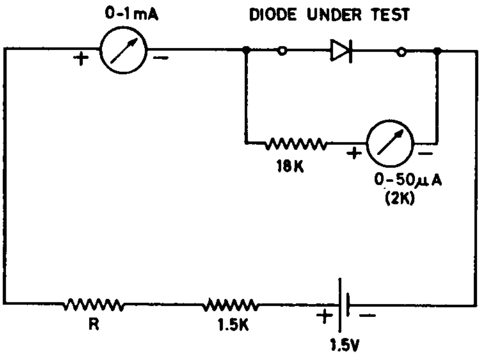


FIG. 8.11. Diode test circuit.

Draw a graph of your results, putting voltage on the horizontal axis, and current on the vertical. Now estimate the gradient of this graph at (a) 100 μA (b) 1 mA. This gradient represents the *dynamic conductance* of the diode at these points; the reciprocal of this quantity, the *dynamic resistance*, is often used as a measure of how close the diode comes to being a true short-circuit when it is conducting. Replot the graph on semilogarithmic paper, putting voltage on the linear axis and current on the logarithmic axis. What do you observe about this graph?

8.2 Repeat experiment 8.1, using the base-to-emitter junction of an *NPN* transistor as a diode. Leave the collector disconnected. Compare your results with those of experiment 8.1.

8.3 Set up a small *NPN* transistor in the circuit of Fig. 8.6, and plot its transfer characteristic. Repeat with collector supply voltages of (a) 15 V (b) 9 V. What is the β of the transistor for each supply voltage?

8.4 Set up the transistor amplifier of Fig. 8.9, and measure its collector voltage. If it is not about 6 V, vary the bias resistor to bring it into this region.

If an oscilloscope and sine wave generator are available, the amplifier may be studied in operation. Connect the oscilloscope to the amplifier input; set it to observe d.c. signals, and carefully note on its screen the position of zero volts, +12 V, and the resting output voltage. Apply an input signal from the sine wave generator through a 10 K resistor, using a frequency of 500 Hz, and adjust the level of the input until a satisfactory sine wave output is observed on the oscilloscope; note carefully the peak and trough voltages of the sine wave on the screen. Record the peak-to-peak input and output voltages, and hence calculate the voltage gain of the amplifier with its 10 K input resistor.

Now increase the input signal slowly, observing the output wave form for signs of distortion. Note carefully at what voltages the amplifier (a) cuts off (b) bottoms. What is the greatest peak-to-peak *swing* your amplifier will deliver without distortion?

This amplifier should be retained for use in later practical sessions.

8.5 Using the tuned circuit of experiment 7.5, set up the simple broadcast receiver of Fig. 8.12. The signals selected by the tuned circuit are amplified by the two coupled SE4002 amplifier stages, and are then rectified by the diode, as will be seen by examining the waveforms at points *A* and *B* with an oscilloscope. The broadcast speech or music is represented by the *average* value of the rectified signal, which is obtained by use of the 100 K resistor and 470 pF capacitor following point *B*. This average waveform may be seen at point *C* on an oscilloscope. If a high-resistance crystal earphone is connected between *C* and earth, it will follow this average waveform, and the broadcast signal will become audible.

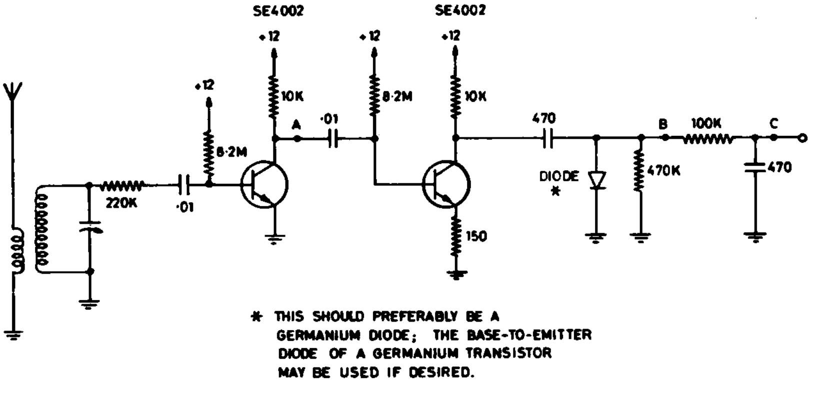


FIG. 8.12. Simple broadcast receiver.

CHAPTER 9

FIELD EFFECT TRANSISTORS

9.1 The field effect transistor

The *NPN* or *PNP* transistor discussed in the last chapter, as we have seen, is controlled by the magnitude of the *current* entering its base, and if voltage amplification is required the voltage source is usually applied through an input resistor. It is, however, possible to make semiconductor devices in which the output current is controlled directly by an applied *voltage*; these are known as *field effect* transistors. There are two basic types of FET available; these are the junction FET (JFET) and the insulated-gate FET (also known as a metal-oxide semiconductor FET, or MOSFET).

9.2 The junction FET

The junction FET consists of a single bar of either *N* or *P*-type semiconductor material, through which the output current to be controlled flows. Attached to the centre of this is a block of the opposite type of semiconductor, which forms a *PN* junction with it; in operation this *PN* junction is always reverse biased, and the block is referred to as the *gate*. The two connectors to the bar of semiconductor material (the *channel*) are the *source* and the *drain* (Fig. 9.1). The channel will carry current in either direction; the magnitude of this current is a function of the drain voltage. All voltages are measured with respect to the source. By the application of a reverse bias voltage to the gate, the centre of the channel is depleted of conductors (compare with Fig. 8.2); its apparent resistance rises, and the drain-to-source current falls. A sufficient gate voltage will cut off the source current entirely. If the gate is forward biased, the current flow in the channel is increased, but the gate then draws current, and its advantage is lost.

The transfer characteristic of a *N*-channel JFET may be drawn by the use of the circuit shown in Fig. 9.2: a *P*-channel FET may be investigated by reversing all supply polarities. Transfer characteristics for the 2N3819 JFET are shown in Fig. 9.3; the dotted curves show the tolerance allowed in transistors of this type. It is at once apparent that individual specimens of the transistor show widely different characteristics; however, the *slope* of the curves, in their relatively linear regions, is much the same. This slope is known as the *transconductance* of the FET, and will be seen to be about 3·5 mA/V. It is usual to operate the FET only in the linear portion of its characteristic.

9.3 The insulated-gate FET

The MOSFET, like the JFET, consists of an *N* or *P*-channel conductor from drain to source; however, it is not a *PN* junction, but consists of an electrode placed very close

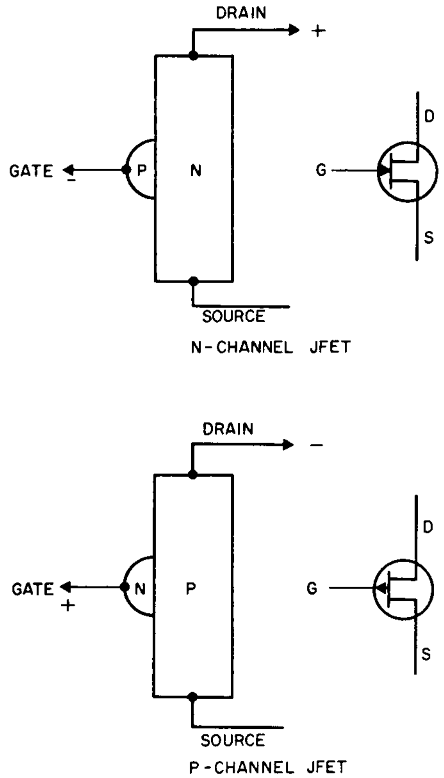


FIG. 9.1. Types of junction FET.

to the channel, and insulated from it by a thin film of metal oxide. Accordingly the gate can never draw current, whether biased positively or negatively to the channel; however, it can still control current flow in the channel, and can in fact be used to either *deplete* or *enhance* the flow. Four types of MOSFET can be made: *N* and *P*-channel depletion types, which are similar to the corresponding JFETs in operation, and *N* and *P*-channel

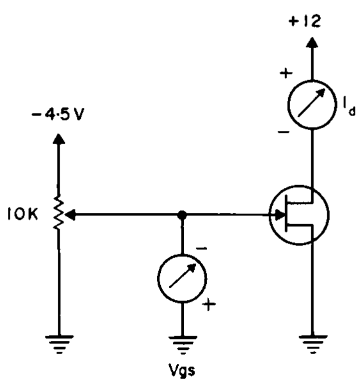


FIG. 9.2. *N*-channel JFET circuit.

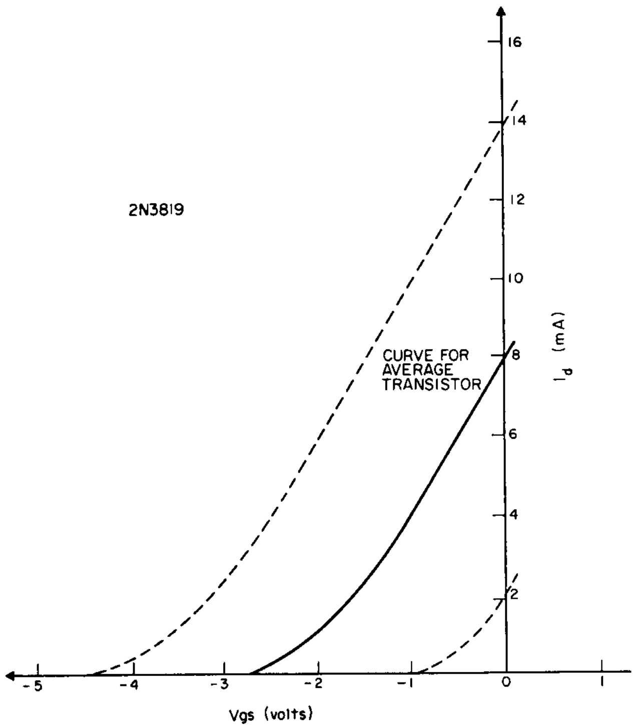


FIG. 9.3. Transfer characteristics of *N*-channel JFET.

enhancement types, whose channels are *non-conducting* unless enhanced by forward bias on their gates.

The gate insulation of MOSFETs is extremely good, typically 10^{14} ohms. This leads to a special problem in handling, since a gate lead left disconnected, even when the transistor is in its original package, can easily accumulate enough static charge on its gate-to-source capacitance to destroy the oxide film. Most modern MOSFETs are protected by the inclusion of internal diodes, connected permanently from gate to source. These permit the application of normal working voltages, but conduct if the voltage on the gate becomes excessive. Types without this protection are always sold with all leads short-circuited together, and this short-circuit must be maintained until the device is finally soldered into place in its circuit. The circuit must include a suitable "leak" resistor or other direct pathway from gate to source. Before using any unfamiliar type, the manufacturer's literature on it should be carefully studied.

9.4 Use of FETs

In their present state of development, FETs are less satisfactory as amplifiers than are bipolar transistors. They require higher voltages, give lower voltage gains, have much wider tolerances, and in the amplification of very small signals generate more internal noise. However, there are certain applications in which they are essential. In particular, there are many circumstances where a voltage to be measured has a very high source resistance; for example, the microelectrodes used to measure the resting and acting potentials of single living cells have a resistance of the order of 10 M. To record a potential of 50 mV with an accuracy of 1 %, Ohm's law shows that the recording device must not take a current in excess of 5×10^{-11} A.

Wherever possible, junction FETs should be used, because of the ease with which they can be handled. MOSFETs are seldom required as amplifiers, although they are used extensively as switches; this will be discussed later.

9.5 FET AMPLIFIER

The circuit of Fig. 9.4 shows a typical junction FET amplifier. In this circuit the wide variation in FET characteristics is compensated by the use of the 100 K resistor between source and earth, which also provides the operating *bias voltage*. An average FET in this circuit passes a drain-to-source current of 40 μA, putting the drain (and output) at +8 V, and the source at +4 V with respect to earth. In the absence of an input signal the gate will be at zero volts with respect to earth, since no current is flowing through the 22 M leak resistor. However, *with respect to the source*, earth is at −4 V; this is the correct bias to cause the FET to draw 40 μA. A specimen of 2N3819 which tends to pass a smaller current will produce less bias across the 100 K resistor, and so the current flowing will not decrease much; a specimen which tends to develop a greater current will produce more bias, and so the current flowing will not increase much.

The resistor from source to earth is *bypassed* by an electrolytic capacitor, which is selected to have a low impedance at even the lowest frequency the amplifier is intended to amplify; this means that the source is held steady at +4 V during operation, and the FET behaves as though it had a drain supply voltage of 8 V, and were biased to the mid-point of its operating range.

This circuit will have a voltage gain of about 20, and a maximum peak-to-peak output of about 2 V. These are much less than the values for the amplifier of Fig. 8.9, but the input resistance of this amplifier is 22 M (it could be higher if required), while the input resistance of the bipolar transistor amplifier is quite low.

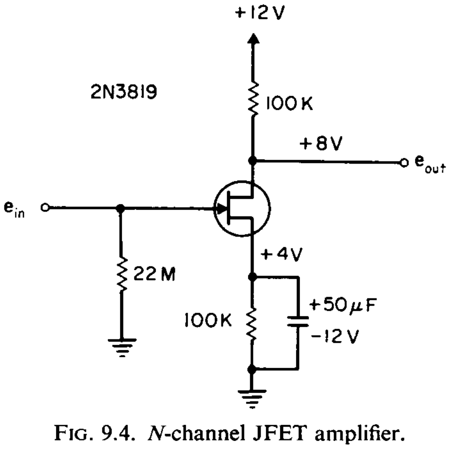


FIG. 9.4. *N*-channel JFET amplifier.

FET amplifier stages may be cascaded in the same fashion as bipolar amplifier stages, or a FET input stage may be used to drive a following bipolar amplifier.

9.6 FET GATE

A FET (often a MOSFET) may be used to *gate* a signal voltage, as in Fig. 9.5. In this way a signal may be passed or blocked, or a sample of the signal taken at some instant. (Notice the symbol used to represent a MOSFET.)

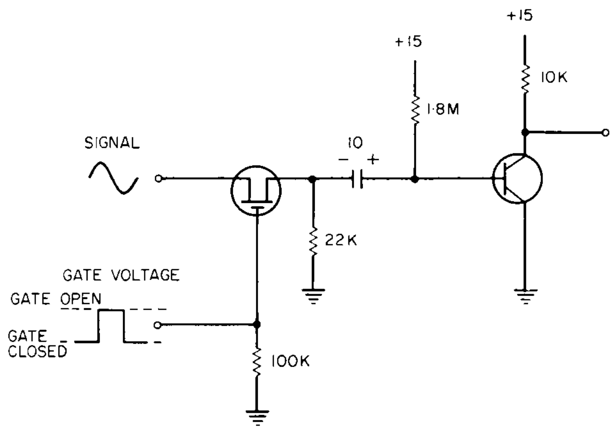


FIG. 9.5. MOSFET gate.

FURTHER READING

Field Effect Transistors, Texas Instruments Series, McGraw-Hill, New York, 1965.
MOSFETS, Texas Instruments Series, McGraw-Hill, New York, 1967.
Application notes are published by all manufacturers of FETs.

PRACTICAL

9.1 Set up a FET in the circuit shown in Fig. 9.2, and plot its transfer characteristic. Repeat this plot for one or two other specimens of FET, and observe the variation in characteristics.
9.2 Connect up the amplifier shown in Fig. 9.4, and measure its drain current. Calculate from this the drain and source voltages with respect to earth. (Why can you not measure these voltages directly with a voltmeter?) Repeat with one or two other specimens of FET, and note the variations in drain voltage.
9.3 Apply an input signal from a signal generator (about 10 mV peak-to-peak at 500 Hz) and observe the output signal at the drain with an oscilloscope; calculate the voltage gain. Now increase the input signal until the output shows distortion, and note the maximum undistorted output attainable.

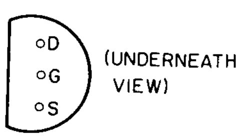


FIG. 9.6. Connections to 2N3819 transistor.

CHAPTER 10

POWER SUPPLIES

10.1 Reliability and safety

To ensure reliability of operation of electronic equipment, particularly following a period of disuse, it is essential to provide for its operation from the a.c. supply lines. Batteries should be avoided unless operation remote from a.c. supplies is envisaged; if they are used, they constitute the most probable cause of failure of the equipment.

The use of the a.c. power lines introduces the hazards of accidental electrocution of the technician, the user, or the patient, and the possibility of introducing a.c. interference into any recordings being taken; with due care in the design of the equipment and selection of components of adequate quality, the probability of these occurrences is negligible. It is strongly recommended that all persons concerned with the operation of electronic equipment familiarise themselves with the emergency treatment for electrocution and the local procedure for summoning aid, and that they do not work on any equipment unless a second person is at hand. In addition, any person concerned with attaching electronic equipment to a patient should be thoroughly acquainted with the risks involved; many of these are by no means obvious. (See Chapter 23.)

10.2 Types of rectifier circuit

The usual supply requirements for electronic circuitry are for one or more low-voltage d.c. sources capable of delivering a current of the order of 1 A or less. These supplies may be obtained by the use of a step-down transformer from the a.c. supply lines, followed by a rectifier circuit and suitable filtering to generate an a.c.-free supply of the correct voltage.

Four types of rectifier circuit are in common use, as shown in Fig. 10.1. In each of these circuits, R_s represents the total internal resistance of the power transformer; occasionally additional resistors may be used, but it is not usually necessary. The transformer secondary voltage is typically 10–50 V r.m.s., as required, and C is electrolytic, of suitable peak voltage rating, and about 2000 μF per ampere of load current to be drawn. The half-wave circuit of Fig. 10.1*A* is the simplest, and all the others are based on it. The waveforms which occur in this circuit when delivering current into a load are shown in Fig. 10.2.

The rectifier conducts only when the rising transformer voltage on the left of the rectifier exceeds the falling capacitor voltage on its right, and so rectifier and transformer current flow only for a short portion of each cycle. During the rest of the cycle, the whole of the load current is supplied by the capacitor alone. The *ripple* (the amount by which the capacitor voltage falls exponentially before the occurrence of the next charging pulse)

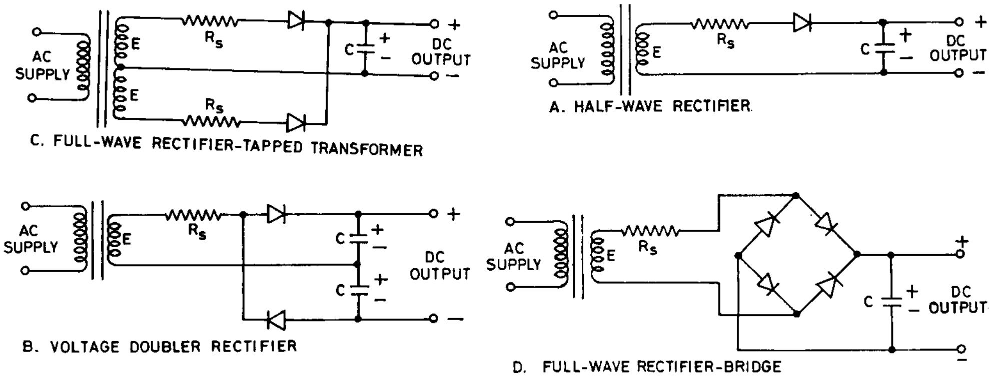


FIG. 10.1. Rectifier circuits.

is governed by the magnitude of this load current; if the load resistance is R, then the capacitor voltage falls with a time constant CR.

This circuit should appeal to those with a physiological background, since it is the exact analogue of the left ventricle (the transformer), the aortic valve (the rectifier), the great arteries (the capacitor) and the peripheral resistance (the load). Cardiac output occurs only when the left ventricular pressure exceeds the aortic pressure, and the circulation is sustained for the remainder of the cardiac cycle by the elasticity of the great arteries. The pulse pressure is analogous to the ripple.

At a phase 180° past the conducting peak the capacitor is still charged almost to the original peak transformer voltage, but the transformer voltage has reversed. For a transformer of peak voltage 50, the situation will now be as in Fig. 10.3. It will be seen that at this instant the rectifier has across it 100 V trying to make it conduct backwards; this is the *peak inverse voltage* which the rectifier must be rated to withstand. It is clearly approximately equal to twice the transformer peak voltage, or three times its r.m.s. voltage.

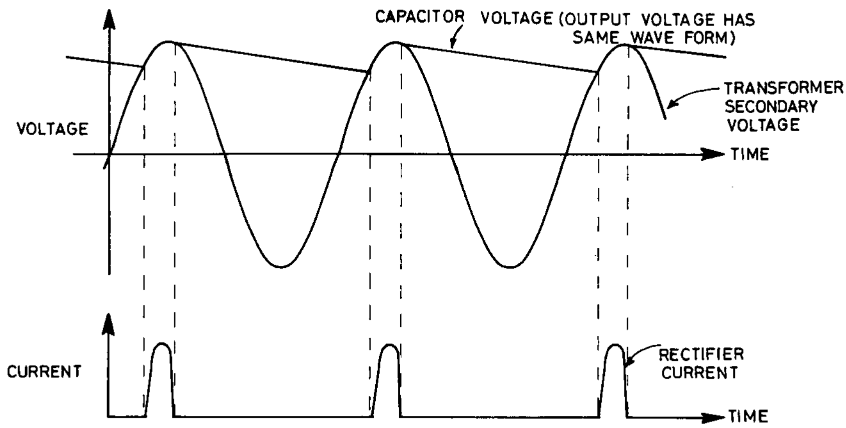


FIG. 10.2. Voltages and currents in half-wave rectifier supply.

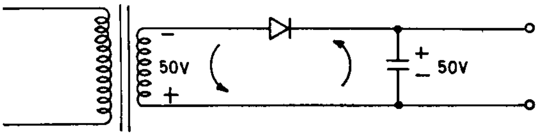


FIG. 10.3. Peak inverse voltage.

The *peak current* which the rectifier passes must be many times the fairly steady output current delivered by the circuit. This peak current is governed by R_s, which must be large enough to keep the peak current within the ratings of the rectifier. (R_s also serves to limit the much more severe *initial surge* when the a.c. supply line is switched on with the capacitor entirely empty.) A typical small rectifier will be rated at 200 V peak inverse, 1 A continuous output, 10 A peak current, 50 A initial surge.

The polarity of the output voltage can be reversed by reversing the diode (requiring the electrolytic capacitor to be reversed also). This uses the negative rather than the positive half cycles from the transformer to charge the capacitor. By combining positive and negative half-wave rectification, the *voltage doubler*, circuit *B*, is obtained. The upper capacitor is charged to the peak of *E* in a manner identical with circuit *A*, and the lower capacitor to the peak of *E* by circuit *A* reversed to give negative half-wave rectification. Since the two capacitors are in series their voltages add, and the transformer peak current flows twice in each cycle, to charge them alternately. The output voltage for no load is twice the peak value of *E*, and the ripple is now at a frequency of twice the supply frequency.

Circuit *C* requires two transformer windings, each of voltage *E*; it is effectively two circuits *A* connected to the one capacitor, and phased so that they charge the capacitor alternately, in successive half cycles; the ripple voltage is consequently halved, and is at twice the supply frequency.

Circuit *D* is the bridge rectifier already considered as a meter rectifier. It is a full-wave rectifier using only one transformer winding (unlike *C*) and one capacitor (unlike *B*). Since rectifiers are inexpensive, it is to be preferred unless some other reason suggests the use of *A*, *B*, or *C*.

10.3 DECOUPLING

Where a number of circuits are to be operated from the one power supply, a surge of current drawn by any one will tend to lower the supply voltage to all the others. This *coupling* between circuits may easily lead to the appearance of unwanted signals at various points in the equipment, and is usually avoided by additional filtering right at each individual circuit. This will usually consist of a *tantalum bead* electrolytic capacitor of about 10 μF connected to earth at the point where the supply enters each individual circuit; the effect may be improved, at the expense of the loss of some voltage, by inserting a low value resistor into the supply lead prior to the capacitor. This process is described as *decoupling*.

10.4 Graphical representation of power supply performance

The performance of a rectifier and filter circuit is best described graphically, by a set of curves showing output voltage and ripple as functions of load current and a.c. supply voltage. A properly-designed supply gives an adequate output voltage even when the a.c. supply voltage is low and the load current is at maximum, and does not overheat or break down when the a.c. supply voltage is high and the load current is zero. A fluctuation of $\pm 10\%$ about the nominal supply voltage is not unusual, and in a building with old wiring or a heavily loaded supply, may be even greater. Thus a nominal 240 V r.m.s. supply may possibly vary between 265 V late at night, and 215 V on a cold day. A typical small 20 V d.c. supply is shown in Fig. 10.4, and the regulation curves for it in Fig. 10.5. The curves for the supply line 10% high and 10% low are shown, with the peak and trough of the ripple voltage indicated in each case.

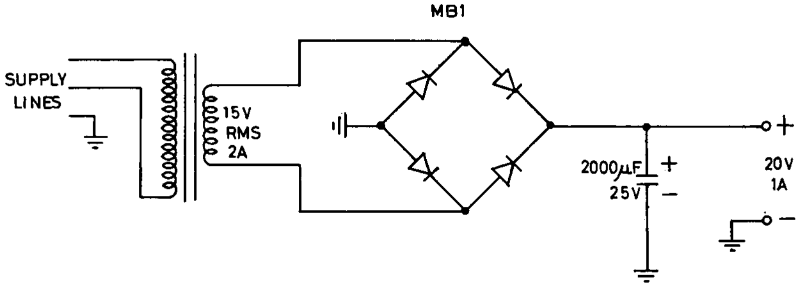


FIG. 10.4. Small d.c. power supply.

10.5 Regulation

For most purposes the variations of output voltage due to changes in a.c. supply voltage or load current, as shown in Fig. 10.5, are too great to be acceptable, and the ripple is also excessive. To produce an acceptable d.c. supply, the supply as already described is followed by an *electronic regulator* circuit. This is supplied with a voltage which under the worst conditions of a.c. supply voltage, load current and ripple, never falls below the desired constant output voltage, and in practice remains at least 2 V above it. The regulator then functions to remove the surplus voltage, and leave precisely the desired value. The regulator to be described in the following sections takes the nominal 20 V output from the unregulated supply, and delivers 15 V, at a maximum load current of 1 A.

10.6 Zener diode regulator

Although the *Zener diode* regulator is neither efficient nor particularly effective on its own, it will be described first, since it is incorporated in the regulator to be discussed in the next session.

If a rising voltage is applied in the reverse direction to a suitably designed diode, it will break down at a peak inverse voltage which is remarkably constant, and maintain this voltage across itself over a wide range of currents. Such diodes are called Zener

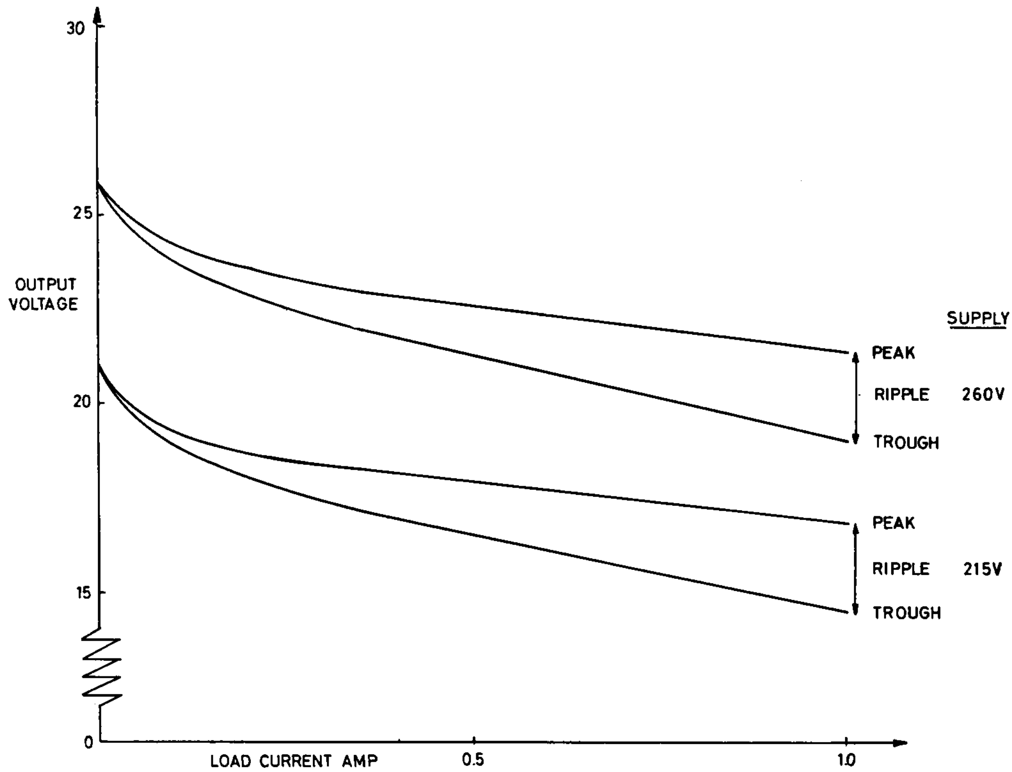


FIG. 10.5. Power supply regulation curve.

diodes. A typical circuit is shown in Fig. 10.6. Zener diodes are available in the voltage range from 3 to 200, and are commonly made in power ratings of 1, 3 and 10 watts; the larger sizes require a heat sink.

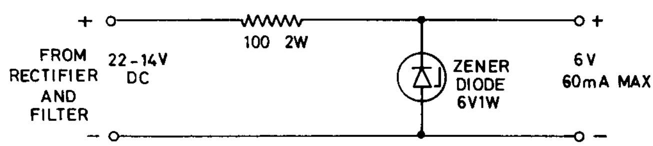


FIG. 10.6. Zener diode regulator.

The regulator is designed by first calculating the maximum current the diode can carry; a 6 V 1 W diode will carry

$$I = \frac{W}{E}$$

$$= \tfrac{1}{6}\ (Amp)$$

Since the diode always has 6 V across it, the series resistor must drop 16 V when the input supply is at its maximum of 22 V; the minimum resistor to restrict the current would then be

$$R = \frac{16}{\frac{1}{6}}$$

$$= 96 \text{ ohms}$$

Select 100 ohms: by Ohm's law, this will pass 160 mA for an input of 22 V, and 80 mA for an input of 14 V. If an output current is drawn, this current will be diverted to the load from the Zener diode, so for a low input the load current must not exceed 60 mA. Over the range from 22 V input and no load to 14 V input and 60 mA load the output voltage will in fact vary by about 0·2 V, since the Zener diode is by no means perfect.

Two semiconductor phenomena are involved in the functioning of Zener diodes. In low voltage diodes the *Zener* effect predominates; at higher voltages the constancy of breakdown is mainly due to the *avalanche* effect. In both cases the breakdown voltage varies with temperature, but the Zener effect diodes fall in voltage with a rising temperature, while avalanche effect diodes rise. By the selection of an intermediate breakdown voltage (6·2 V) a temperature-independent diode is obtained. These are always used when the highest stability is required.

10.7 Principle of the series regulator

The development of a typical *series regulator* may be understood by reference to Figs. 10.7 to 10.11. In Fig. 10.7 it is easy to see that a human operator could hold the output voltage of the regulator constant at 15 V by observing the voltmeter, and adjusting the series resistor R in such a direction as to correct any variations in output voltage due to change in input voltage or load current. In Fig. 10.8 the same arrangement is shown, with a large NPN silicon transistor substituted for the variable resistor; its effective series resistance is controlled by varying its base current manually with resistor R1, and so correcting output voltage variations as before.

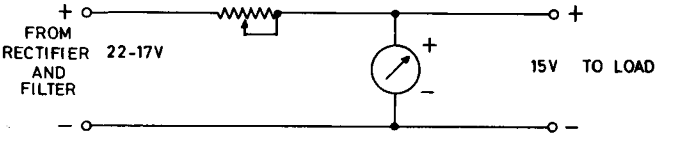


Fig. 10.7. Manual series regulator.

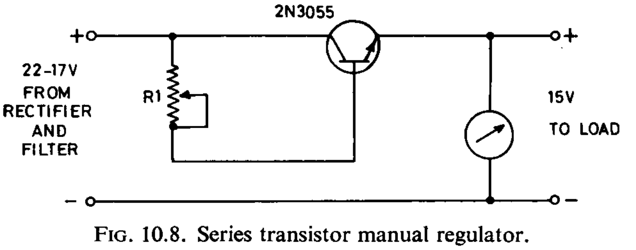


Fig. 10.8. Series transistor manual regulator.

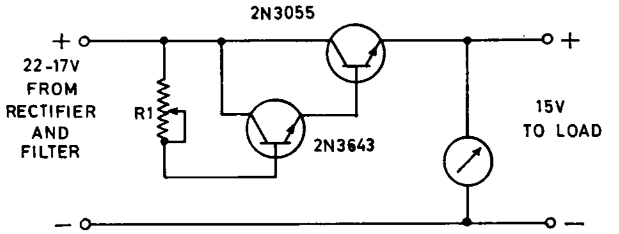


FIG. 10.9. Dual series transistor manual regulator.

If we desire a maximum output current of 1 A, the 2N3055 will require a base current of about 50 mA (minimum $\beta = 20$). This can be supplied from the emitter of a second smaller transistor, whose small base current is in turn controlled by the manually operated variable resistor. If its β is 100, only 0·5 mA will now be required to control 1 A (Fig. 10.9). Since deviations of the voltmeter from 15 V tell the operator that a re-adjustment is necessary, a meter to indicate *deviations from* 15 *V* rather than the actual 15 V would be useful. This may be done as in Fig. 10.10. A 6·2 V Zener diode is maintained in operation by the current through R2. By designing the voltage divider R3, R4 suitably, its tap can be put at 6·2 V when the output voltage is 15 V; the voltmeter will

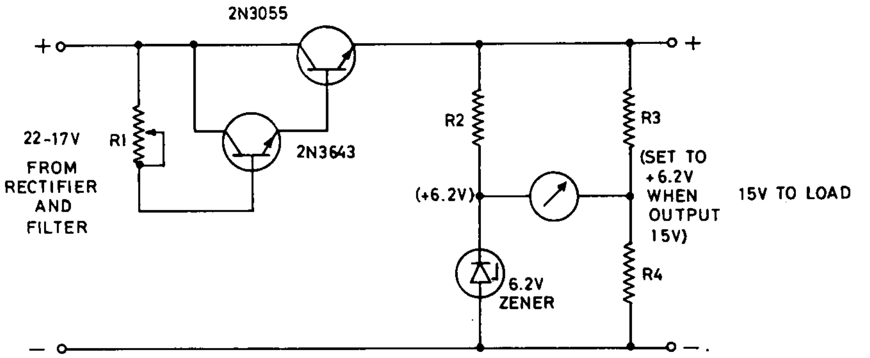


FIG. 10.10. Voltage deviation sensing circuit.

now read zero. Should the output voltage vary, the meter will immediately deviate from zero, and the direction of deviation will indicate to the operator the direction of the manual correction required in R1. A simple addition to the circuit of Fig. 10.10 will render the whole system automatic, and allow the human operator to be dispensed with. All that is necessary is to arrange that a deviation in output voltage will cause an appropriate readjustment of the base current of the 2N3643, and this can be done by replacing the voltmeter with a further transistor, as shown in Fig. 10.11. Since the SE 4002 has its emitter connected to a constant 6·2 V, and its base to the voltage divider R3, R4, a rise in output voltage of the regulator will cause its base current to rise. Its collector current will rise, and this will divert current previously supplied through R5 to the base of the 2N3643. This will in turn reduce the base current of the 2N3055, its resistance will rise,

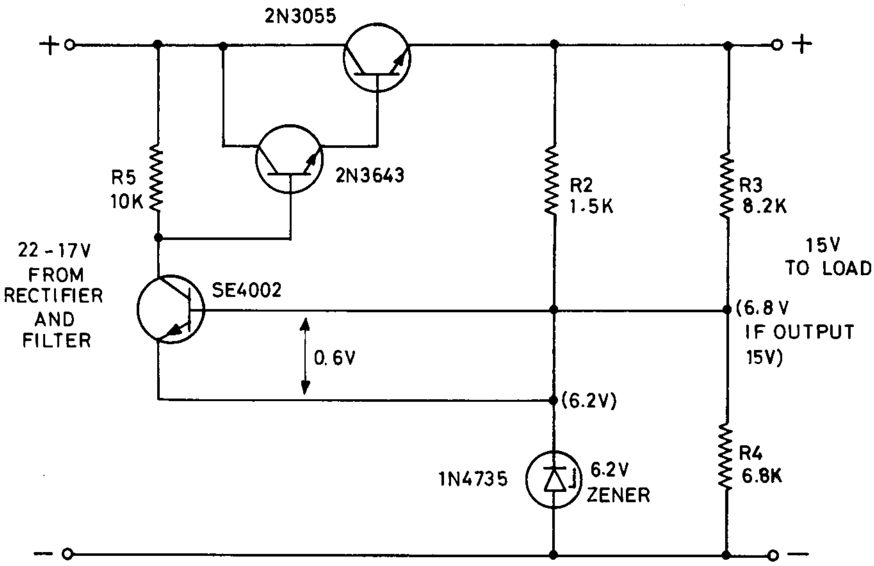


FIG. 10.11. Automatic series regulator.

and the original output voltage rise will be checked. Should the output voltage tend to fall, the reverse sequence will occur. The SE4002 acts as a *comparator* between the desired voltage and the actual voltage.

10.8 Practical series regulator

While the design of Fig. 10.11 is quite functional, a number of improvements can be made. The final design adopted for construction in the practical session is shown in Fig. 10.12. Fig. 10.12 differs from Fig. 10.11 (a) by the addition of an output capacitor

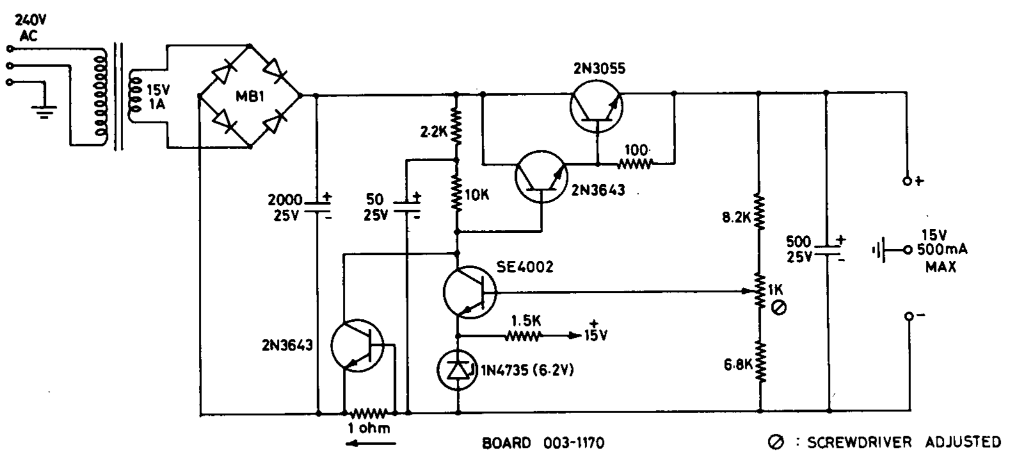


FIG. 10.12. Practical series regulator.

to supply sudden demands for current, which may be too fast to allow the regulator to compensate, (b) by the introduction of a small adjusting potentiometer in the voltage divider, to permit the voltage to be set at exactly 15 V, (c) by a decoupling circuit of 2·2 K and 50 μF to reduce ripple in the base supply of the 2N3643, (d) by a 100 ohm resistor in the emitter circuit of the 2N3643, which improves the regulation at small load currents, and (e) by the introduction of a *current overload protection circuit.* The 2N3643, a silicon transistor, requires its base to be about 0·6 V positive to its emitter before its collector commences to conduct. Current returning from the load to the rectifier is made to pass through a 1 ohm resistor; provided this current is less than 0·6 A, the 2N3643 connected across the 1 ohm resistor is inoperative. Should the current exceed this value, this 2N3643 diverts extra current from the base of the series 2N3643, raises the resistance of the 2N3055, and drops the output voltage drastically, thus protecting the regulator.

Fig. 10.13 shows how the components of this regulator may be assembled on a printed card, designed to plug into a strip connector. The transformer, rectifiers, filter capacitors, and the 2N3055 on its heat sink are mounted separately. Fig. 10.14 is a photograph of the complete assembly.

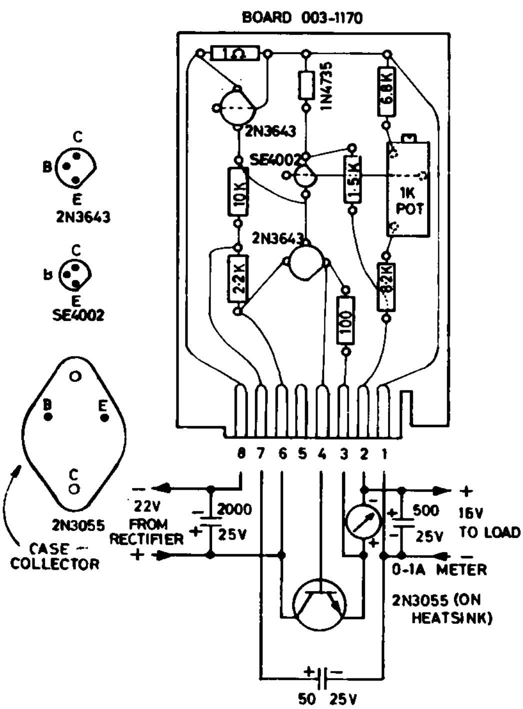


Fig. 10.13. Printed circuit card for regulated supply.

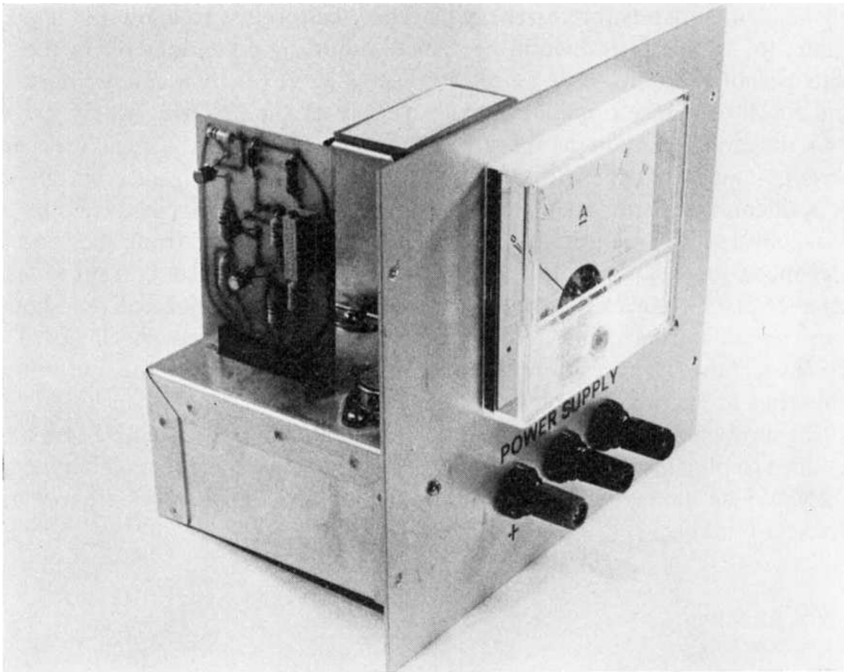


FIG. 10.14. Complete 15 V 0·5 A series regulator.

10.9 COMMERCIAL REGULATOR PACKAGES

In modern circuit practice the most commonly required supply voltages are +15, −15 and +5, with respect to earth. Complete regulator units are available in a single package for each of these voltages; they require only to be fastened to a heat sink, and to be supplied with a suitable filtered unregulated d.c. supply. There is little doubt that these units will completely supplant discrete wired assemblies such as have been described above.

FURTHER READING

Hewlett-Packard Application Note AN718, "Patient Safety".
STRONG, *Biophysical Instrumentation*, Tektronix, Beaverton, 1970.

PRACTICAL

The construction of a regulated power supply to deliver 15 V d.c. at currents up to 0·5 A is the first major project in this book. Two such supplies will be required for a number of the experiments in later practical sessions.

10.1 Commence construction of the power supply by wiring up the transformer, bridge rectifier, and filter capacitor.

10.2 Temporarily connect a d.c. voltmeter across the unregulated output, and a milliammeter in series with a dummy load across the output. (A suitable dummy load for testing small regulated or unregulated supplies is shown in Fig. 10.15. Such a device is often required.) Plug a Variac (a continuously variable

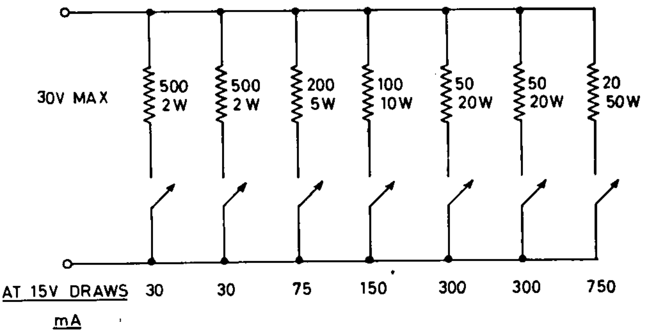


FIG. 10.15. Dummy load for low voltage supply tests.

power transformer) into the a.c. supply line, and attach an a.c. voltmeter to its output. Set it to the nominal a.c. supply voltage (240 V or other) and plug your power supply into it. Plot the d.c. output voltage curve over the load current range from 0 to 1 A, checking that the a.c. supply voltage is constant before each reading.

10.3 Connect an oscilloscope across the dummy load. Switch its input to d.c.; set the Variac to deliver a line voltage 10% above the nominal value, and plot a pair of curves showing the peak and trough values of the output voltage plus ripple as a function of load current, as in Fig. 10.5. Repeat for a line voltage 10% below the nominal value. These curves supply all the data necessary to design a regulated supply; they are always required when a new type of transformer, rectifier, or filter capacitor is to be used, to ensure that the regulator will, under all circumstances, receive sufficient input voltage.

Remember that although the secondary circuit operates at low voltage, the *primary wiring is at a lethal voltage*. Unplug the equipment before attempting *any* modifications.

10.4 Carefully assemble the components of the series regulator on the printed circuit board, soldering the joints neatly and quickly with a clean iron, and a minimum application of heat and solder.

10.5 Complete the wiring of the 2N3055 series transistor. Note that this uses the chassis as a heat sink, but is insulated from it. Make the appropriate connections to the socket for the printed card. Check the whole circuit carefully; notice that there should be *no* physical connection between the chassis and any point in the circuit. The third terminal on the front panel of the power supply is connected to the chassis, and either the positive or the negative output terminal of the supply may be linked to this, as required.

10.6 Test the whole supply by use of a Variac, dummy load, and suitable meters. Ensure that the overload protection is functioning. Set the output voltage to 15·0 V precisely by means of the potentiometer on the circuit card.

10.7 Check the output at full load for residual ripple, using an oscilloscope.

10.8 Measure the dissipation (volts × amperes) in the 2N3055 at full load, using an a.c. supply voltage 10% high.

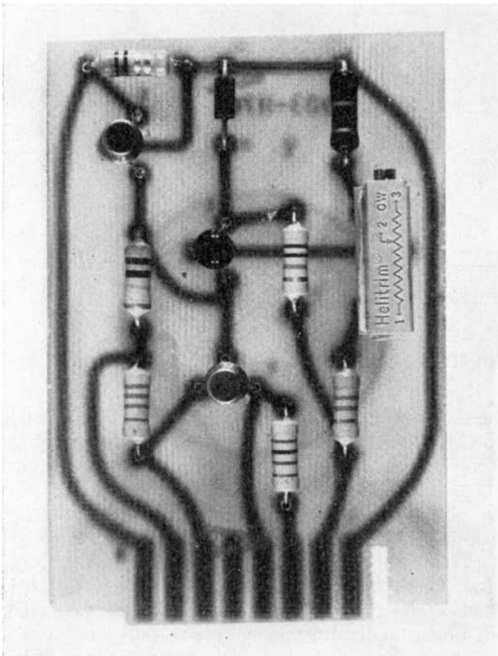


FIG. 10.16. Printed circuit card for regulated supply.

CHAPTER 11

OPERATIONAL AMPLIFIERS

11.1 The operational amplifier

Although any desired amplification may be produced by the use of a chain of field effect or bipolar transistor amplifiers, as discussed in Chapters 8 and 9, it is now more usual to employ complete packaged amplifiers. Such an amplifier consists of a considerable number of transistors, with their associated resistors and capacitors. In modern practice these are all formed automatically on a single silicon chip, typically having an area of 1 square millimetre, and this chip is then mounted in a single package, which may be similar to a large transistor, or may be moulded into a small plastic rectangle.

These assemblies are known as *operational amplifiers*, since they may be used to carry out various mathematical operations on a signal with high precision.

Operational amplifiers have a gain of the order of 100,000; however, they are never used in this fashion, but in conjunction with external resistors, which provide *feedback*, and reduce the gain to some lower value. Typical amplifiers are shown in Fig. 11.1, together with a matchstick.

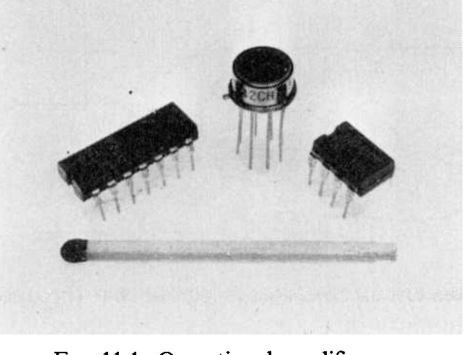


Fig. 11.1. Operational amplifiers.

An operational amplifier is usually provided with *two* input connections and one output connection, together with connections for a positive and a negative supply (most commonly +15 volts and −15 volts). In addition, other connections to allow external modification of the amplifier characteristics may be brought out. These will be discussed later.

The two input connections are known respectively as the *non-inverting* and the *inverting* input. When both inputs are at earth, the output is designed to be at zero with respect to earth. A small positive signal on the non-inverting input causes the output terminal to move positively; the same positive signal on the inverting input causes the output terminal to move negatively, by an equal amount. Both inputs are designed to have a high resistance, and thus to take a negligible current from the sources of signal connected to them; the output has a low source resistance.

The operational amplifier is represented diagrammatically by a triangular symbol, as shown in Fig. 11.2.

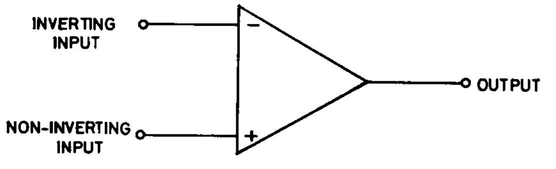


FIG. 11.2. Operational amplifier symbol.

11.2 PRINCIPLE OF INVERTING AMPLIFIER

The simplest use of an operational amplifier is in the *inverting* configuration; Fig. 11.3 shows the principle of this connection. Using the values shown, this circuit will have a gain of precisely −1. If the input is at zero, the output will be at zero. If the input is raised to +1 V, the output moves to −1 V, and so on. If the input is driven from a small sinusoidal voltage source, the waveform is reproduced at the output with an identical amplitude, but with a 180° phase difference.

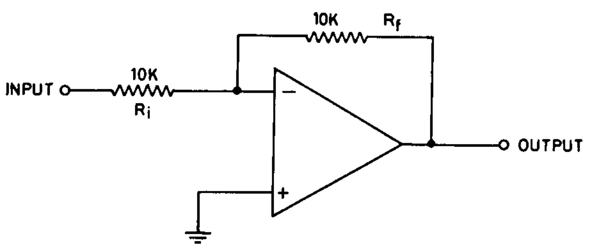


FIG. 11.3. Principle of inverting amplifier.

To understand how this circuit operates, suppose that the operational amplifier alone has a gain of 100,000. Suppose that a d.c. voltage is applied to the input of the circuit, and that it has a value such that the output moves by +1 V. We will now calculate this value using Ohm's law.

Since the output moves by +1 V, and the amplifier has a gain of 100,000, the inverting input of the amplifier itself must have moved by 1/100,000 V, or 0·00001 V, in a negative direction. This means that there is now a voltage of 1·00001 V across R_f, the *feedback* resistor. By Ohm's law, a current of 100·001 μA must be flowing through it from right to left. This current is coming from the output; where is it going? The inputs of the operational amplifiers have a very high resistance, so it is not going there; it must be continuing on through the 10 K input resistor R_i, and from there into the d.c. driving source. In passing through the input resistor, it will by Ohm's law produce a drop of 1·00001 V across it. Since, as we have seen, the inverting input of the operational amplifier is at $-0{\cdot}00001$ V, the input of the whole circuit is clearly at $-1{\cdot}00002$ V, and this is the value we set out to calculate.

Thus $-1{\cdot}00002$ V at the input has produced $+1{\cdot}00000$ V at the output; the voltage gain is $1{\cdot}00000/-1{\cdot}00002$, or $-0{\cdot}99998$. For practical purposes, it may be taken as $-1{\cdot}0$.

We have of course assumed that both resistors are precisely 10 K; they must be selected to high accuracy on a Wheatstone bridge.

The inverting input of the operational amplifier clearly never departs from zero volts with respect to earth by more than a few microvolts; it is sometimes described as a *virtual earth.*

Although this unity gain amplifier gives no voltage gain, its output resistance is low (a fraction of an ohm), and its input resistance is quite high (in fact, equal to R_i), so it provides both current and power amplification. It may be used in place of a transformer to match a high impedance voltage source to a low impedance load. It is also commonly used when a signal identical to but inverted with respect to another signal is called for.

If in Fig. 11.3, R_f is increased to 100 K, and the previous argument is adapted correspondingly, it will be seen that the circuit has a gain of almost exactly -10. For any other values of R_f and R_i, the gain is given almost exactly by

$$G = -\frac{R_f}{R_i}, \tag{11.1}$$

11.3 Practical operational amplifiers

Operational amplifiers may be divided into two broad classes, those with bipolar transistor inputs, and those with FET inputs. The latter are considerably more expensive, and tend to have other minor disadvantages, but the currents taken by their input terminals are much less. For practical purposes, one preferred type in each class is probably adequate for a research workshop. In this book the μA741C will be used as the bipolar preferred type, and the LH0042C as the FET type. Table 11.1 gives abridged

Table 11.1

	μA741C	LH0042C	Unit
Supply voltages	+15, −15	+15, −15	V
Input bias current	80	0·05	nA
Input offset current	20	0·01	nA
Input resistance	2	1,000,000	M
Output voltage swing	±12	±12	V
Output short-circuit current	25	20	mA
Permissible duration of output short-circuit	Any	Any	
Voltage gain without feedback	200,000	100,000	
Small signal bandwidth	1	1	MHz
Slew rate for unity gain amplifier	0·5	3	V/μS
Power supply current	1·7	2·8	mA
Power dissipation	85	120	mW

specifications for the two types. The μA741C was developed by Fairchild Semiconductor, and is manufactured by themselves and a number of other firms; the LH0042C is manufactured by National Semiconductor.

The pin connections for the two amplifiers are identical, as are the precautions to be taken in using them. It is essential to bypass each of the two power supplies close to the amplifier with a *tantalum* electrolytic capacitor; failure to do this is certain to result in instability of operation. The connections to the two input terminals must be kept as short as possible, particularly for the LH0042C.

The *input bias current* for an operational amplifier is the average of the input currents drawn by its two input terminals; to avoid this current in the μA741C affecting the operation of the external circuit it is usual to make the resistance seen by both amplifier input terminals the same. (The bias current of the LH0042C is for most purposes negligible.) The circuit of Fig. 11.3 may be modified as shown in Fig. 11.4 for this purpose; R_b is made equal to R_i and R_f in parallel. The *input offset current* is the typical *difference* that may exist between the currents of the inverting and non-inverting inputs; in a circuit such as that of Fig. 11.4 it would cause the output voltage for different speci-

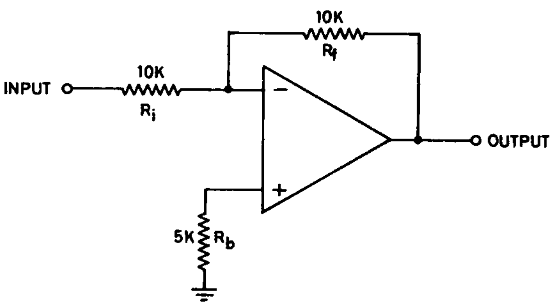


FIG. 11.4. Compensation for input bias current.

mens of operational amplifier to differ from zero when the circuit input was at zero. If this variation is significant in a particular application, it may be corrected over a small range by the use of an *offset potentiometer* of 10 K between pins 1 and 5, with its wiper returned to the −15 V supply. *The offset current of the μA741C restricts the maximum resistance an input may see to about* 220 *K*. If larger values are called for in a circuit, the LH0042C should be used.

The basic circuit of a μA741C or LH0042C is shown in Fig. 11.5, and a suitable small printed circuit board on which to assemble it in Fig. 11.6. In schematic drawings, it is conventional to omit power supply connections. Offset controls are generally drawn if they are to be used.

11.4 FREQUENCY RESPONSE AND SLEW-RATE

An operational amplifier is limited in regard to the highest frequency it can reproduce. For a small signal input and output (say 100 mV or less) this limitation is expressed as the *bandwidth for a unity gain amplifier*. The bandwidth is inversely proportional to the gain used; if a μA741C or LH0042C is set up to have unity gain, it will handle all fre-

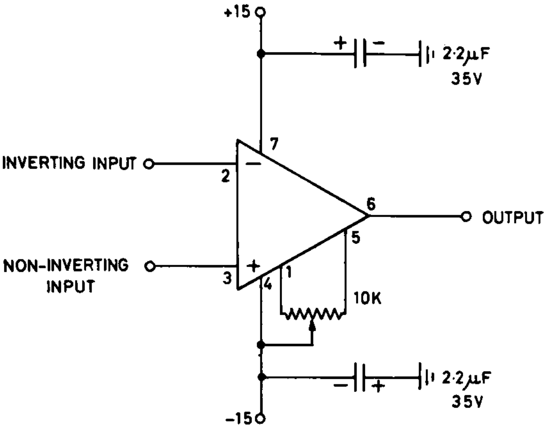


FIG. 11.5. Basic circuit for μA741C or LH0042C.

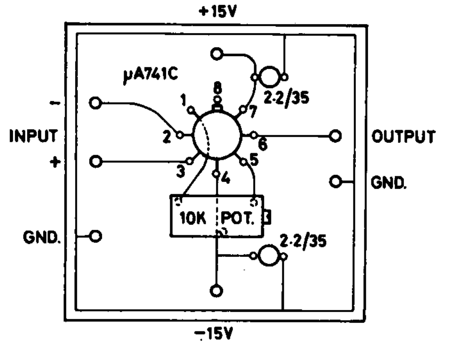


FIG. 11.6. Printed board for circuit of Fig. 11.5.

quencies up to 1 MHz. For a gain of 2, it will respond up to 0·5 MHz, and so on; gain may be traded off against bandwidth.

For a large abrupt change in input, the ability of the output to respond is much less than the small signal bandwidth would lead one to suppose. This ability is expressed in terms of *slew rate,* in volts per microsecond change at the output for an abrupt input change.

Other operational amplifiers are available which have better small signal bandwidths and slew rates than the μA741C or LH0042C. At present all of them require additional compensating capacitors connected externally. Many types have disadvantages which may not become apparent until they are actually used in a particular circuit.

11.5 Summing amplifier

An inverting amplifier may be set up with two or more inputs, as shown in Fig. 11.7. In this case, the output voltage e_0 is given by

$$e_0 = -\left(\frac{R_f}{R_1}e_1 + \frac{R_f}{R_2}e_2\right). \tag{11.2}$$

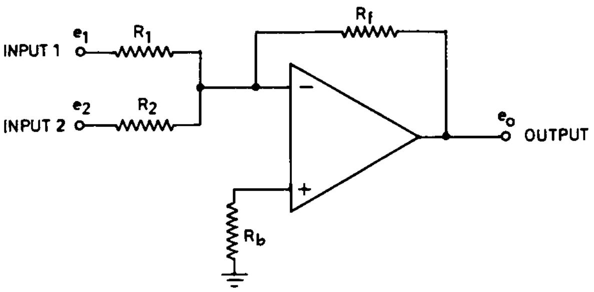


FIG. 11.7. Summing amplifier.

If for example $R_f = R_1 = R_2 = 100$ K

$$e_0 = -(e_1 + e_2).$$

The amplifier provides a convenient and accurate means of adding two voltages, without either affecting the other. Either voltage may be multiplied by its own scale factor R_f/R prior to summing, and this scheme may be extended to any number of inputs. R_b is as before made equal to the resistance the inverting input sees; in the example given it should be 33 K.

11.6 Non-inverting amplifier

Where a large input resistance is required, or an amplifier which does not invert its input signal, it is possible to inject the signal into the non-inverting terminal, and to ground the normal inverting circuit input; this gives the arrangement of Fig. 11.8. Appropriate

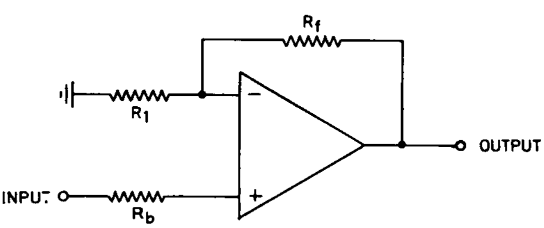


FIG. 11.8. Non-inverting amplifier.

values for this circuit are shown in Table 11.2. The gain is now given by

$$G = 1 + \frac{R_f}{R_i}. \tag{11.3}$$

R_b is as before made equal to R_f and R_i in parallel.

Where a gain of unity is required with a large input resistance, the very simple circuit of Fig. 11.9 may be used; this arrangement is known as a *voltage follower*. This circuit is often used in the input stage of biological amplifiers.

TABLE 11.2

Gain	R_i	R_f	R_b	Bandwidth	R_{in}(μA741C)
10	1 K	9 K	900	100 kHz	400 M
100	100	9·9 K	99	10 kHz	280 M
1000	100	99·9 K	99·9	1 kHz	80 M

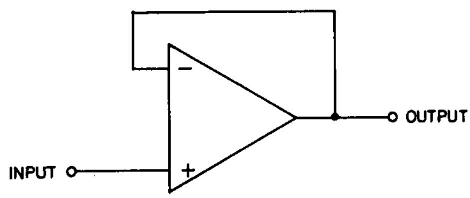


FIG. 11.9. Voltage follower.

11.7 DIFFERENCE AMPLIFIER

The difference between two signals may be obtained by using the circuit of Fig. 11.10. This circuit has a gain of -1 from e_1 to e_0; in the absence of resistor R2 it would have a gain of $+2$ from e_2 from e_2 to e_o (eqn. (11.3)). However, R1 and R2 form a voltage divider with a gain of 0·5, so with R2 present the gain from e_2 to e_0 becomes $+1$.

Accordingly,

$$e_0 = e_2 - e_1. \tag{11.4}$$

All four resistors must of course be matched with great care, and must be highly stable with time.

The difference amplifier may readily be given a gain greater than 1, by suitably increasing the ratios R4/R3 and R2/R1.

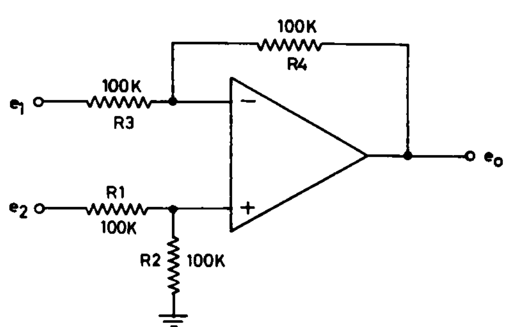


FIG. 11.10. Difference amplifier.

11.8 Integrator

One of the most valuable applications of the operational amplifier is as an *integrator*. In this configuration the output voltage is proportional to the *area* accumulating under the input. For example, if the input represents rate of flow of air expired by a patient, the output is proportional to total volume expired. Fig. 11.11 shows a typical circuit.

Suppose that initially the capacitor is completely discharged, and that the circuit input, the amplifier inverting input, and the output are all at zero. Now let the circuit input be taken abruptly to +1 V, and held there. By Ohm's law, a current of 10 μA will commence to flow from the circuit input towards the amplifier non-inverting input. Since this has a very high resistance, the current is caused to flow into the capacitor, so this must commence to charge at 1 V/sec ($i = C\dot{e}$). The amplifier non-inverting input is a virtual earth, so the left hand side of the capacitor remains at zero, and the right hand side and the amplifier output move negatively at 1 V/sec.

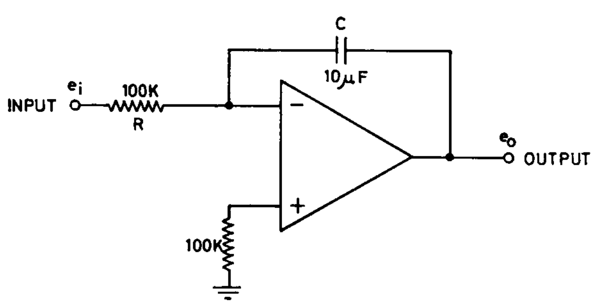


Fig. 11.11. Integrator.

If the +1 V is left on for 2 seconds, the output will be at −2 V; if it is left on for 5 seconds, at −5 V, and so on. It can easily be seen that for any input voltage left on for any time, the output will be at a voltage representing (input voltage × time); even if the input voltage is fluctuating, the output voltage will still total up the effective area which has accumulated. In calculus notation, it can be shown that

$$e_0 = \frac{1}{RC}\int_0^t e_i \, \mathrm{d}t. \tag{11.5}$$

In the circuit of Fig. 11.11, $RC = 1$, and

$$e_0 = \int_0^t e_i \, \mathrm{d}t.$$

This circuit is frequently used to determine areas under varying signals, and to generate linearly rising or falling voltages. A shorting switch across the capacitor is usually incorporated, to reset the output to zero at the beginning of a measurement; a FET may be used as a gate in this position.

FURTHER READING

GILES J. F., *Fairchild Semiconductor Linear Integrated Circuits Applications Handbook*, 1967.
HERBST L. J., *Discrete and Integrated Semiconductor Circuitry*, Chapman & Hall, London, 1969.
TOBEY, GRAEME and HUELSMAN, *Operational Amplifiers*, McGraw-Hill, New York, 1971.
Manufacturers' data sheets and application notes.

PRACTICAL

The pin connections and layout of the operational amplifier card are shown in Figs. 11.5 and 11.6.

11.1 Assemble the amplifier card as shown in Fig. 11.6, using a μA741C. Connect the two 15 V power supplies, returning their earth connections to the earth pin. Wire the card as in Fig. 11.4, to form a unity gain inverting amplifier, and provide a voltmeter to measure the output.

Earth the circuit input, and observe the output voltage; note that the offset potentiometer has very little effect on the output in this configuration.

Apply various positive and negative input voltages from a battery, from $+12$ V to -12 V. Tabulate and graph input voltage against output voltage. Should the measured gain not be precisely -1, find out why, and correct the fault.

11.2 Reconnect the card to give an amplifier with a gain of exactly -10, and repeat the calibration procedure of experiment 11.1.

11.3 Connect the card to form a summing amplifier, and repeat; earth any unused input.

11.4 Connect the card to form a non-inverting amplifier with a gain of 2, and repeat. Also repeat using the voltage follower connection.

11.5 Connect the card to form a difference amplifier. Measure the gain from each input terminal separately to the output, earthing the unused terminal. By how much do the *amplifier* inputs move in each of the two cases? Show by applying two different input voltages simultaneously that the output is in fact their difference. What happens if *the same* voltage is applied simultaneously to both inputs? By how much do the *amplifier* inputs move in this case?

11.6 Connect the card to form an integrator; provide a switch to short the capacitor as required. With the input earthed and the capacitor not shorted, adjust the offset potentiometer until the output is not drifting either up or down. Apply input voltages of $+1$, -1, $+2$ and -2 V, and for each measure the rate of rise of output voltage. With the capacitor initially discharged, apply each of these voltages for 3 seconds, and note the output reading in each case. Repeat for other periods of time, and show that output $=$ (input $\times$ time).

Using a 10 K potentiometer and a 3 V battery generate a fluctuating voltage of $\pm$ 1·5 V, and observe the output voltage produced.

CHAPTER 12

FREQUENCY RESPONSE AND FILTERS

12.1 NEED FOR FILTERING

If a varying signal from a biological source, such as an electrocardiogram or blood pressure wave, is analysed suitably, it will be found to contain a range of frequencies, some present in greater amplitudes than others. For research purposes, all of these must be amplified and reproduced on a recorder or oscilloscope, accurately and without changes in relative phases one to another; for clinical purposes, some components of the signal may have little or no diagnostic significance, and may be removed. Further, a desired biological signal will often be accompanied by other undesired signals, or by extraneous noise such as alternating voltage picked up from power supply lines, or thermal or other noise generated in the resistance of the patient or during amplification of the signal. If these undesired signals are at frequencies outside the range of the desired signals, simple methods of *filtering* the desired signals may be sought. For all these purposes, a knowledge of the *frequency response* characteristics of circuits and recording devices is of the utmost importance. While a detailed study of the topic is very involved, and lies outside the scope of this book, sufficient basic principles can easily be established to avoid the worst pitfalls.

12.2 FREQUENCY RESPONSE OF IDEAL RECORDING SYSTEM

An ideal recording system would amplify and record only those frequencies desired, and would completely reject all others; this can be represented graphically as in Fig. 12.1. The signal shown has frequency components 0·1 to 100 Hz. Accordingly, the recording system must transmit any frequency in this *pass band* with constant amplification, and should not transmit any other frequency at all. The recording system must also not change the phase of any component in the pass band, or the position of peaks and notches in the recording will shift relative to the main signal waves. Outside the pass band, phase changes do not matter, since no signal is transmitted anyway. In practice a less stringent phase response requirement is acceptable; if the phase of various frequency components within the pass band is shifted by the recording system, but *every component is shifted in phase by an amount directly proportional to its frequency*, the final recorded signal will be reproduced exactly, but with a time delay after the incoming signal. This time delay usually does not matter at all. Such *linear phase shift* systems are often used.

12.3 SINGLE RC LOW-PASS FILTER

It is usual to consider the upper and lower ends of the pass band separately. A *low-pass* (or *high-cut*) filter is one which determines the upper limit of the pass band. The simplest

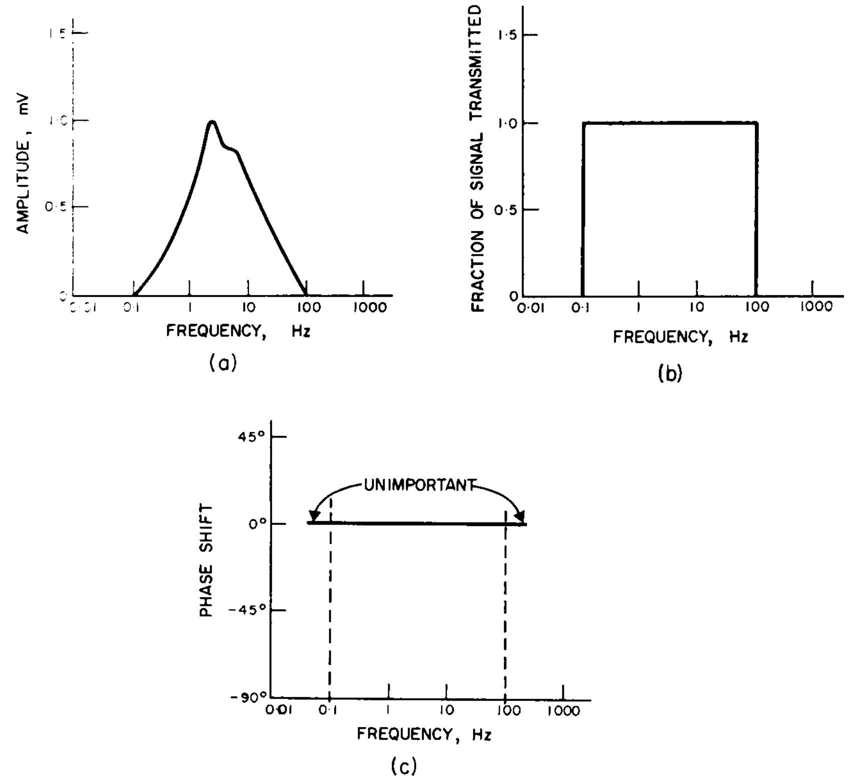


FIG. 12.1. Amplitude and phase response of ideal system. (a) Signal to be recorded. (b) Ideal amplitude response of system. (c) Ideal phase response of system.

and most fundamental type is the RC voltage divider network of Fig. 12.2. The behaviour of this filter can easily be analysed. The impedance of the capacitor at a frequency f is:

$$Z_c = \frac{1}{2\pi fC}$$

and the total impedance of the series combination of R and C is given by

$$Z_{\text{total}} = \sqrt{(R^2 + Z_c^2)}. \tag{12.1}$$

Since the network is merely a voltage divider,

$$e_{\text{out}} = \frac{Z_c}{Z_{\text{total}}} e_{\text{in}}. \tag{12.2}$$

The *voltage gain* G of the network (which will be less than 1) is given by

$$G = \frac{e_{\text{out}}}{e_{\text{in}}}$$

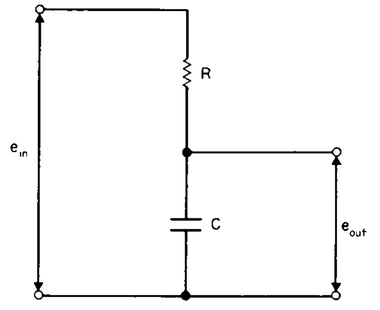


FIG. 12.2. Simple low-pass filter.

$$= \frac{Z_c}{Z_{\text{total}}} \qquad \text{(from eqn. (12.2))}$$

$$= \frac{Z_c}{\sqrt{(R^2 + Z_c^2)}} \qquad \text{(from eqn. (12.1))}$$

$$= \frac{1}{\sqrt{[(R^2 + Z_c^2)/Z_c^2]}}$$

$$= \frac{1}{\sqrt{[(R^2/Z_c^2) + 1]}} \qquad (12.3)$$

Now since $Z_c = 1/2\pi fC$, at low frequencies it is very large, and falls as f rises. At some frequency, which we will call f_0, Z_c will become equal to R. At this frequency

$$G = \frac{1}{\sqrt{(1 + 1)}}$$

$$= \frac{1}{\sqrt{2}}$$

$$= 0{\cdot}707. \qquad (12.4)$$

This frequency is known as the *cut-off frequency* of the filter. By definition, at f_0

$$R = Z_c$$

$$= \frac{1}{2\pi f_0 C},$$

so

$$f_0 = \frac{1}{2\pi RC}. \qquad (12.5)$$

What happens to the *phase* of e_{out} as the frequency changes? The *current* through the series combination leads e_{in} by the angle a, which as we have seen in Chapter 6, is given by

$$a = \text{artan}\,\frac{Z_c}{R}.$$

This current flows through C, and so it must lead e_{out} by 90°; therefore e_{out} lags behind e_{in} by (90° − a). By drawing a right angled triangle, as in Fig. 12.3, it can be seen that the angle b is equal to (90° − a). Accordingly,

$$b = \text{artan}\,\frac{R}{Z_c}. \tag{12.6}$$

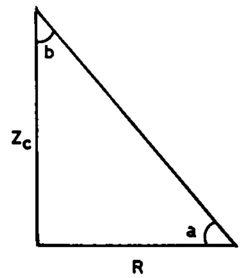


FIG. 12.3. Phase of e_{out}.

We can now graph both the gain and phase shift of this filter as functions of frequency, and compare them with the ideal characteristics of Fig. 12.2; this is done in Fig. 12.4. It will be seen that if f_0 is taken as the upper limit of the pass band, this filter gives quite a good approximation to a linear phase shift within the pass band. Its amplitude characteristic, however, leaves much to be desired; it falls off quite slowly above f_0, allowing much undesired material with frequencies above f_0 to pass.

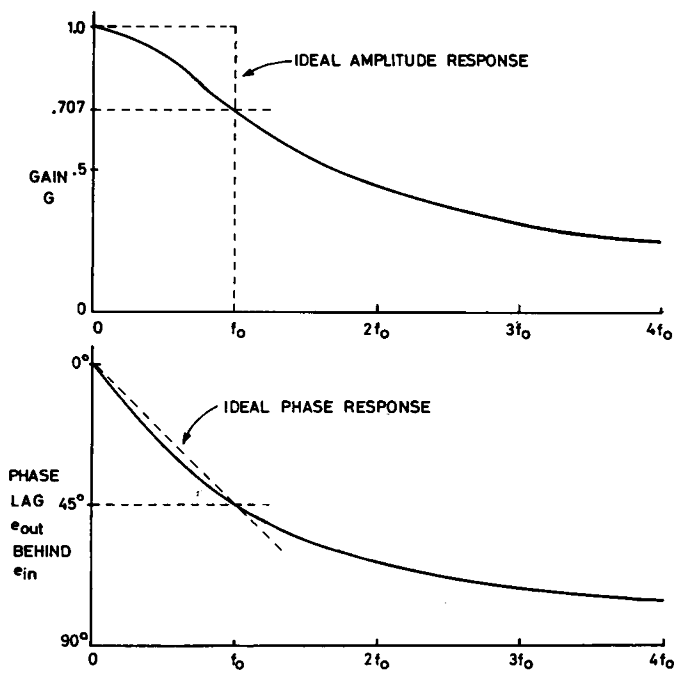


FIG. 12.4. Gain and phase shift of single RC low-pass filter.

12.4 Practical low-pass filter

In modern practice *active* filters are commonly used; these usually incorporate an operational amplifier. A low-pass filter design of high quality suitable for biological work is shown in Fig. 12.5. In this filter, which is based on a unity-gain non-inverting amplifier

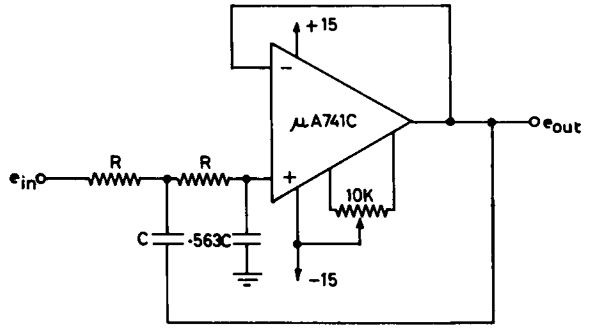


Fig. 12.5. Low pass unity gain filter.

the values of R and C are selected on a Wheatstone bridge to 1 % or better accuracy:

$$f_0 = \frac{0{\cdot}212 \times 10^6}{RC}, \tag{12.7}$$

where C is in pF, R in kilohm, and f_0 in kHz. The design is suitable for frequencies up to 100 kHz; R should not exceed 100 K. In this filter, G at f_0 is 2/3.

The characteristics of this filter are shown in Fig. 12.6; its advantages over the simple design are obvious.

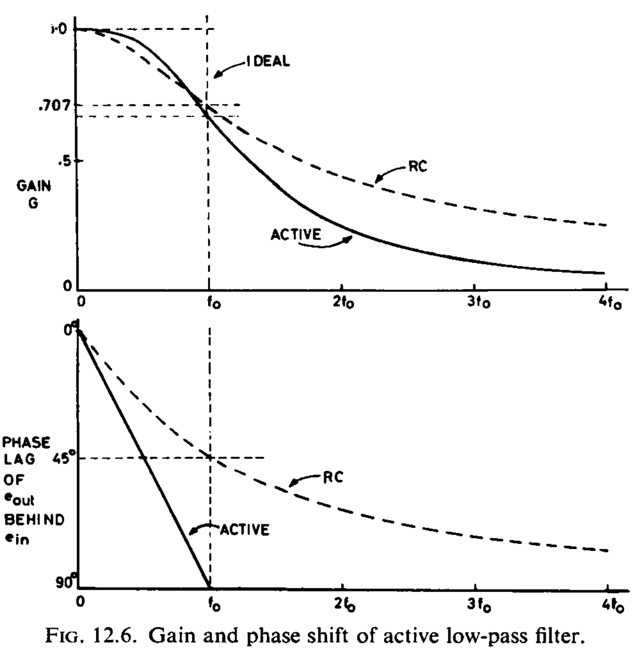


Fig. 12.6. Gain and phase shift of active low-pass filter.

12.5 Single RC high-pass filter

A *high-pass* (or *low-cut*) filter is one which determines the lower limit of the pass band. The simplest type is the RC voltage divider network of Fig. 12.7. Following the same procedure used in §12.3 above for the low-pass filter, the voltage gain G is found to be

$$G = \frac{1}{\sqrt{[(Z_c^2/R^2) + 1]}} \tag{12.8}$$

The cut-off frequency f_0 is again

$$f_0 = \frac{1}{2\pi RC}, \tag{12.9}$$

and the gain at f_0 is again 0·707.

The *current* through the series combination, as before, leads e_{in} by the angle a. This current flows through R, and so e_{out} must be in phase with it; therefore e_{out} leads e_{in} by a:

$$a = \operatorname{artan} \frac{Z_c}{R}. \tag{12.10}$$

If the gain and phase shift of this filter are graphed as functions of frequency, its properties are seen to be the converse of those of the low-pass filter.

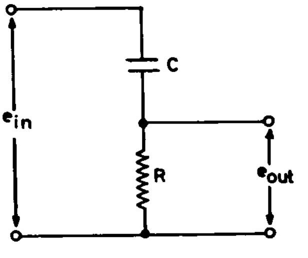


Fig. 12.7. Simple high-pass filter.

12.6 Practical high-pass filter

The active high-pass unity gain filter corresponding to the low-pass design shown previously is shown in Fig. 12.8. The cut-off frequency for this filter is

$$f_0 = \frac{0{\cdot}119 \times 10^6}{RC} \tag{12.11}$$

where C is in pF, R in kilohm, f_0 in kHz. R should not exceed 100 K. The gain at f_0 is again 2/3.

The characteristics of this filter are the converse of those shown in Fig. 12.6.

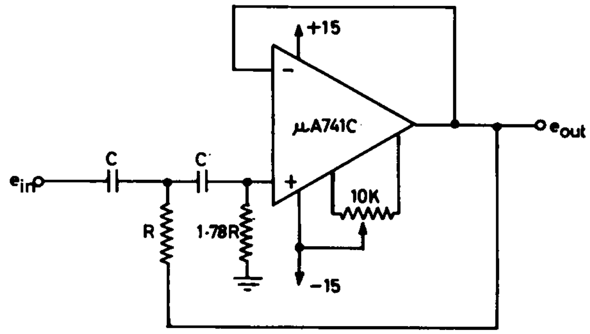


FIG. 12.8. High pass unity gain filter.

12.7 DECIBELS

There are a number of conveniences arising from the expression of amplifier and filter gains as the logarithm of the ratio rather than as the ratio itself; the main one is that the overall gain of a system can be obtained by addition rather than multiplication. The gain in *decibels*, A, is defined as

$$A = 20 \log G. \tag{12.12}$$

For the simple high or low-pass filter above, at f_0, $G = \frac{1}{\sqrt{2}}$, so

$$\begin{aligned} A &= 20 \log \frac{1}{\sqrt{2}} \\ &= -\ 20 \log \frac{1}{\sqrt{2}} \\ &\doteqdot -\ 10 \log 2 \\ &= -\ 3 \text{ db.} \end{aligned}$$

f_0 is often referred to as the 3 *db point* in a filter of this type.

If the gain curve of Fig. 12.6 is replotted, using decibels on the vertical axis and a logarithmic scale of frequency on the horizontal, an interesting result occurs, as shown in Fig. 12.9. It will be seen that the output of the RC filter outside the passband falls off at a rate which soon becomes 6 db per octave (per twofold change in frequency), or 20 db per tenfold change; the active filter falls off at *twice* this rate.

Fig. 12.10 shows the corresponding graph for the high pass RC and active filters; the curves are seen to be symmetrical with those of Fig. 12.9.

12.8 TRANSIENT RESPONSE OF SIMPLE LOW-PASS FILTER

A somewhat different approach to describing the behaviour of a filter or of a whole recording system is in terms of its output when an abrupt change in input is applied; this is known as its *transient response*. It can be shown that this response measures the

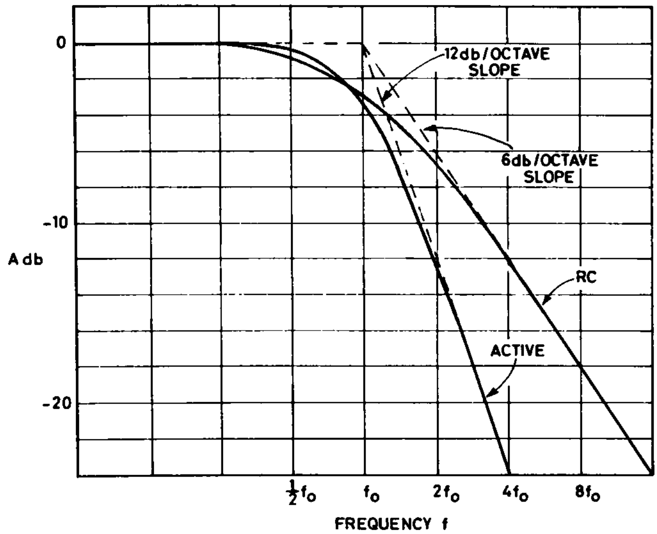


FIG. 12.9. Low-pass filter curves on db-log frequency scale.

same qualities as does a measurement of amplitude and phase over a wide range of frequencies. The transient measurement is much simpler to make, but the results are sometimes harder to interpret.

If an abrupt 1 V step is applied to the input of the simple low-pass filter of Fig. 12.11, the capacitor will charge exponentially towards 1 V with a time constant of *RC*.

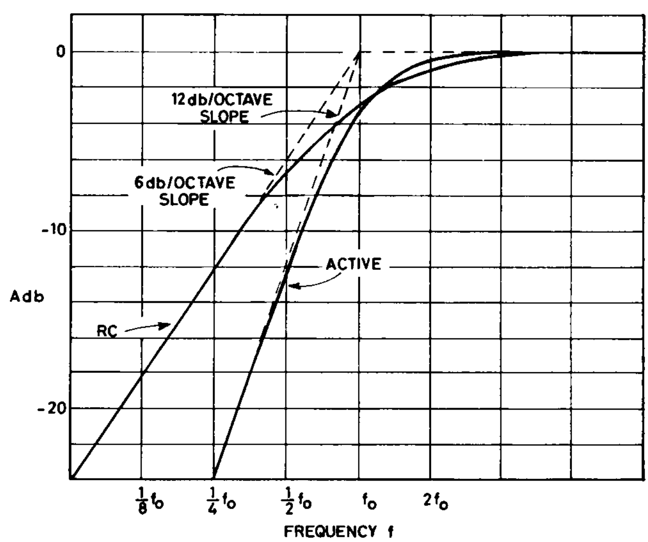


FIG. 12.10. High-pass filter curves on db-log frequency scale.

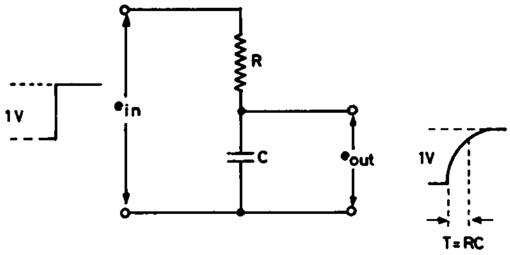


FIG. 12.11. Transient response of low-pass filter.

The *rise time* of a filter is defined arbitrarily as the time taken for the output to rise from 10 to 90% of its final value; for the simple low-pass filter this time will be 2·2 time constants.

The initial portion of the rise is approximately linear; over this region the output closely approaches the integral of the input, and the filter is sometimes described loosely as an integrating network. The relationships between f_0, the rise time, and the time constant are summarised in Fig. 12.12.

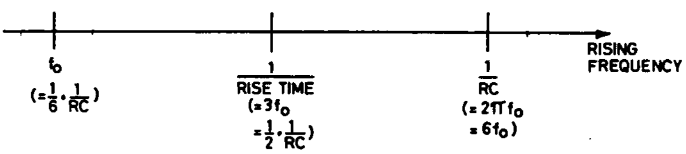


FIG. 12.12. Relationship between cut-off frequency, rise time, and time constant of a simple low-pass filter.

12.9 Transient response of simple high-pass filter

The transient response of a high-pass filter can be readily deduced. If an abrupt 1 V step is applied to the input of Fig. 12.13, the capacitor will initially remain uncharged, and the full 1 V will appear across R and the output. As the capacitor charges exponentially, it will take more and more of the 1 V, leaving less and less across the input. The output will fall to zero exponentially with a time constant of RC.

It will be clear that if only a short rectangular pulse is applied to the filter, the output voltage will not have fallen much by the time the pulse terminates, and the result will be as in Fig. 12.14.

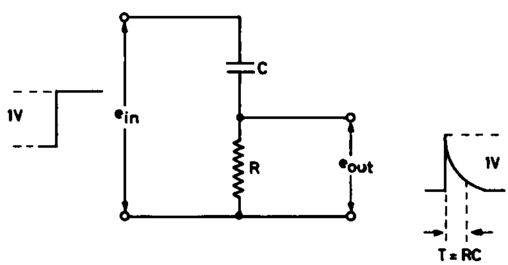


FIG. 12.13. Transient response of high-pass filter.

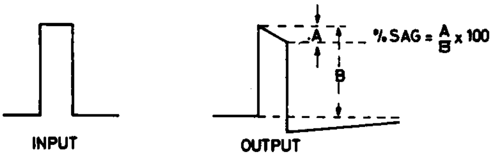


FIG. 12.14. Output of high-pass filter with time constant long compared with applied pulse.

The distortion produced is generally expressed as *percentage sag* in the top of the pulse. The initial part of the decay is approximately linear, and if it continued at this rate would fall to zero in one time constant. The approximate rule holds that the percentage sag is equal to the percentage of a time constant occupied by the pulse. For example, if the time constant is 1 sec, a pulse of 0·01 sec duration would have 1% sag. Conversely, if a pulse of 100 msec is to be passed with not more than 5% sag by a certain high-pass filter, the filter time constant must be at least twenty times 100 msec, or 2 sec.

If the filter time constant is made very short compared with the pulse duration, a rectangular pulse will be reproduced as in Fig. 12.15.

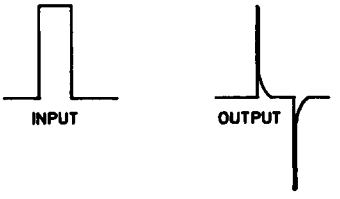


FIG. 12.15. Differentiation of pulse.

The output is an approximation to the rate of change or differential of the input, and the filter is sometimes loosely described as a differentiating circuit.

The relationship between f_0, the time constant, and the time for 5% sag can be summarised by Fig. 12.16.

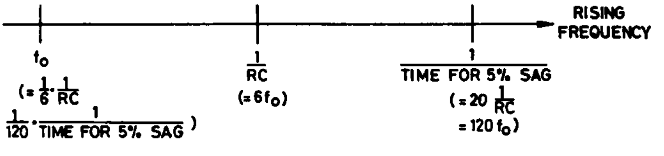


FIG. 12.16. Relations between cut-off frequency, 5% sag time, and time constant for a simple high-pass filter.

12.10 Application of filter concept

Filters may be used deliberately, to separate a desired signal from either undesired signals or extraneous noise lying outside the pass band of the desired signal. For example, a high-pass filter may be used to remove slow skin potential changes from an electrocardiogram, or a low-pass filter may be used to remove muscle potentials from the same electrocardiogram. As will be seen in Chapter 27, the use of filtering is absolutely essential when attempting to extract signals from noise by averaging the results of repeated experiments.

Apart from their deliberate use, low-pass filters inevitably occur in amplifiers, as a result of stray capacitances. For example, in the transistor amplifier of Fig. 12.17, the stray capacitance C, due to wiring capacitance to earth and the internal capacitance of

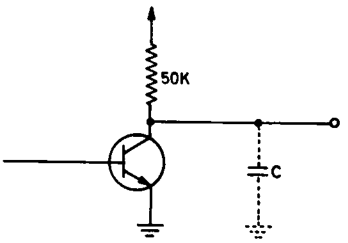


FIG. 12.17. Stray capacitance in transistor amplifier.

the transistor, may well amount to 10 pF. In conjunction with the transistor collector load of 50 K, this gives a filter whose cut-off frequency is

$$f_0 = \frac{1}{2\pi \times 10^4 \times 5 \times 10^{-11}}$$

$$= 318 \times 10^3 \text{ Hz}.$$

Above this frequency amplification will fall off rapidly.

A similar situation arises in mechanical recording systems, such as heat or ink-writing recorders, in which the mechanical components form a low-pass filter. In this case the cut-off frequency usually lies between 50 and 100 Hz.

High-pass filters are frequently included in transistor amplifiers to remove effects of slow voltage changes due to temperature effects on the transistors; care must be taken to avoid removing clinically significant components of a biological signal at the same time.

12.11 FREQUENCY RANGE OF BIOELECTRIC SIGNALS

Table 12.1 shows the frequency range required for accurate reproduction of various types of bioelectric signal. For any particular type of signal, the use of apparatus with a range in excess of this will merely increase the spurious noise level. Some reduction in the

TABLE 12.1. FREQUENCY RANGE OF BIOLOGICAL SIGNALS

Type of potential	Upper limit		Lower limit	
	3db point	Rise time	3db point	Time for 5% sag
EMG, nerve action potential	6 kHz	60 μsec	10 Hz	800 μsec
Phonocardiography	1 kHz	350 μsec	0·5 Hz	16 msec
Pulse pressure transducer	50 Hz	7 msec	0·2 Hz	40 msec
ECG	150 Hz	2·3 msec	0·03 Hz	280 msec
EEG	50 Hz	7 msec	0·05 Hz	160 msec

range given in the table can often be tolerated, particularly when the apparatus is used for human diagnostic or monitoring purposes, and the choice lies between obtaining a degraded but clinically useful signal, or one obscured by noise. The ultimate test is, of course, whether the wave-form being examined is degraded in a clinically misleading fashion. Biological amplifiers are frequently provided with a switched selection of high- and low-pass filters to limit the frequency range to that required. The circuit of a modern biological amplifier intended to deliver an output of 1 V, and to be used after a suitable preamplifier, is shown in Fig. 12.18. In this amplifier, the first stage is a voltage follower, to prevent changes in the resistance of the input voltage divider affecting the high-cut filter; it acts as a *buffer* stage. An LH0042C is used, so that changes in input resistance will not produce output voltage changes due to bias current.

The two filter stages are also shown in Figs. 12.5 and 12.8; the capacitors are mounted on two-bank eleven position rotary switches. The second stage uses an LH0042C to avoid a change in output voltage caused by bias current, when the low cut is switched to d.c.

The two amplifier stages are conventional inverting stages; two are used, each with a gain of 10, to make the whole amplifier unit non-inverting; moreover, one stage with a gain of 100 would not give the desired frequency response.

The unit is intended to accept signals in the range 10 mV to 1 V peak-to-peak from a biological preamplifier, and to deliver 1 V peak-to-peak output for subsequent observation or storage. With the exception of the three selector switches and their associated components, the whole amplifier may be assembled on a simple printed board.

FURTHER READING

Burr-Brown Research Corp., Operational Amplifier Active RC Networks, 1966.

TOBEY, GRAEME and HUELSMAN, *Operational Amplifiers*, McGraw-Hill, New York, 1971.

There is a vast quantity of literature on filters; most of it has been rendered obsolete by the introduction of active filters using operational amplifiers.

PRACTICAL

12.1 Set up a simple RC low-pass filter, using $R = 33$ K, $C = 0{\cdot}01$ μF. At what frequency is Z_c equal to R? Drive the filter from a signal generator, and use an oscilloscope to measure its frequency response and phase shift in the range 10 Hz to 10 kHz. Plot these characteristics (i) on linear graph paper, (ii) on 3-decade semi-logarithmic paper, first converting gain to decibels. Compare the curves with the theoretical curves given in this chapter.

12.2 Repeat experiment 12.1, setting up the components as a high-pass filter.

12.3 Using a square-wave generator and oscilloscope, measure the rise time of the low-pass filter of 12.1 above, and the percentage sag in a known time of the high-pass filter of 12.2 above. From your results, calculate f_0 in each case.

12.4 Test a commercial electrocardiograph for (a) frequency response, and (b) rise time and sag. (The voltage calibrator constructed in Chapter 1 is ideal for (b)). Compare your results with the figures given in Table 12.1. If you are familiar with clinical electrocardiography, consider in what way an ECG wave-form would be unacceptably distorted by an amplifier with (a) inadequate high-frequency response, (b) inadequate low-frequency response.

12.5 Construct an active low-pass filter using the circuit of Fig. 12.5, and measure its response with a signal generator and oscilloscope.

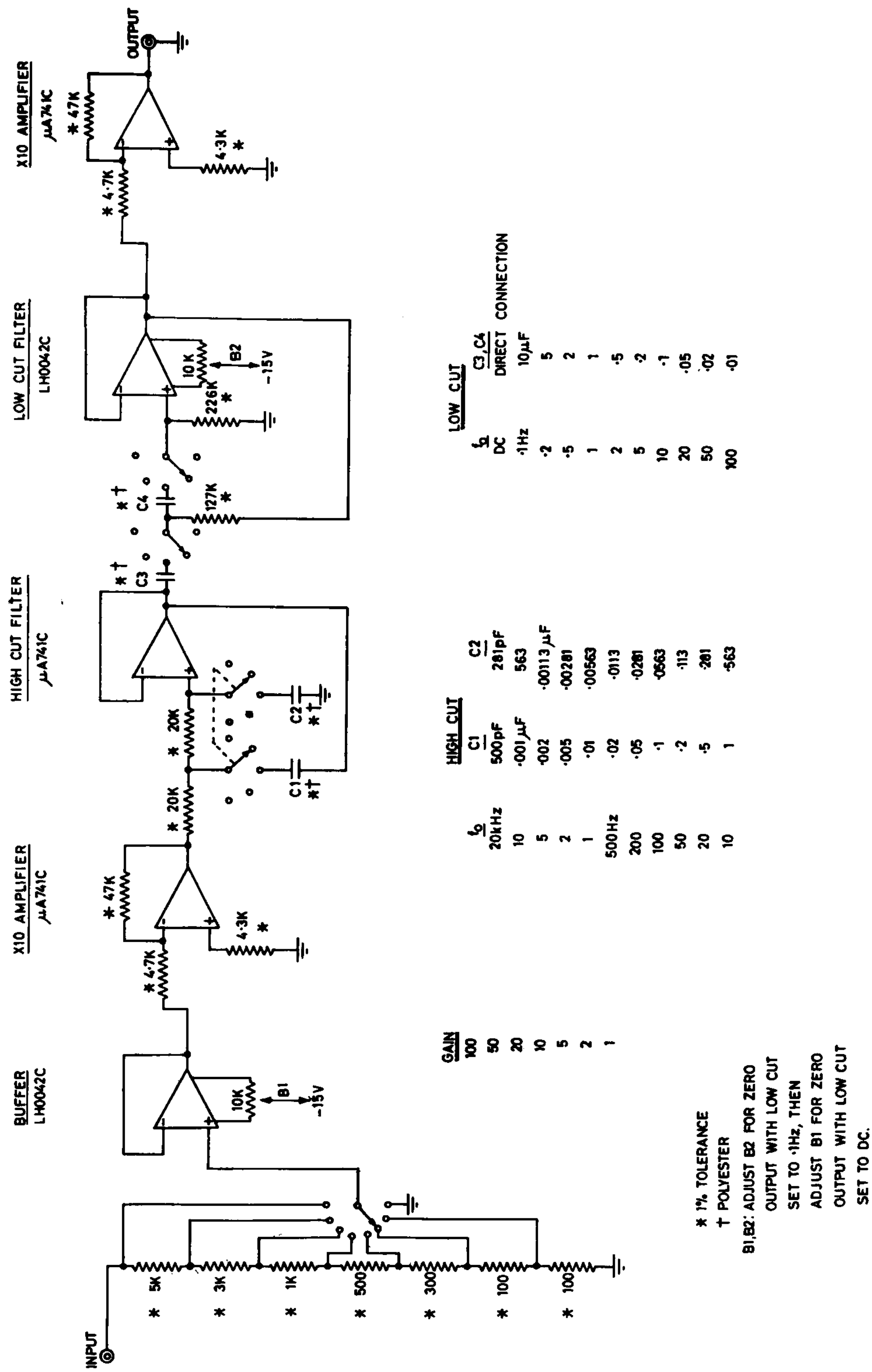


FIG. 12.18. Biological amplifier.

CHAPTER 13

THE CATHODE RAY TUBE

13.1 The cathode ray oscilloscope

The cathode ray oscilloscope is one of the most valuable and commonly used pieces of measuring equipment. In modern practice it is the central feature of the instrument development or repair workshop, and is common in the research laboratory and the operating theatre. Although it would be hopelessly uneconomic to attempt the construction of one's own oscilloscope, it is of considerable importance to understand the central features of the instrument, and the degree of trust which may be placed on the readings obtained from it. For this reason, this and the following five chapters will be directed to the construction of a small laboratory oscilloscope. In passing, a number of circuit concepts already discussed will be shown in practice, and a number of new concepts introduced; these have much wider applications than those shown here.

13.2 Commercial cathode ray oscilloscopes

It should be stated at the outset that a good oscilloscope is expensive. What is being purchased is not so much an assemblage of components, as the calibrations on the front panel, and the knowledge that these calibrations can be trusted, month after month and year after year. Every 2 or 3 years the oscilloscope should be serviced and recalibrated by the manufacturer or his agent; if this is done the calibrations in a good instrument should be accurate to $\pm 2\%$.

13.3 The cathode ray tube

The heart of the oscilloscope is the cathode ray tube; a typical tube is shown in section in Fig. 13.1. The tube consists of two parts. (i) The *electron gun* assembly produces a thin beam of electrons; these pass down the tube through the vacuum within the glass envelope, and strike the screen at the end. This screen is coated with a fluorescent material, which glows at the point where the beam strikes it. (ii) The *deflector plates* are arranged in two pairs, and the beam passes between the pairs in turn on its way to the screen.

The electron gun comprises several parts. The source of the electron stream is the *cathode*; this is a small hollow nickel cylinder coated with a mixture of rare earth oxides. Within the cathode, and insulated from it, is the *filament*, a loop of resistance wire through which a current is passed to bring the whole cathode assembly to a dull red heat. When this occurs, the oxide coating emits a cloud of free electrons into the surrounding vacuum. Although individual electrons are continually being emitted and

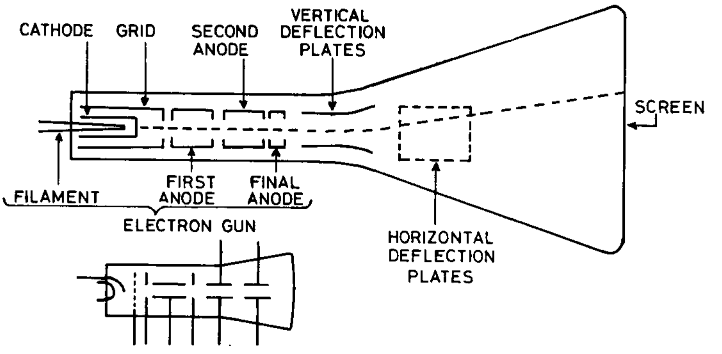


FIG. 13.1. Typical CRT structure and symbol.

falling back into the cathode, a supply is always maintained in the space about the cathode. The cathode structure is surrounded by a larger metal cylinder, with a small hole in the centre of its cap; this is known as the *grid.* Beyond the grid is a succession of *anodes*; either two or three may be used, and a three-anode tube is shown in Fig. 13.1. Each anode is a hollow cylinder with an axial hole through it. In use the first anode is made a thousand or more volts positive to the cathode, and this attracts electrons from the cloud about the cathode; they pass out through the hole in the end of the grid in a thin stream, at considerable velocity. The diameter of this stream may be controlled by placing a *negative* voltage on the grid; this "pinches off" the electron flow, and if sufficient (about 40 V negative to the cathode) will stop the flow completely. The stream of electrons leaving the grid has sufficient velocity to carry it right through the holes in the first anode, out the far side, and through the second anode also. The second anode is made less positive to the cathode than the first anode; this is insufficient to retard electron beam much, but it does form an electrostatic "lens" between the second and first anodes, causing the electron stream to converge. By altering the second anode voltage, the focal length of this lens can be varied, and the electron beam brought to a point focus on the fluorescent screen. The final anode again is at a high voltage; and the electrons are accelerated further, so that they leave it in the form of a thin high-velocity pencil. A two-anode tube omits the second anode, and focusing is done between first and final anodes.

The beam then passes between the first pair of deflection plates. If an e.m.f. is applied between these, the beam is repelled from the more negative and attracted to the more positive one, so that its course is changed in proportion to the e.m.f. applied, and the spot it produces on the screen is moved along a line at right angles to the pair of plates. The beam then passes between the second pair, which are at right angles to the first pair; this enables the spot to be deflected along a line on the screen at right angles to the first one. By suitable combinations of voltages on the two pairs of plates (known as the *vertical* or *Y plates*, and the *horizontal* or *X plates*) the spot can be located anywhere on the fluorescent screen.

The CRT *must be handled carefully.* If the glass envelope is scratched or strained, knocked sharply or dropped, a quite nasty implosion may result, with the accompaniment of a lot of very sharp flying fragments of glass.

The fluorescent material used to coat the CRT screen is known as the *phosphor*. Various phosphors are used, depending on the purpose for which the tube is required. The commonest are:

P1, P31 Green fluorescence, persisting for about 15 milliseconds after the beam is removed. Used for visual observation of medium speed phenomena.

P2 Green fluorescence, with a fainter afterglow persisting for about 2 seconds. Used for visual observation of both fast and slow speed phenomena.

P7 Blue fluorescence with brilliant yellow afterglow, persisting for about 10 seconds. Used for visual observation of slow phenomena, and very common in patient monitoring equipment. This is an excellent phosphor, but is *permanently damaged* if the spot is allowed to remain in one place for any length of time.

P11 Blue trace, persisting only for about 5 microseconds. Highly actinic, and used for photographic recording.

None of these phosphors is ideal for the display and study of biological signals, unless photography is used. In modern practice, two methods of retaining a sample of a biological signal on an oscilloscope screen are available. In the first, a *storage* tube is used. These tubes contain a charge storage grid, located behind the screen, and one or more *flood guns* to cover the screen with a uniform electron beam. The signal to be stored is written on the screen once, in the manner described above; the pattern is then retained for periods of an hour or more by the action of the flood guns. It may be erased as required by turning the flood guns off briefly. Storage tubes are quite effective, but are rather expensive, and have a rather limited life.

The second method is to store the sample of biological signal in a *computer memory*, and then to replay it rapidly over and over again on an ordinary oscilloscope, giving the illusion of a stored pattern. This method is rapidly coming into favour as computer memories become cheaper and cheaper.

13.4 CRT VOLTAGE SUPPLIES

The various voltages required to operate the CRT are normally derived from a single power supply by means of a voltage divider. It is convenient, and customary, to operate the CRT with the deflection plates near earth, and driven by suitable amplifiers. This means that the final anode is also near earth, while the grid, cathode, and filament are at a considerable *negative* voltage to earth; this voltage ranges from about 1000 V for a 3-inch diameter CRT up to many thousand volts for larger tubes. The current required is that carried by the electron beam, and is quite small, of the order of 100 μA or less. This *e.h.t.* (*Extra High Tension*) supply can be derived directly from a transformer and rectifier, or may be obtained by first generating a high frequency a.c., transforming this to a high voltage by a small transformer, and then rectifying the output. This saves considerable weight, and where high voltages are required removes the risk of receiving a lethal shock, since the filter capacitors can be much smaller than those required at supply line frequencies.

A typical circuit arrangement for a 3-inch 3-anode CRT is shown in Fig. 13.2. It will be seen that the first and final anodes are connected together within the tube, and are taken to an *astigmatism potentiometer*; they are typically set at about +75 V with respect to earth. The purpose of this adjustment will be explained later. The second anode is taken to the *focus potentiometer*, and is set at about −700 V with respect to earth; the cathode and filament are taken to −900 V, and the grid to the *intensity potentiometer*, at about −925 V. However, within the tube, the behaviour of the electron stream is governed by voltages *with respect to the cathode*. With respect to the cathode, the grid is at −25 V, the second anode at +200 V, and the first and final anodes at +975 V, so the voltages are as discussed at the beginning of this section. Connections to the four deflection plates will be discussed in Chapter 14; for the moment each may be connected to earth through a 100 K resistor.

Fig. 13.3 shows a typical power supply for a 3-inch oscilloscope. The power transformer is specially designed to produce the minimum of external alternating magnetic field, as this can easily distort the electron beam in the CRT; it has several secondary windings. The e.h.t. supply is a simple half-wave rectifier, using a high-voltage diode rectifier; the main filter capacitor is followed by a simple low-pass filter to further reduce

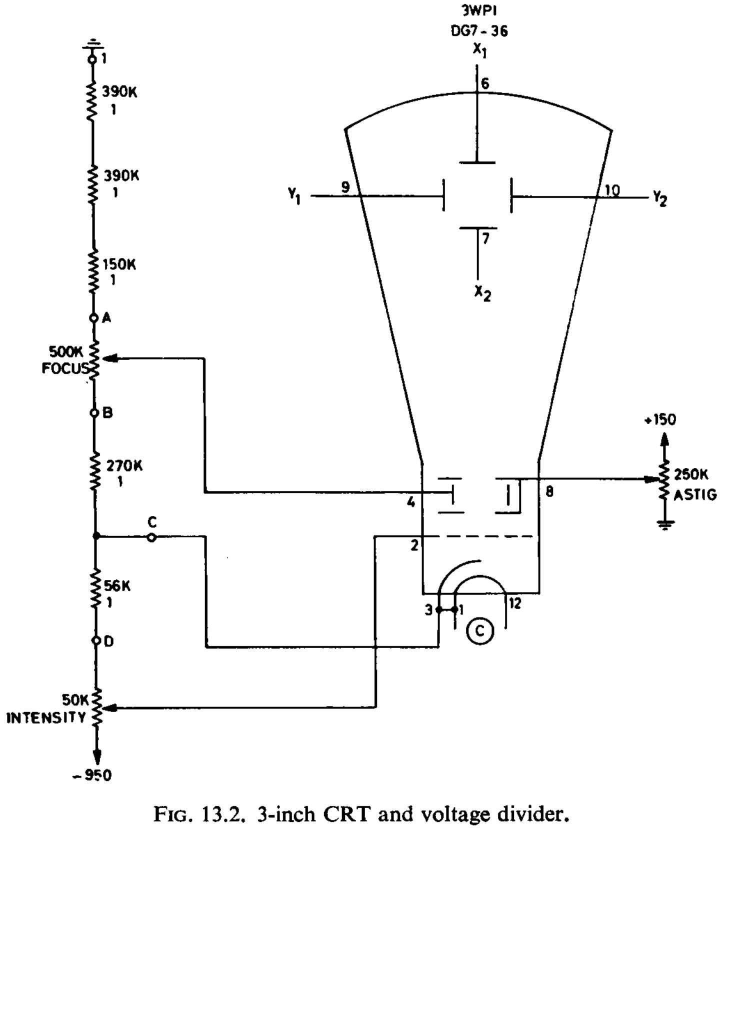


FIG. 13.2. 3-inch CRT and voltage divider.

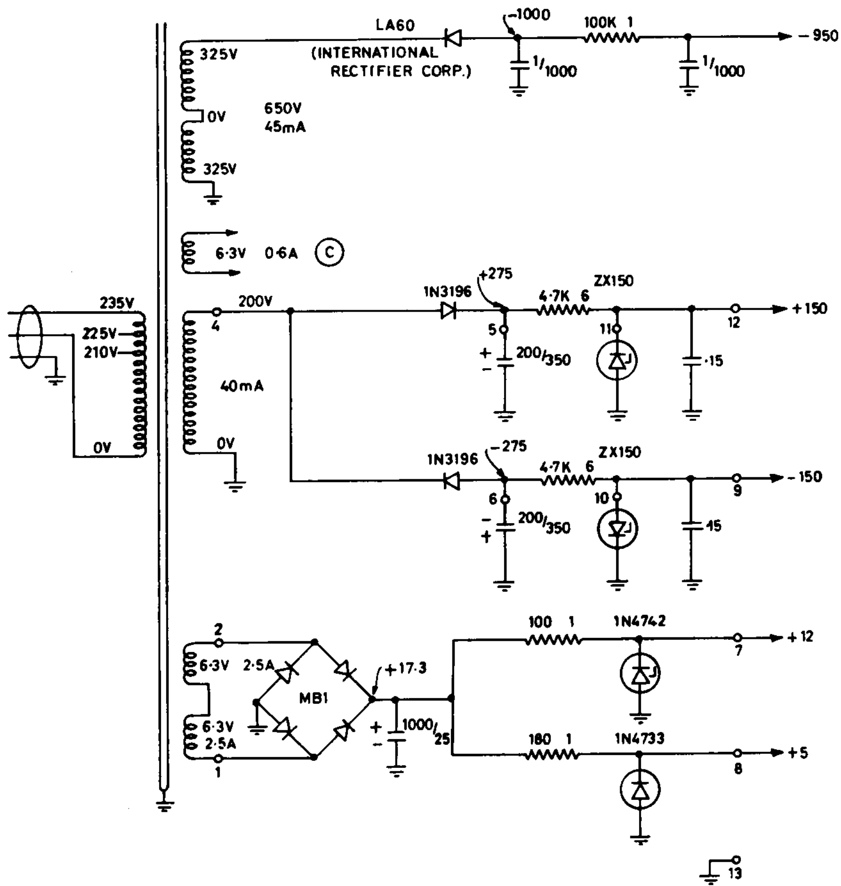


FIG. 13.3. Power supply for 3-inch CRO.

the ripple voltage. *This supply can deliver a very severe shock, and possibly a fatal one.* The filament of the CRT is heated by a.c. from a 6·3 V winding insulated to withstand the full e.h.t. supply. Further Zener-regulated d.c. supplies of +150 V, −150 V, +12 V and +5 V are shown; their use will be discussed in subsequent chapters.

13.5 Magnetic shielding

It is essential to reduce to a minimum any stray magnetic fields which may reach the CRT, since these will certainly distort the electron beam. It is usual to surround the CRT with a *magnetic shield,* consisting of a special steel alloy. The most commonly used material is known as mumetal. This is most effective, but cannot be cut, drilled, or even knocked sharply, without losing a good deal of its efficiency. (After fabrication, it is heat-treated in a reducing atmosphere.)

13.6 Deflection sensitivity of a CRT

With all four deflection plates connected to earth, the spot on the CRT screen should be central. It requires quite large voltages to move the spot to the margin of the screen in the X or Y direction; typical sensitivities are of the order of 1·5 V/mm of deflection. Since all tubes are designed to give a deflection accurately proportional to the applied voltage, it is possible to calibrate the screen in volts, in either the X or Y direction, and use it to measure either d.c. voltages, or the peak-to-peak value of a.c. voltages.

13.7 Magnetic deflection of CRT

Instead of using X and Y deflection plates, it is possible to deflect the beam by a magnetic field, using X and Y *coils*; in this case deflection is proportional to the current through the coils. Due to the considerable inductance of any practical deflection coils, this method is suitable for use only at low frequencies. It is used in some cathode ray electrocardiographs, and in all television and radar equipment.

FURTHER READING

Arthur, *Power Supply Circuits*, Tektronix Inc., Beaverton, Oregon, 1967.
Arthur, *Cathode Ray Tubes*, Tektronix Inc., Beaverton, Oregon, 1967.

PRACTICAL

Front and side views of a 3-inch oscilloscope suitable for class construction are shown in Figs. 13.4, 13.5, and 13.6. Progressive stages in its assembly are discussed in the following chapters.

13.1 Wire the EHT power supply, CRT filament supply, CRT voltage divider and CRT in accordance with Figs. 13.2 and 1.33. Wire the under side of the circuit card sockets as in Fig. 13.9. Plug two cards made up as in Fig. 13.10 into the sockets marked "X output" and "Y output", and connect the X and Y plates of the CRT to them. For the time being, connect the first and final anodes to earth. Carefully check the wiring (the CRT can be destroyed by a false connection). Switch on the power supply; when the CRT has warmed up, adjust the focus and intensity controls until a fine focused spot is obtained.

N.B. *After switching off*, and before touching any part of the circuit, *discharge the EHT capacitors* with an insulated screwdriver.

13.2 Using known d.c. voltages applied to (a) each deflection plate separately (b) across each pair of plates, calibrate the screen for deflection sensitivity in both X and Y directions. Does it matter whether method (a) or method (b) is used? Is the deflection sensitivity the same in the X and the Y directions?

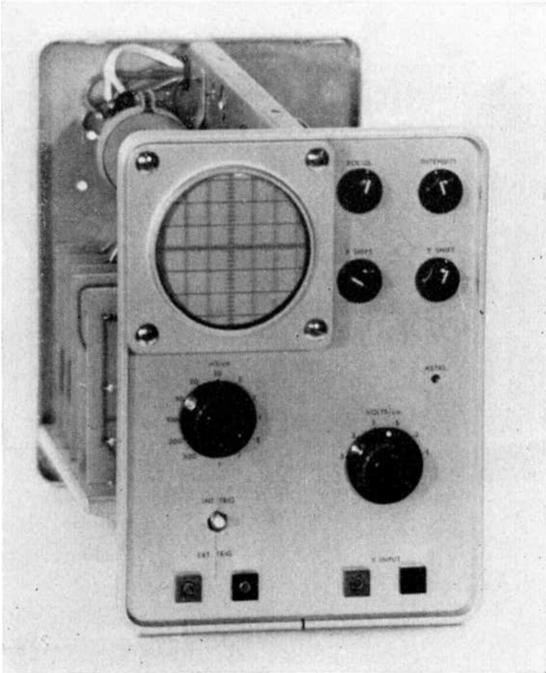


FIG. 13.4. Front view of oscilloscope.

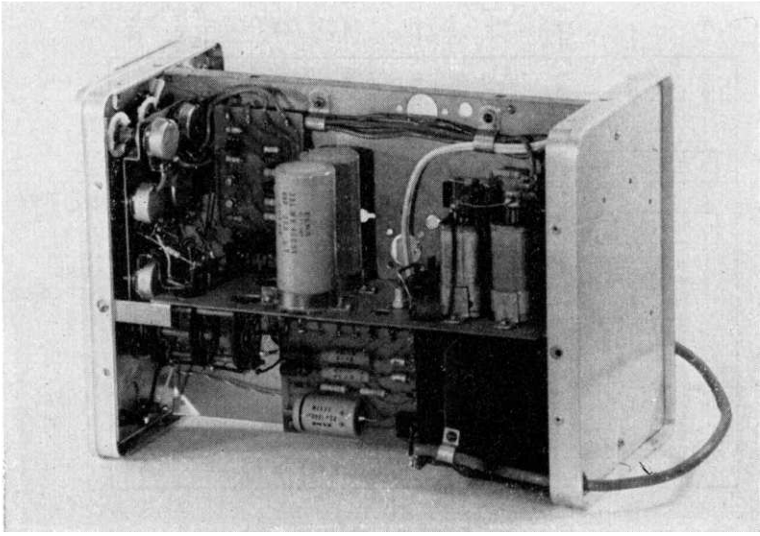


FIG. 13.5. Right side of oscilloscope.

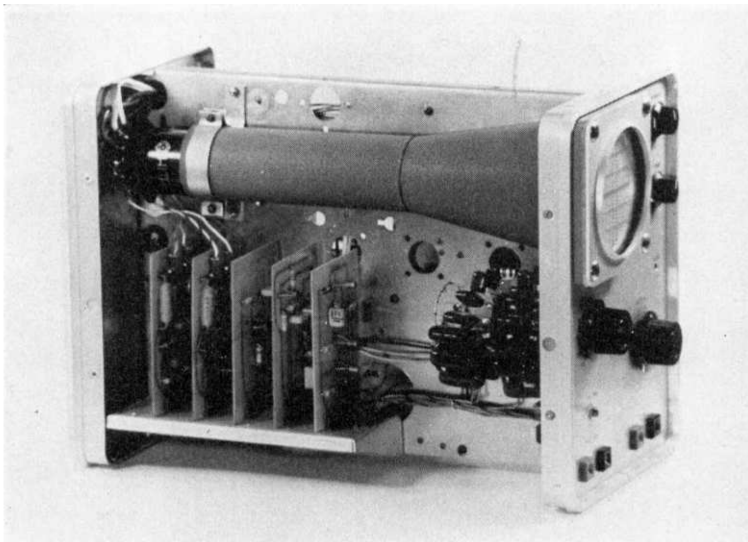


FIG. 13.6. Left side of oscilloscope.

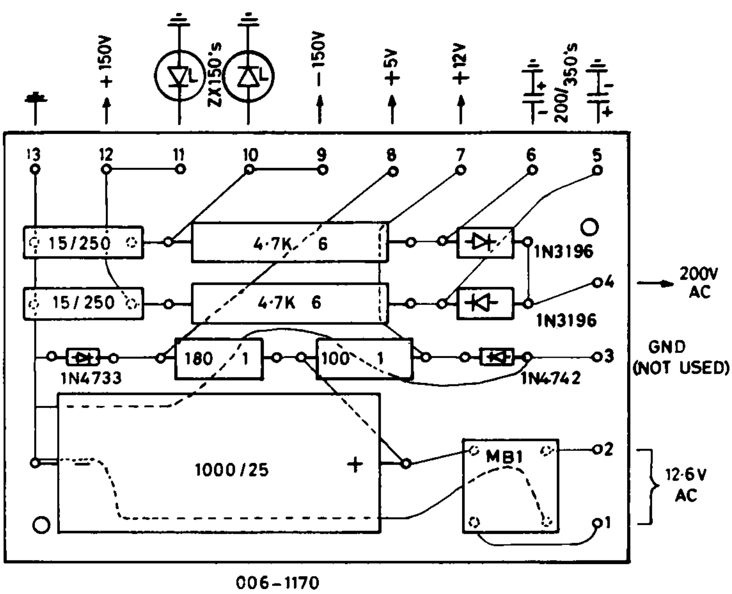


FIG. 13.7. Power supply card.

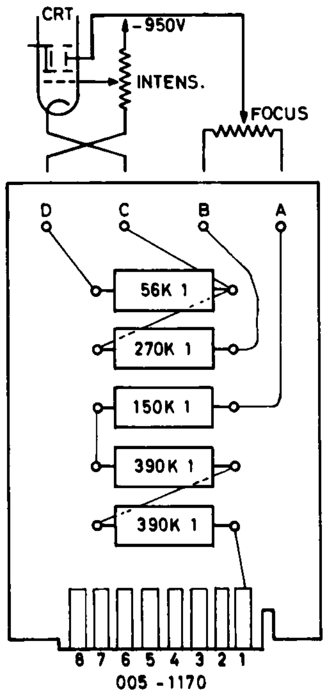


Fig. 13.8. EHT divider card.

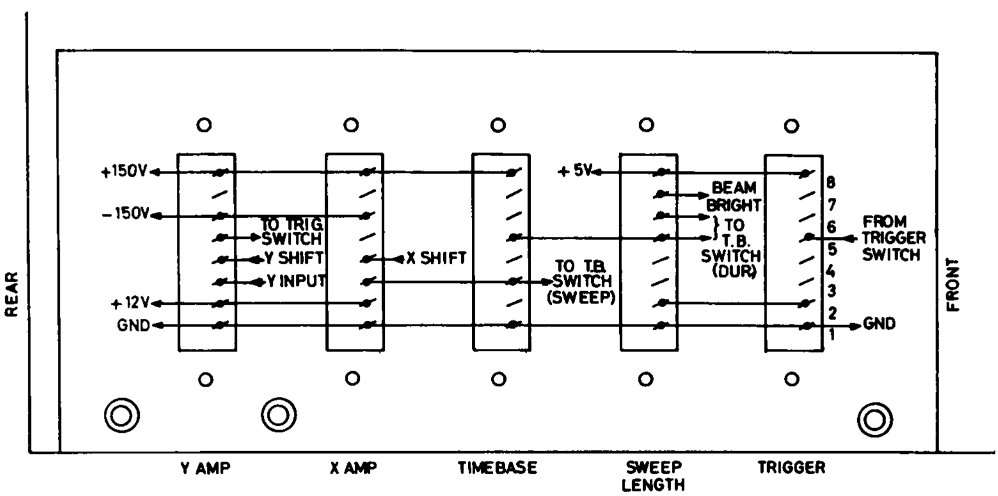


Fig. 13.9. Underside of circuit card sockets.

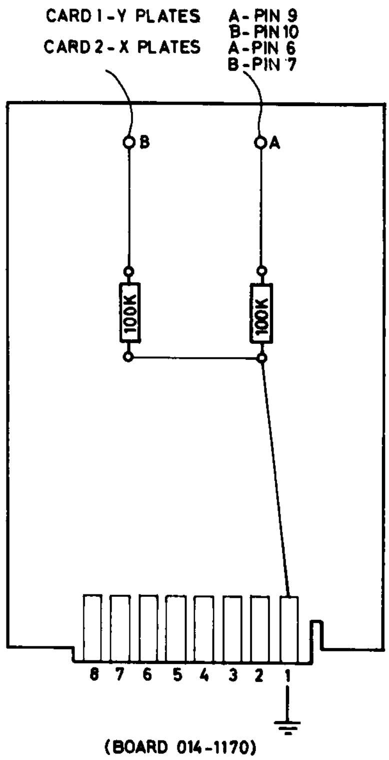


FIG. 13.10. Dummy X and Y driver cards.

CHAPTER 14

OSCILLOSCOPE AMPLIFIERS

14.1 Amplifier requirements for an oscilloscope

As was discussed in the previous chapter, the deflection plates in a CRT require considerable applied voltages to deflect the beam from one side of the CRT to the other; at 1·5 V/mm, a 3-inch tube will evidently require a 120 V change in potential between a pair of plates to utilise the whole screen. Since most voltages to be measured by the oscilloscope are much smaller than this, it will be necessary to employ amplifiers of accurately known and stable gain, and having a frequency response much better than the range of frequencies to be measured, between the signal source and the deflection plates. Operational amplifiers are suitable for this purpose, and are often used to obtain some amplification; however, no standard transistor operational amplifier can provide the 120 V of signal that is required to drive the deflection plates, and special amplifiers are needed. These vary in complexity; the simplest type, on which most output amplifiers are based, is the *balanced amplifier* (known in British terminology as the "long-tailed pair").

14.2 Principle of the balanced amplifier

The principle of the balanced amplifier may be understood by reference to Fig. 14.1. Assume initially that the two transistors Q1 and Q2 are identical, and the base of Q1 is connected to earth; the circuit is then completely symmetrical. The collector currents will be identical, and the two deflection plates will be at the same potential. What will the potential be? Since both bases are at earth, and point A is returned to -150 V through the "long tail", both base-to-emitter diodes are forward biased, and Q1 and Q2 will commence to draw collector current. These currents will return to the -150 V supply through the long tail, and will produce a voltage drop in it. Point A will thus rise in potential until the bases of Q1 and Q2 assume their usual base-to-emitter voltages of about 0·5 V: i.e., until A is at $-0{\cdot}5$ V to earth. This means that the voltage across the long tail is $(150 - 0{\cdot}5)$, and thus the current through it is by Ohm's law 149·4/20, or 7·475 mA. This arrangement is self-stabilising; should the collector currents tend to rise for any reason, the potential of point A will rise, and tend to cut off Q1 and Q2, thus reducing the variation. The converse will occur if the currents tend to fall.

Since the circuit is symmetrical, half the current in the long tail will come from Q1, and the other half from Q2; they will each be 3·74 mA. Assuming a β of 100 for Q1 and Q2, this implies a base current of 37·4 μA into each transistor, and a voltage drop of $3{\cdot}74 \times 20$ (say 75 V) across each collector load. The CRT deflection plates will then each be at $150 - 75$, or 75 V.

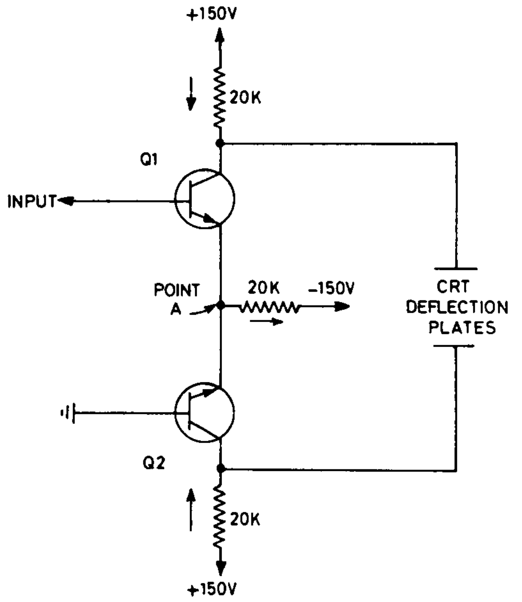


FIG. 14.1. Principle of balanced amplifier.

How accurate was our previous estimate of 0·5 V as the base-to-emitter voltage of Q1 and Q2? This can be determined from a graph of base-to-emitter voltage against base current for the transistor type used. Such a curve is shown in Fig. 14.2. For the estimated base current of 37·4 μA, the base-to-emitter voltage is in fact 0·50 V.

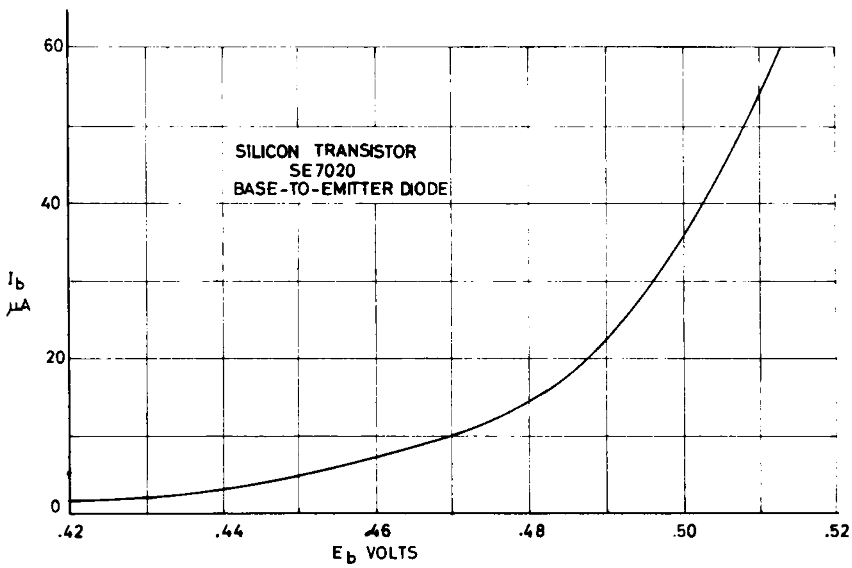


FIG. 14.2. Base current as a function of base voltage.

Now let us apply an input to the amplifier, raising the base current of Q1 somewhat. The collector current of Q1 will rise, causing the voltage drop in the collector load of Q1 to rise; the upper CRT deflection plate will thus become *less positive*. This increase in current will also pass through the long tail, and as we have seen, point A will commence to rise. This will at once *reduce* the base-to-emitter voltage of Q2; its base current will fall, its collector current will fall, the drop across its collector load will decrease, and the lower CRT deflection plate will become *more positive*. Let us assume that this last movement is from +75 to +100 V, and work backwards to deduce what must have happened at the other plate, and how much input current must have been supplied.

Since there is now 50 V across the collector load of Q2, its collector current must be 50/20, or 2·5 mA, and hence its base current must be 25 μA. From Fig. 14.2, its base-to-emitter voltage must now be 0·492 V; since its base is at earth, point A must be at −0·492 V. By Ohm's law, the current through the long tail is (150 − 0·492)/20, which is not significantly different from its initial value (7·475 mA). The current through Q1 must then be (7·475 − 2·5), or 4·975 mA. This produces in its collector load a drop of 4·975 × 20, or 99·5 V, so the upper deflection plate will be at +50·5 V (having moved from +75 V) Since the collector current of Q1 is 4·975 mA, its base current must be 49·75 μA. From Fig. 14·2, its base-to-emitter voltage must be 0·508 V; since point A is at −0·492 V, its input voltage must be +0·016 V.

We can best see what has happened by tabulating (Table 14.1). A rise of 0·016 V and 12·4 μA at the input has produced a practically symmetrical change in deflection plate voltage of 50 V, enough to produce a 3·3 cm shift in the spot on the fluorescent screen.

TABLE 14.1

	Input to Q1 V	Input to Q1 μA	Upper plate V	Lower plate V
Before application of signal	0	37·4	75	75
After application of signal	0·016	49·8	50·5	100
Change	0·016	12·4	−24·5	+25·0

14.3 PRACTICAL BALANCED AMPLIFIERS

The circuits of the two balanced amplifiers to be used in the oscilloscope under construction are shown in Fig. 14.3 (Y amplifier) and Fig. 14.4 (X amplifier). They are identical in form, but utilise slightly different component values. They are assembled on identical printed cards, as shown in Figs. 14.5 and 14.6.

The two circuits differ from that of Fig. 14.1 in that they have each a network of three resistors replacing the simple long tail. This is an expedient which allows the gain of such an amplifier to be reduced; it will be seen that if the resistor between the emitters is reduced to zero, the circuit reduces to that of Fig. 14.1. As the resistor is increased, the amplifier gain falls, but its constancy of gain and independence from variations in individual transistor characteristics is greatly improved. The X amplifier in fact requires an input of about 5 V to move the spot right across the screen, while the Y amplifier

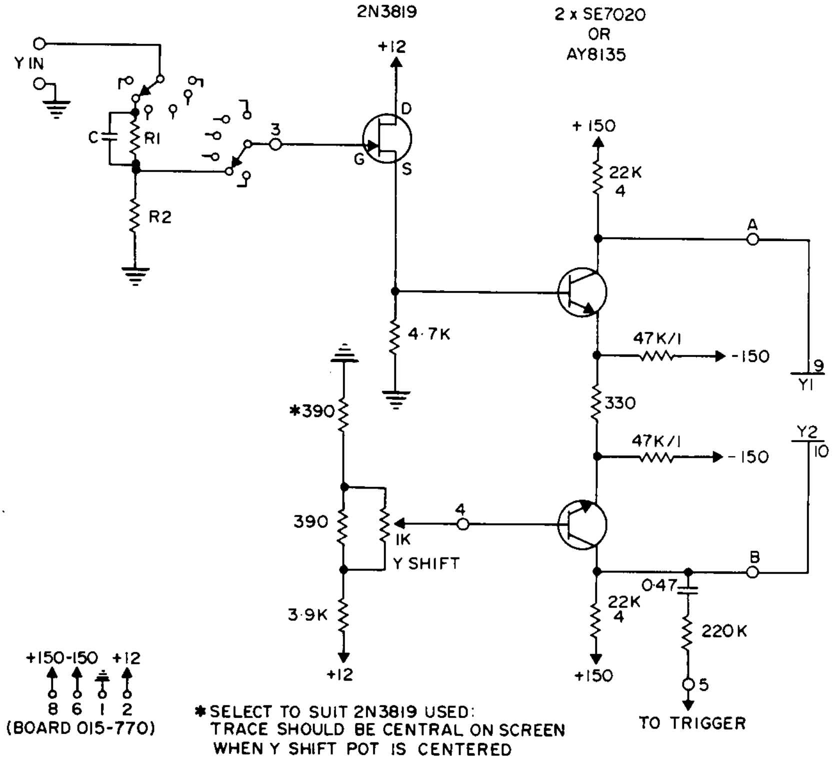


FIG. 14.3. *Y* amplifier.

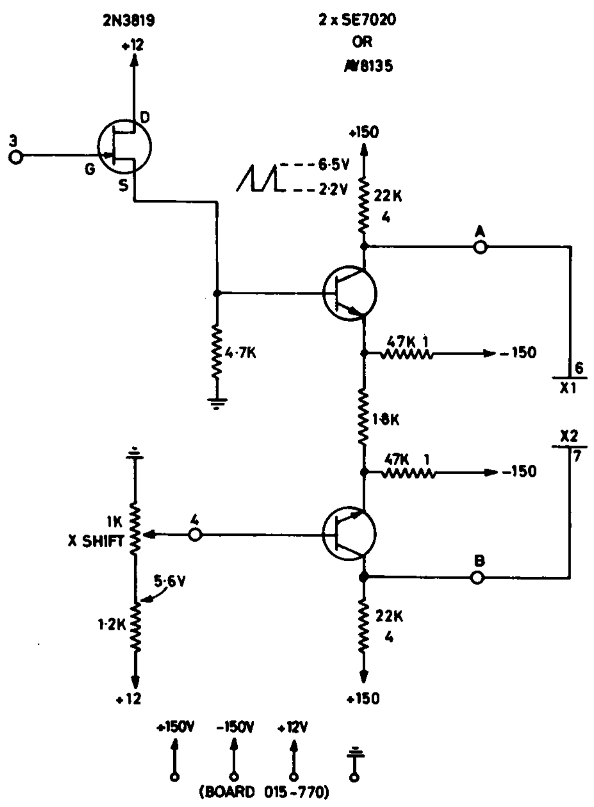


FIG. 14.4. *X* amplifier.

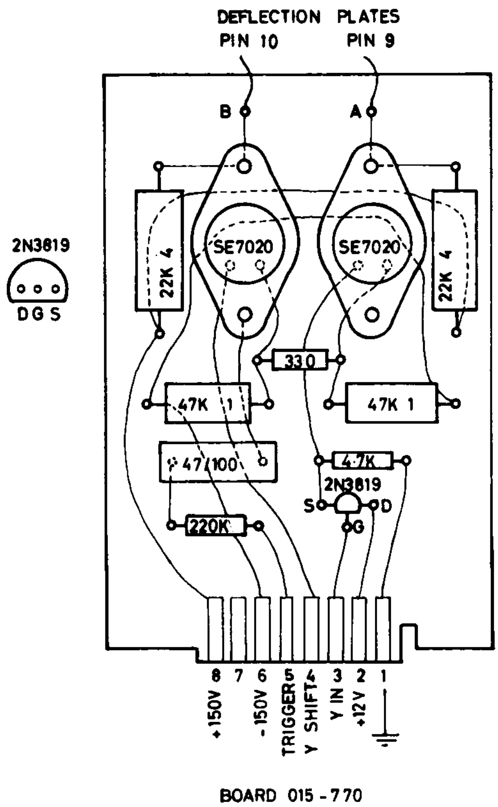


FIG. 14.5. Card for *Y* amplifier.

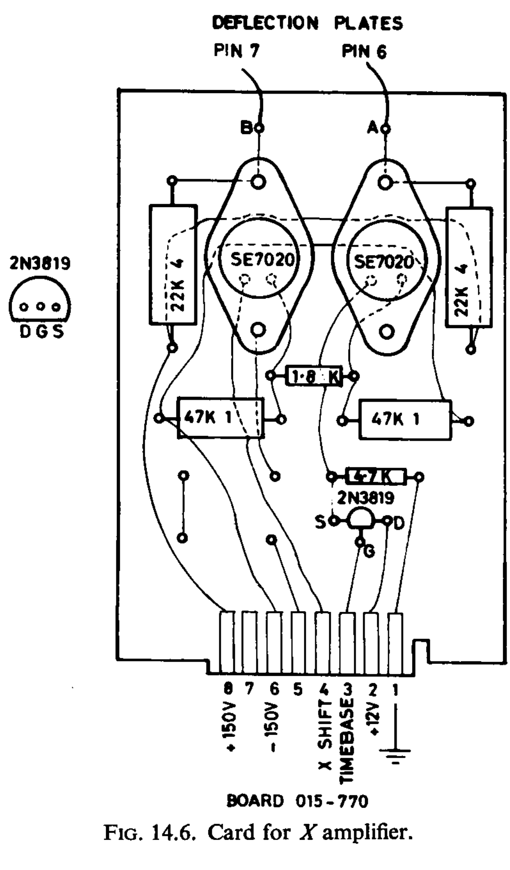


FIG. 14.6. Card for *X* amplifier.

requires about 0·5 V. The circuit can be analysed by the same method that was used in §14.2.

The base of Q2 is connected in each case to an adjustable voltage divider, which allows the spot to be centred in both X and Y directions; these *shift* voltages are in fact combined with (or more correctly, subtracted from) any signals applied to the base of Q1.

14.4 Source follower

The input signals of both the X and Y amplifiers are provided by means of field effect transistors; these are not connected as simple amplifiers, but as *source followers*. The basic source follower is shown in Fig. 14.7, using a N-channel junction FET.

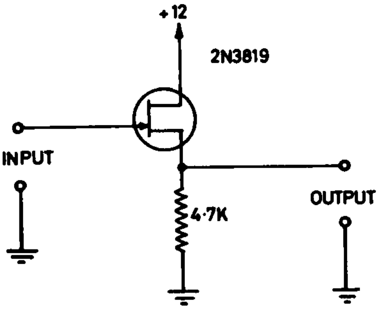


Fig. 14.7. Source follower.

This is an extremely useful circuit, having the property of a very high input impedance (typically 10^9 ohms) and a low output impedance (typically 500 ohms). Its voltage gain from input to output is about 0·9, the output voltage being always above the input. Curves for the 2N3819 transistor are shown in Fig. 14.8, the upper and lower curves

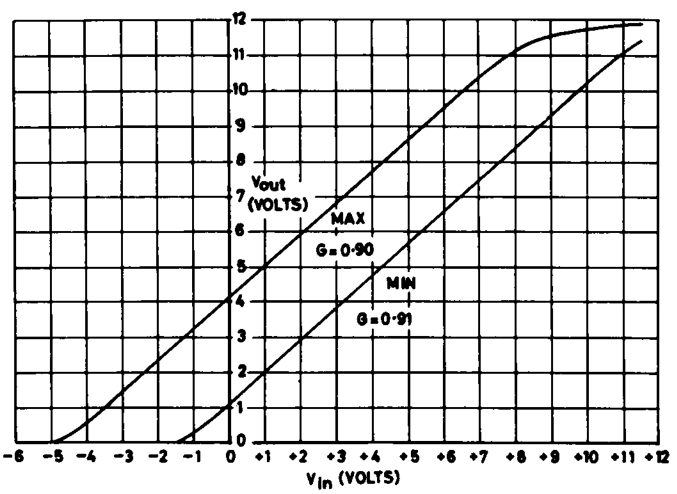


Fig. 14.8. 2N3819 FET as a source follower.

representing specimens of 2N3819 at the extremes of the range for this type. It will be seen that in the range of inputs from 0 to +7 V the curves are quite linear, and that the voltage gain is very close to 0·90 for all specimens.

To understand the operation of a source follower, reference should be made to the FET transfer characteristics shown in Fig. 9.3. Let us start by assuming an *ouput* voltage of, say, 4 V. By Ohm's law the current through the 4·7 K resistor, which is the drain-to-source current, I_D, must be 4/4·7, or 0·81 mA. Referring to Fig. 9.3 for the average FET, this requires a gate-to-source voltage V_{GS} of 2·3 V; i.e., the gate is 2·8 V below the output voltage of 4·0, putting it at +1·7 V. This must be the input voltage required to produce the output of 4 V originally assumed.

By assuming a series of output voltages in this way, and deducing the input voltage for each, the curves of Fig. 14.8 may be built up. It will be seen that the output voltage *follows* changes in the input voltage with considerable accuracy.

A bipolar transistor may be used in the same way, as an *emitter follower*. In this case the input impedance is much lower, but may be adequate for many purposes. If the emitter resistor is R_L, the input impedance is given with good accuracy by βR_L.

14.5 *Y* INPUT ATTENUATOR

It is usual to provide an oscilloscope with a switch on its *Y* amplifier, to allow the selection of any of a series of fixed voltage calibrations on the CRT screen. This is done by selecting in turn from a series of carefully adjusted voltage dividers; in the simple oscilloscope being constructed these are in the *Y* input circuit.

If a simple voltage divider were used in each switch position, the stray capacitances round the switch and wiring would effectively form a high-cut filter in conjunction with the divider, as shown in Fig. 14.9(a). By deliberately introducing a suitable compensating

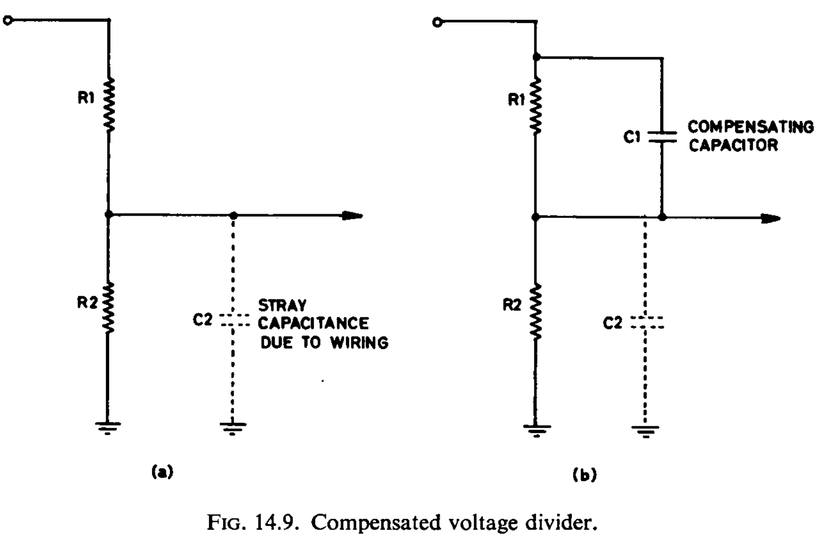


FIG. 14.9. Compensated voltage divider.

capacitor, as in Fig. 14.9(b), the effect of this stray capacitance may be balanced out: the value required is

$$C1 = \frac{R2}{R1}\,C2$$

The correct value is normally found by trial and error, by feeding a square wave into the input, and observing its reproduction on the CRT screen. It should neither undershoot nor overshoot, as shown in Fig. 14.10.

The values used for the Y attenuator switch in the oscilloscope under construction are shown in Table 14.2.

TABLE 14.2. Y ATTENUATOR VALUES

Switch position	$R1$	$R2$	C
100 mV/cm	0	1 M	—
200 mV/cm	470 K	470 K	18 pf
500 mV/cm	820 K	220 K	4·7 pf
1 V/cm	900 K	100 K	1·5 pf
2 V/cm	900 K	47 K	—
5 V/cm	1 M	22 K	—

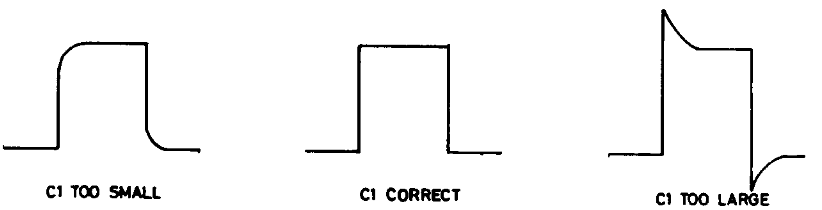


FIG. 14.10. Adjustment of compensated divider.

FURTHER READING

MILLMAN, J. and TAUB, H., *Pulse, Digital, and Switching Waveforms* (2nd ed.) McGraw-Hill, New York, 1973.

PRACTICAL

14.1 Set up a source follower on the bench, using a 2N3819, and following the diagram of Fig. 14.7. Vary the input voltage by means of a potentiometer, and plot input against output voltage. Compare your graph with Fig. 14.8. Insert a 10 M resistor in the gate lead, and verify that the input–output curve is unchanged. What conclusion can be drawn from this result?

14.2 Assemble the oscilloscope X and Y amplifier cards, following Figs. 14.3 to 14.6. After careful checking, plug them into their sockets in the oscilloscope. Temporarily ground the input to the X amplifier, and switch on. Set focus, intensity, and astigmatism, check that the X and Y shift controls operate, and test the Y amplifier and attenuator by the application of suitable d.c. voltages.

CHAPTER 15

TIME BASE GENERATION

15.1 The function of a time base

The most common use of an oscilloscope is for the study of voltages varying as a function of time. For this purpose the screen of the CRT is used as a graphical display, in which the X axis represents time on a suitable scale, and the Y axis voltage. This voltage may be a suitably amplified version of an electrical signal generated by the body, such as the electrocardiogram or electroencephalogram, or it may be generated by a *transducer*, which converts a variable such as blood pressure into a proportional voltage. To produce a linear time scale, the spot on the screen is moved in the X direction from left to right at a constant velocity. On reaching the right side of the screen, the spot is caused to *fly back* very rapidly to the left, and start a new *traverse* or *sweep*. If the phenomenon being examined is a *transient* one, occurring only once, the next sweep may be caused to show a new sample of it; this is true of most biological signals. If the phenomenon is *recurrent*, such as a sinusoidal or other periodic signal, it is usual to adjust the sweep repetition frequency to be a simple fraction of the frequency of the phenomenon being observed. In this way a few cycles of the phenomenon are traced out over and over again, giving an apparently stationary pattern on the screen. The signal applied in this way to the X plates is known as a *time base*.

To generate a linear sweep velocity, the voltage on the X plates must rise linearly; the flyback is then produced by a rapid fall to the original value. The *sawtooth* waveform required is shown in Fig. 15.1.

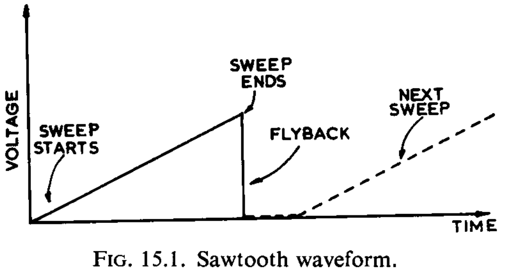


Fig. 15.1. Sawtooth waveform.

15.2 Generation of a sawtooth waveform

A simple approximation to a linearly rising voltage can be made by utilising the early part of the exponentially rising voltage across a capacitor which is charging from a voltage source through a resistor. Such an arrangement is shown in Fig. 15.2. When the

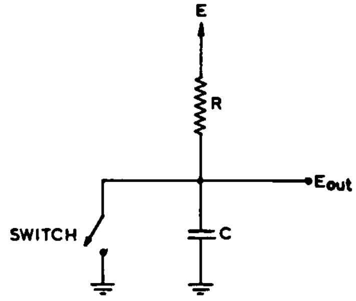


FIG. 15.2. Basic time base.

switch is closed the voltage across the capacitor is held at zero, and a current E/R flows through the resistor R and the switch to earth. When the switch is opened, this current flows into the capacitor, which commences to charge exponentially towards E; the voltage across it is

$$E_{\text{OUT}} = E(1-\epsilon^{-t/RC})$$

as was discussed in Chapter 6. The initial part of this charging curve is very nearly a linear rise, as will be seen from Table 15.1.

TABLE 15.1. INITIAL PORTION OF EXPONENTIAL CURVE

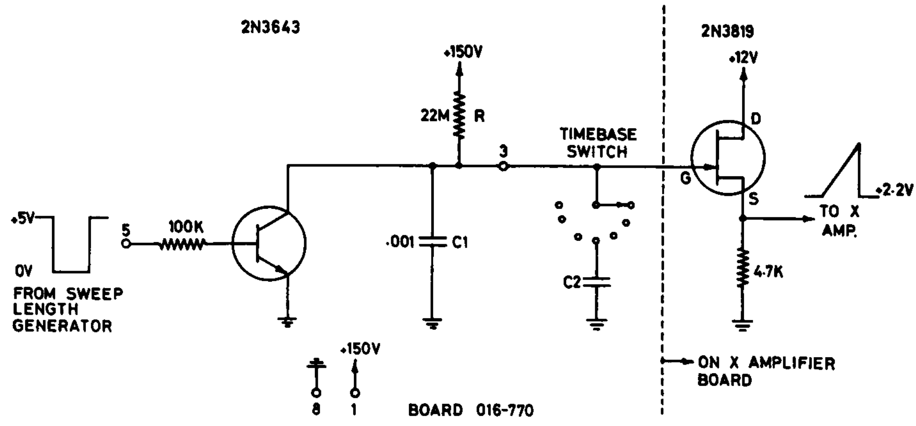
t	0	0·01RC	0·02RC	0·03RC	0·04RC	0·05RC
E_{OUT}	0	0·010E	0·020E	0·030E	0·039E	0·048E

The maximum deviation of the exponential curve from the straight line drawn from 0 to 0·048 E is about 1 %, so the curve is for most purposes an adequate approximation.

Oscilloscopes used for studying biological signals often require very slow sweep speeds; for a total sweep duration of 10 sec, a time constant of 100 sec or longer will be required. Typically, a value for R of 20 M, and a value for C of 5 μF would be used. Considerable care must be taken to avoid leakage of current around the capacitor, and the capacitor itself must be of high quality, with a polyester dielectric. Electrolytic capacitors are quite unsuitable for this purpose; their leakage is excessive, and their capacitance varies with age and temperature.

15.3 PRACTICAL TIME BASE

The circuit of a simple but practical time base, which will be incorporated into the oscilloscope under construction, is shown in Fig. 15.3, and the corresponding printed card and its connections in Fig. 15.4. It will be seen that a range of sweep speeds is provided, by the selection of any one of a bank of suitable capacitors; the smallest capacitor (0·001 μF) is permanently connected, and any one of the others is added in parallel with it, as required. The output voltage is connected to the gate of the source follower already incorporated into the X amplifier card, so the rising voltage is trans-



SWEEP VELOCITY		R	C_{TOTAL} (C1+C2)	
·5	SEC/CM	22M	5	MFD
·2	"	"	2	"
·1	"	"	1	"
50	MSEC/CM	"	·5	"
20	"	"	·2	"
10	"	"	·1	"
5	"	"	·05	"
2	"	"	·02	"
1	"	"	·01	"
·5	"	"	·005	"
·2	"	"	·002	"
·1	"	"	·001	"

FIG. 15.3. Time base.

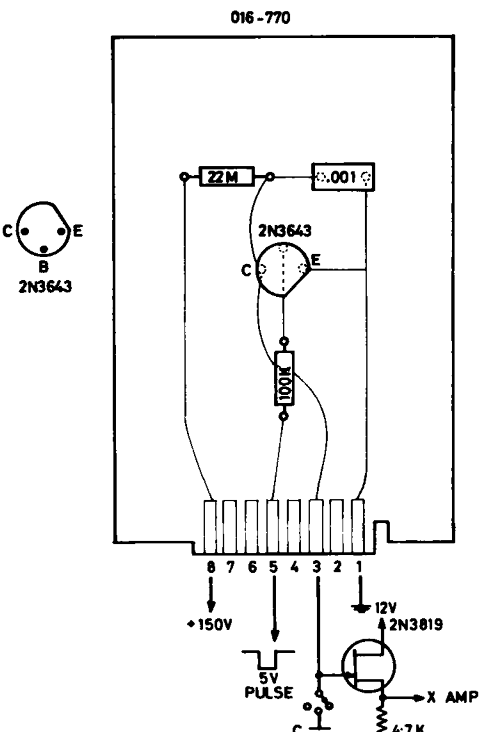


FIG. 15.4. Time base card.

ferred to the output amplifier without diverting any current from the charging capacitor The switch of Fig. 15.2 is replaced by a bipolar transistor. If the input terminal of this circuit (pin 5) is held at +5 V, a base current of 50 μA flows into the transistor, which becomes a good conductor from collector to emitter, and C is prevented from charging. If pin 5 is switched to ground (0 V) the base current ceases, the transistor becomes a very good insulator from collector to emitter, and C commences to charge. It will continue to do so until pin 5 is again returned to +5 V, when flyback will occur, as C discharges abruptly through the transistor. (Notice that the capacitor cannot charge to more than about 10 V; at this point the gate of the source follower will move positive to its source, and gate current will flow, stopping any further rise. Since only about a 4 V rise is required to generate the full sweep, this does not matter.)

FURTHER READING

KINMAN, K. A., *Sweep Generator Circuits*, Tektronix, Beaverton, Oregon, 1969.

MILLMAN, J. and TAUB, H., *Pulse, Digital and Switching Waveforms* (2nd ed.), McGraw-Hill, New York, 1973.

STRAUSS, L., *Wave Generation and Shaping* (2nd ed.), McGraw-Hill, New York, 1970.

PRACTICAL

15.1 Assemble the circuit of Fig. 15.3 on the printed card. Connect the output pin (3) on its socket to the time base range switch and the input pin of the X amplifier socket (pin 3). (Remove the temporary ground from pin 3 of the X amplifier socket.) Temporarily connect pin 5 of the time base socket to +4·5 V from a battery, and ground the other end of the battery. Switch on the oscilloscope, switch the time base range switch to its lowest velocity setting, and remove the +4·5 V from pin 5. The spot should sweep across the screen from left to right at 0·5 sec/cm, flying back to the left as soon as the +4·5 V is reconnected to pin 5. Check the other velocity settings in the same way; for the fastest settings it will be very difficult to observe the sweep, but the spot should vanish as soon as the 4·5 V supply is removed, and return when it is restored.

N.B. When the oscilloscope is being constructed by groups of students, it is suggested that the timing capacitors in the time base circuit, and also those used on the same switch in Chapter 16, should be selected and mounted on the switch assembly prior to the practical session.

CHAPTER 16

PULSE GENERATION—THE ASTABLE MULTIVIBRATOR

16.1 REQUIREMENTS

Frequently, the need arises to generate a rectangular *pulse* of voltage or current, having a predetermined magnitude and duration; very often a succession of such pulses is required to be produced at a known rate. In the oscilloscope under construction, the time base generator of Chapter 15 requires to be turned on and off repetitively if a succession of sawtooths is to be generated. In the biological context, single pulses or a succession of pulses are often required to stimulate a tissue into carrying out its normal function. Tissue stimulation will be discussed in Chapter 22.

Pulse generation is very commonly done by the use of the *multivibrator* circuit, first described by Abraham and Bloch in 1919. Multivibrators may be classified as *astable*, *monostable*, or *bistable*, depending on whether they produce a continuous train of pulses, a single pulse started by a trigger pulse, or a single pulse which is started by one trigger pulse and ended by a second. The outputs and trigger requirements of the three types are shown in Fig. 16.1, which also shows input requirements and output for a variant of the bistable multivibrator, the *Schmitt trigger*. This last circuit starts its output pulse whenever its input voltage exceeds a predetermined *upper threshold*, and ends it when the input next falls below a predetermined *lower threshold*.

The astable multivibrator will be discussed in this chapter, the monostable and Schmitt trigger in Chapter 17, and the bistable in Chapter 20.

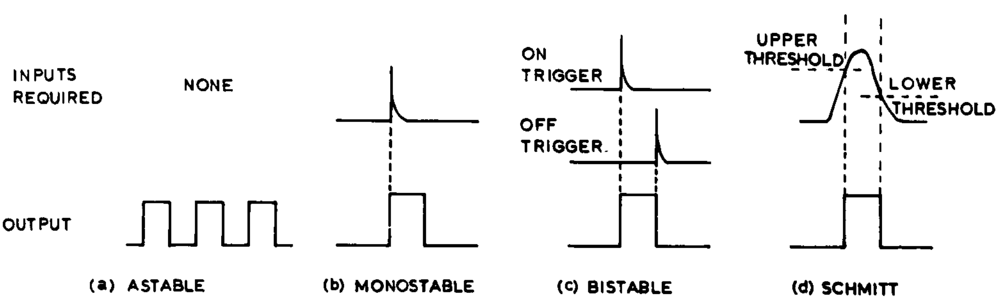


FIG. 16.1. Inputs and outputs of multivibrator circuits.

16.2 THE TRANSISTOR AS A SWITCH

In the discussion in the foregoing chapters, we have seen various ways in which the transistor can be used as an amplifier, with a base bias current such that the collector voltage lies about halfway between zero and the supply voltage. An equally common use for the transistor is as a switch, and one example of this mode has already been seen in Chapter 15. In the switching mode the collector current is either completely cut off or fully turned on, and the transition from one state to the other is made as rapidly as possible.

For the collector current to be completely turned off, it is necessary to reduce the base current to zero, and this is usually done by applying either zero or a reverse voltage to the base-to-emitter diode. Care must be taken, however, not to exceed the breakdown voltage of this diode at any time; 5 V in the reverse direction is the usual maximum permissible value. With the collector current turned off, there is no voltage drop across the collector load, and the output voltage is equal to the supply voltage, as in Fig. 16.2(a).

A fully on transistor has a collector-to-emitter voltage of 0·1 V or less, so practically the whole supply voltage will appear across the collector load under these circumstances (Fig. 16.2(b)). The collector current can readily be calculated by Ohm's law; in Fig. 16.2(b) it will be (12 − 0·1)/1 or 11·9 mA. The base current to produce this may then be found by dividing the collector current by β; assuming $\beta = 400$ for the example of Fig. 16.2(b), a base current of 30 μA would be required. Increasing the base current beyond this value cannot increase the collector current; the transistor is said to be *saturated*. In practice a margin of safety is always allowed; in the case discussed here a base current of about 100 μA would be desirable.

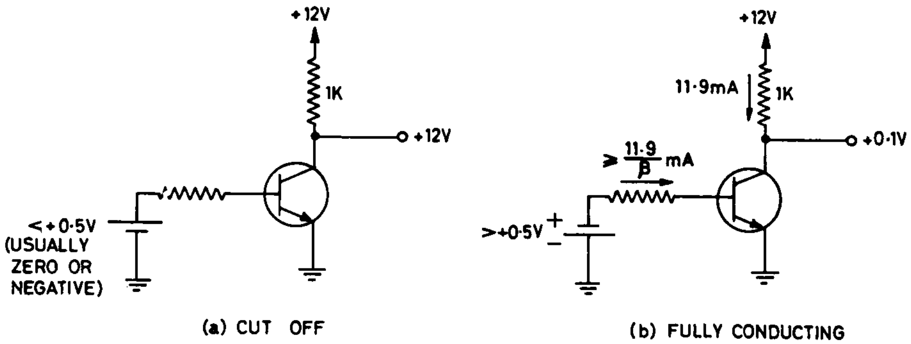


FIG. 16.2. The transistor as a switch.

It should be noticed that the power dissipated in a cut off transistor is zero, and that in a saturated transistor is very small, since the collector-to-emitter voltage is very small; quite large currents can be handled by small transistors. During the process of transition from one state to the other, however, the power dissipation is momentarily much larger, and for any given number of transitions per second this sets an upper limit to the current-carrying capacity.

16.3 THE ASTABLE MULTIVIBRATOR

The circuit of a typical astable multivibrator is shown in Fig. 16.3. The circuit as shown is completely symmetrical. It will alternate between a condition in which Q1 is cut off and Q2 is saturated, and a condition in which the reverse is true, at a rate governed by the time constants R_1C_2 and R_2C_1. To understand its operation, assume that initially

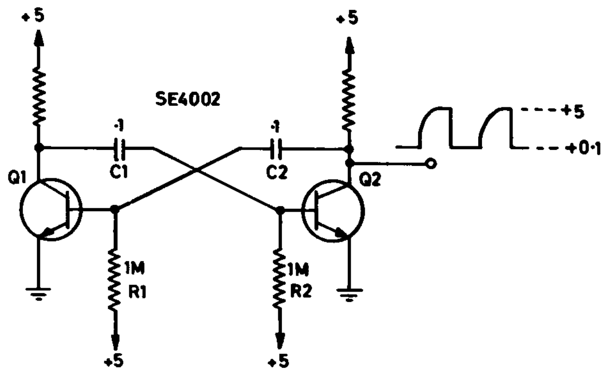


FIG. 16.3. Astable multivibrator.

Q1 is cut off and Q2 saturated, as in Fig. 16.4. The collector of Q2 will be at +0·1 V, and its base at about +0·6 V, since it is saturated. Q1 is cut off, so its collector will be at +5 V, and its base at some voltage lower than +0·5 V, or even negative. (The exact value will be determined later.) The first three of these voltages are quite stable, and can be maintained indefinitely; that at the base of Q1 cannot. It appears at the top of R1, which has +5 V at its lower end. So by Ohm's law, a current must be flowing through R1. It is not entering the base of Q1 (the base-to-emitter diode is reverse biased) so it must be altering the charge on C2, and causing the base of Q1 to move exponentially towards +5 V, with a time constant of R_1C_2. These conditions are shown by point 1 in each of the diagrams of Fig. 16.5.

After a short time the base of Q1 will cross the +0·6 V value (point 2 in Fig. 16.5). As soon as this occurs, Q1 will be forward biased, and will commence to conduct. Its collector voltage will move negatively from +5 V, and since C1 cannot change its state of charge instantly, the base of Q2 also moves negatively. This reduces the collector current of Q2, causing its collector voltage to move positively. This in turn drives a surge of current into the base of Q1 by way of C2, charging C2 as it does so, and the whole

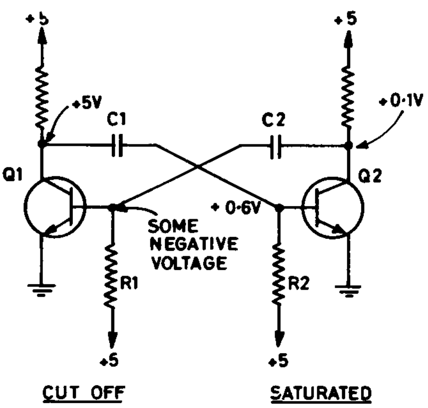


FIG. 16.4. Astable multivibrator: assumed initial state.

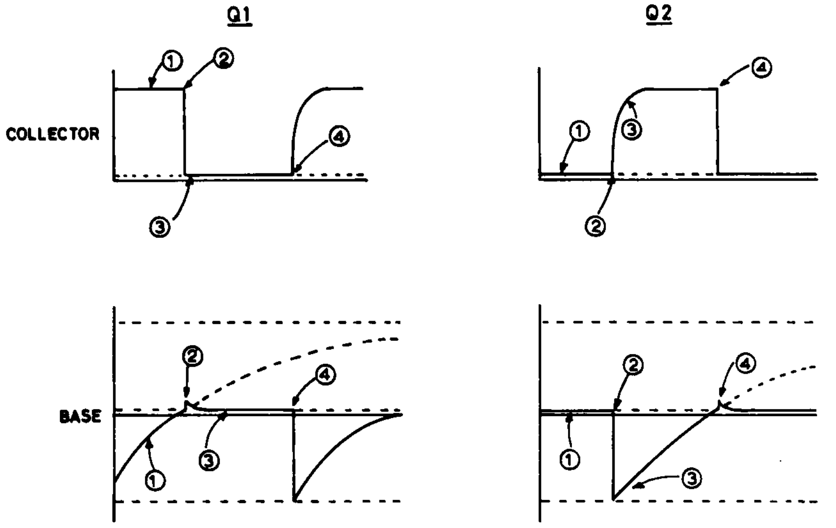


FIG. 16.5. Wave forms in astable multivibrator.

process becomes cumulative. In a period of less than a microsecond Q1 is driven into saturation, with its base-to-emitter diode forward biased. Its collector goes to +0·1 V, a change of 4·9 V in the negative direction. This sudden change moves the base of Q2 4·9 V negatively also, from its original +0·6 V to −4·3 V, and cuts Q2 off hard (point 3 in Fig. 16.5). The two transistors have now changed roles with each other. C1 commences to charge through R2 just as C2 did previously through R1; presently the circuit reverts to its original state by the reverse of the process described above (point 4 in Fig. 16.5). This alternation of roles proceeds indefinitely, generating a series of almost rectangular pulses at the collector of each transistor. They will have an amplitude of 4·9 V, and a duration governed by R_1C_2 and R_2C_1.

The complete charging curve would run from −4·3 V to +5 V, a range of 9·3 V. The circuit actually changes state at +0·6 V, a change of 4·9 V. This is 4·9/9·3, or 0·53, of the total range, and reference to Table 6.2 shows that this takes 0·8 of a time constant. Thus the values of C1 and C2 can be calculated to give any required pulse duration. If it is desired to construct a multivibrator with a range of pulse durations, it is usual to have a series of fixed values of C1 and C2 selected by a switch. It should be noted that R1 and R2 can also be varied, but the maximum value that can be used is that which just permits sufficient base current to saturate the transistors, as discussed in § 16.2 above.

It is also possible to make the two time constants different, to produce ratios of on to off times other than 1:1, but ratios in excess of about 5:1 cannot be achieved without interfering with the normal operation of the circuit.

In modern practice astable multivibrators are frequently constructed from single integrated packages; this will be discussed in Chapter 20.

16.4 Differentiation of output pulse

It frequently happens that a train of sharp pulses at regular intervals is required, either for timing other events, or for triggering a subsequent monostable or bistable multivibrator. Where high accuracy of timing is not required, an astable multivibrator may be used as a pulse generator, with its output differentiated by a simple high-pass RC filter, as discussed in Chapter 10. Fig. 16.6 shows how the output of the circuit of Fig. 16.3 will appear before and after differentiation in this fashion. The positive-going edge of each pulse produces little output, since it is relatively slow. The negative-going edge is reproduced as a sharp pip of height almost equal to that of the original pulse, and having a duration governed by the time constant of the differentiating circuit.

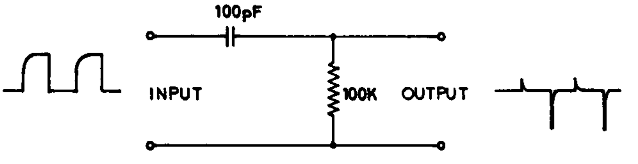


Fig. 16.6. Differentiating circuit.

FURTHER READING

Millman, J. and Taub, H., *Pulse, Digital and Switching Waveforms* (2nd ed.), McGraw-Hill, New York, 1973.

PRACTICAL

16.1 Assemble the circuit of Fig. 16.3 on a printed card, as shown in Fig. 16.7. What pulse duration, and what pulse repetition rate, do you expect it to produce? Plug the card into a socket to which you have attached a 5 V power supply or a suitable battery, and use an oscilloscope to examine the waveforms produced. Sketch them, showing the amplitudes and durations concerned.

16.2 Plug the card into the oscilloscope under construction, in the pulse generator socket. Switch on, and select a sweep velocity such that the sweep on the oscilloscope screen just fills the screen. Apply a signal from an oscillator to the *Y* input terminals, and adjust its frequency until a stationary pattern is produced. Is the pattern easy to hold stationary? What modification to the arrangement would be necessary to be able to study *any* frequency produced by the oscillator?

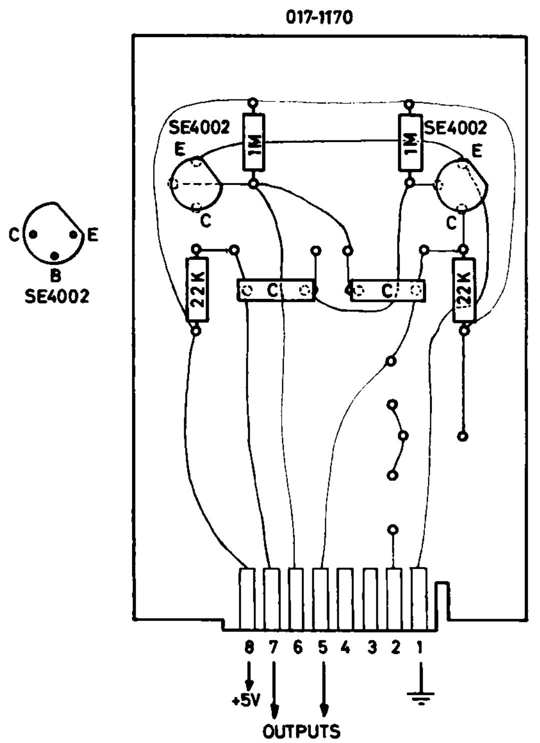


FIG. 16.7. Astable multivibrator card.

CHAPTER 17

PULSE GENERATION—THE MONOSTABLE MULTIVIBRATOR AND SCHMITT TRIGGER

17.1 Applications of a monostable multivibrator

There are many applications in which a single pulse of predetermined duration is required each time a trigger pulse occurs. It may be necessary to increase the duration of the original pulse to operate some device which responds only relatively slowly; it may be required to operate a device for a definite time on receipt of a trigger pulse, or it may be required to generate a trigger pulse a definite time after the receipt of another trigger pulse. All these functions may be carried out by a monostable multivibrator.

17.2 Basic monostable multivibrator

The circuit of a typical monostable multivibrator is shown in Fig. 17.1. It will be seen that the circuit is very similar to that of the astable multivibrator; one RC cross-coupling is replaced by a direct connection from the collector of Q2 to the base of Q1. In the quiescent state Q2 is held saturated by the base current supplied through the base resistor R; its base will be at +0·6 V, and its collector at +0·1 V or less. This voltage will be transferred to the base of Q1 through the 100 K cross-coupling resistor, and since it is less than +0·6 V, Q1 will be cut off. No current will flow through the 100 K resistor, since the base-to-emitter diode of Q1 is reverse biased. The collector of Q1 will draw no current, and will thus be a +5 V. All these voltages are stable, and will be maintained indefinitely unless the circuit is triggered. These conditions are shown by point 1 in Fig. 17.2.

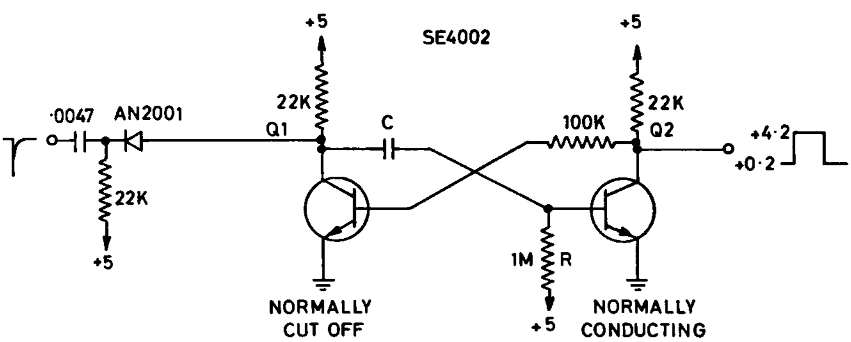


Fig. 17.1. Basic monostable multivibrator.

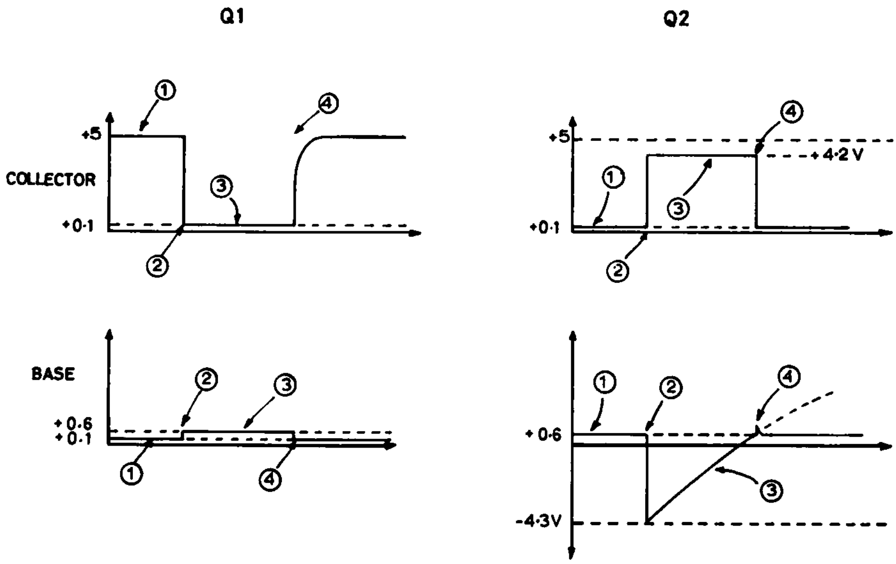


FIG. 17.2. Waveforms in monostable multivibrator.

If now a short negative pulse is applied to the trigger input terminal, it will carry the collector of Q1 negative. Since C cannot change its state of charge instantly, the base of Q2 is also carried negative. The collector current of Q2 falls, and causes the collector voltage to rise; this rise is transferred to the base of Q1. As soon as the base of Q1 moves above +0·6 V, Q1 commences to conduct; its collector voltage falls further, and the whole action becomes cumulative (point 2 in Fig. 17.2). In less than a microsecond Q2 is fully cut off, and Q1 is driven to saturation. The collector of Q1 will have fallen to +0·1 V, a drop of 4·9 V, and carried the base of Q1 4·9 V negative also by way of C, from its previous value of +0·6 V to −4·3 V; this will cut Q2 off hard. The collector of Q2 will draw no current, and the base of Q1 will thus be connected to +5 V through the 100 K cross-coupling resistor and the 22 K collector load of Q2. The base of Q1 will be at +0·6 V, so by Ohm's law a base current of (5 − 0·6)/(100 + 22), or 37 μA through the 22 K collector load will drop 0·8 V, putting the collector of Q2 at +4·2 V (point 2 in Fig. 17.2).

Three of these four voltages are stable; however, the voltage at the base of Q2 is not. The resistor R has now (5 + 4·3), or 9·3 V across it, and a current of 9·3 μA will flow through it, by Ohm's law. This current must alter the charge on C, since the base of Q2 is reverse biased and draws no current. C will charge exponentially through R towards +5 V (point 3 in Fig. 17.2). As soon as the base of Q2 reaches +0·6 V, Q2 will commence to conduct, its collector voltage will fall, and by a cumulative action the circuit will rapidly reset to its quiescent state (point 4 in Fig. 17.2). The time for which the circuit remains in the triggered state will be given by 0·8 RC, as in the case of the astable multivibrator. The output voltage at the collector of Q2 is a positive-going rectangular pulse, 4·0 V high. A similar negative-going pulse may be derived from the collector of Q1 if required.

17.3 Triggering circuit

In Fig. 17.1 the input trigger pulse is applied through a differentiating circuit and a diode; the resistor of the differentiating circuit is returned to +5 V. In the quiescent state of the multivibrator, the collector of Q1 is at +5 V also, so there is no voltage across the diode, and it is not conducting. As soon as the voltage at the trigger input falls by more than 0·6 V, the diode commences to conduct, and the collector voltage of Q1 is made to follow the trigger pulse down. This fall is transferred to the base of Q2 through the capacitor C; the collector current of Q2 is reduced, and the cumulative action which results in the circuit entering the triggered state commences. Once triggering has been initiated, the monostable multivibrator will ignore any further trigger pulses until it resets.

17.4 Monostable multivibrator as time base switch

It is clear from Chapter 16 that there are a number of disadvantages in the use of the astable multivibrator as a time base switch. Even assuming that its frequency and pulse duration are made continuously variable to allow for the examination of waveforms of any frequency, it cannot be used to examine waveforms which occur at irregular intervals, a very common requirement. Moreover, it requires the simultaneous selection of both cross-coupling capacitors by a range switch; since the time base velocity capacitor must also be switched, this calls for a triple bank of switched capacitors, all carefully selected.

If a monostable time base switch is used instead, and is triggered either by the stimulus which initiates a sequence of events to be studied, or by the beginning of a sequence itself, the display of the sequence cannot fail to be stationary on the screen even if it recurs at irregular intervals. All that is required is that the one capacitor in the multivibrator should be selected by a second bank on the time base capacitor switch. The multivibrator capacitor must be such that the sweep *duration* determined by the multivibrator is in inverse proportion to the sweep *velocity* determined by the time base. In this way the sweep *length* will remain constant on the oscilloscope screen, no matter what velocity is selected.

The circuit of the monostable multivibrator to be used in the oscilloscope under construction is shown in Fig. 17.3, and its printed card in Fig. 17.4.

In this circuit the shortest switching time is given by the 0·001 μF coupling capacitor, and longer times are obtained by adding additional capacitors in parallel between pins 6 and 5. The 470 pF capacitor in parallel with the 100 K cross-coupling resistor serves to speed up the transfer of voltage from the collector of Q2 to the base of Q1. In the oscilloscope under construction, the time base switching waveform is taken from pin 5; the positive-going output from pin 7 will be used later (Chapter 18).

If this circuit is substituted for the astable multivibrator of Chapter 16 in the oscilloscope under construction, no time base sweep at all will occur until the multivibrator is triggered by an incoming short negative-going pulse on pin 2. When this occurs the time base will generate *one sweep only*, at a velocity set by the time base capacitor selected; as the end of the sweep is reached, the multivibrator will reset, flyback will occur, and then the circuit will remain quiescent, with the spot on the left side of the screen, until triggered again.

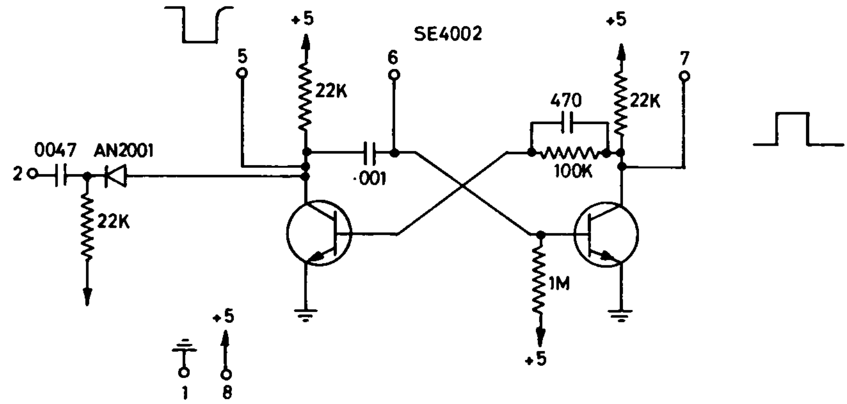


FIG. 17.3. Monostable multivibrator for oscilloscope.

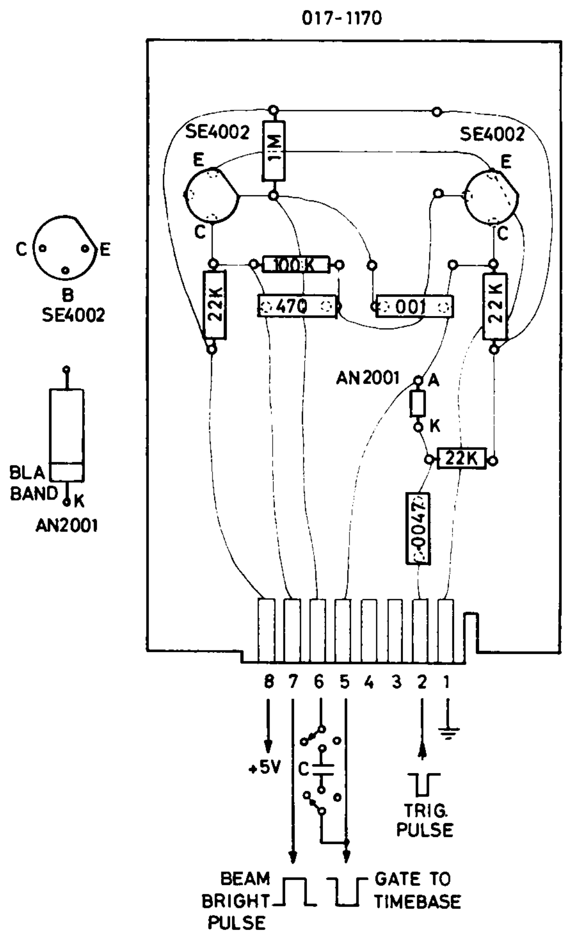


FIG. 17.4. Printed card for monostable multivibrator.

17.5 Monostable multivibrator as a pulse delay generator

If a monostable multivibrator is followed by a differentiating current and diode, the combination will produce a negative-going output trigger pulse delayed on the input trigger pulse by the duration of the monostable pulse, as shown in Fig. 17.5. This is a very common method of producing an event at a predetermined time after an initiating trigger pulse; by use of a series of monostable multivibrators a whole sequence of events may be initiated, each at its own desired time. In modern practice monostable multivibrators are usually purchased as single integrated circuit packages; this will be discussed in Chapter 20.

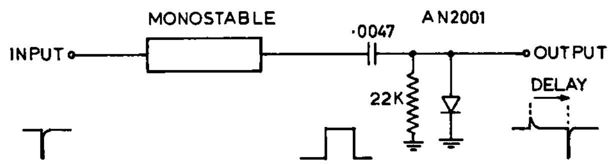


Fig. 17.5. Monostable multivibrator as a pulse delay generator.

17.6 Schmitt trigger

The Schmitt trigger, first described by O. H. Schmitt in 1938, carries out the function of a cumulative change of state whenever its input voltage rises above a predetermined upper threshold level, and a cumulative reset whenever its input voltage falls below a predetermined lower threshold level. The output is identical, no matter how slowly the input voltage is changing. This is a most valuable property, and the incorporation of a Schmitt trigger in the oscilloscope under construction will permit the monostable multivibrator to be triggered from a predetermined point on any input voltage waveform, no matter how slowly it may be varying.

The operation of the Schmitt trigger may be understood by reference to Fig. 17.6 Suppose that initially the input voltage on the base of Q1 is low enough to cut Q1 off

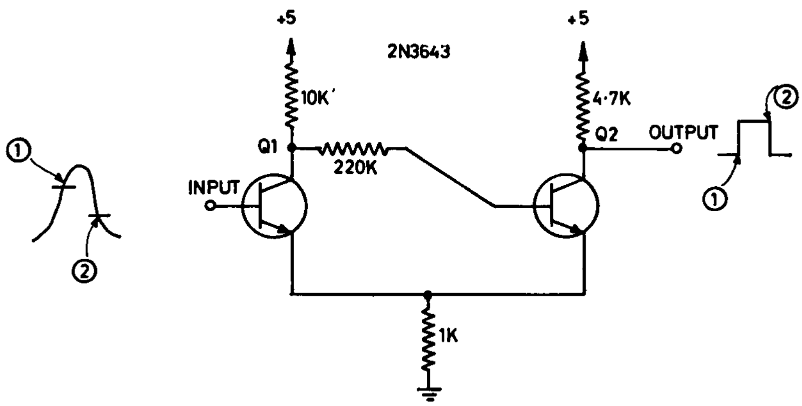


Fig. 17.6. Basic Schmitt trigger circuit.

Under these conditions the collector of Q1 is drawing no current, and Q2 is held in saturation by its base current, supplied from +5 V through the collector load of Q1 and the cross-coupling resistor in series (10 K + 220 K). If Q2 is saturated, there is a voltage drop of say 0·1 V across it, and thus there will be 4·9 V across its collector and emitter loads in series (4·7 K + 1 K). By Ohm's law, this will put its emitter at 0·84 V, and its collector at 0·94 V. For saturation its base voltage will be (0·84 + 0·6) V, or 1·44 V, and its base current by Ohm's law will be (5 − 1·44)/(220 + 10), or 15·5 μA. Its collector current will be (5 − 0·94)/4·7, or 864 μA, so it is certainly saturated.

Remembering that the emitter voltage of Q2 is also that of Q1 (0·84 V), we can see that so long as the base voltage of Q1 remains below the value required to make Q1 conduct (0·84 + 0·6, or 1·44 V), the circuit conditions will remain unaltered. This value, 1·44 V, is the *upper threshold.* As soon as 1·44 V is exceeded, Q1 will commence to conduct. Its collector voltage falls, dropping the base voltage of Q2, and very quickly Q2 comes out of saturation. Its *emitter* current falls, and so does the drop across the 1 K emitter resistor (Q1 is now contributing a small additional current to this resistor, but this addition is much smaller than the reduction due to Q2). The falling voltage across the 1 K resistor is of course also a falling emitter voltage on Q1, and this turns Q1 on harder still. The action becomes cumulative, and in less than a microsecond Q1 saturates and Q2 cuts off.

With Q1 alone conducting, the voltage across the 1 K resistor may now be calculated. The current through Q1 is clearly (5 − 0·1)/(10 + 1), or 0·44 mA, by Ohm's law. So the emitter of Q1 will now be at 0·44 V. As long as the voltage on the base of Q1 remains above (0·44 + 0·6), or 1·04 V, Q1 will continue to conduct; this is the *lower threshold.* If it falls below this value, Q1 will cut off, and by cumulative action the circuit will very rapidly revert to its original state.

Schmitt triggers are usually purchased as single integrated circuit packages; this will be discussed in Chapter 20.

17.7 SCHMITT TRIGGER IN TIME BASE CIRCUIT

The circuit is modified from that of Fig. 17.6 by the addition of a 100 pF "speed-up" capacitor across the 220 K cross-coupling resistor, and by the addition of an input emitter follower. This is necessary because the base of Q1 must always be driven from a low resistance source for proper action of the trigger circuit. Fig. 17.7 also shows the external connections to permit triggering either from an external waveform or from the signal under examination.

It is possible to add many other desirable features to an oscilloscope time-base, but all oscilloscopes use an arrangement similar to that described in the foregoing chapters. Controls to alter the threshold of the Schmitt trigger, or to permit triggering from either positive or negative-going signals, are among the most common.

FURTHER READING

Sweep Generator Circuits, Tektronix, Beaverton, Oregon, 1968.

MILLMAN, J. and TAUB, H., *Pulse, Digital and Switching Waveforms* (2nd ed.), McGraw-Hill, New York, 1973.

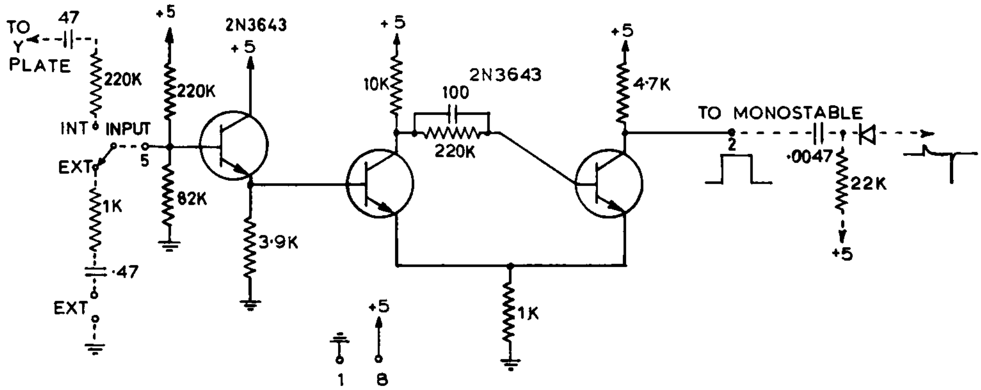


FIG. 17.7. Schmitt trigger as used in oscilloscope.

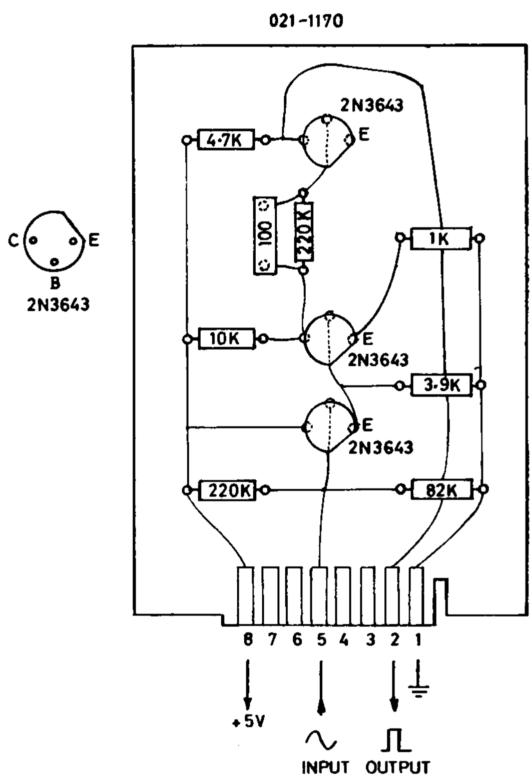


FIG. 17.8. Card for Schmitt trigger.

PRACTICAL

17.1 Assemble the monostable multivibrator of Fig. 17.3 on its card. Plug it into a socket on the bench, to which a 5 V supply has been connected, and test that the collector voltages are as expected. Connect a 1 μF condenser in parallel with the 0·001 on the card (across pins 5 and 6); trigger the circuit by the application of a negative pulse from a battery, and check that it cycles correctly.
17.2 Install the monostable multivibrator card in the oscilloscope in place of the astable multivibrator previously used. Connect up the capacitor selector switch to it. Switch on, and test the whole circuit by triggering the multivibrator from a battery or pulse generator.
17.3 Assemble the Schmitt trigger of Fig. 17.7 on its card. Plug it into a socket on the bench, to which a 5 V supply has been connected, and by means of a battery and a potentiometer at its input determine its upper and lower threshold voltages. Note the value of emitter voltage for each of the two states of the circuit.
17.4 Install the Schmitt trigger card in the oscilloscope. Connect up the input circuitry, and test the whole time base circuit on both internal and external triggering. The complete time base circuit is shown in Fig. 17.9.

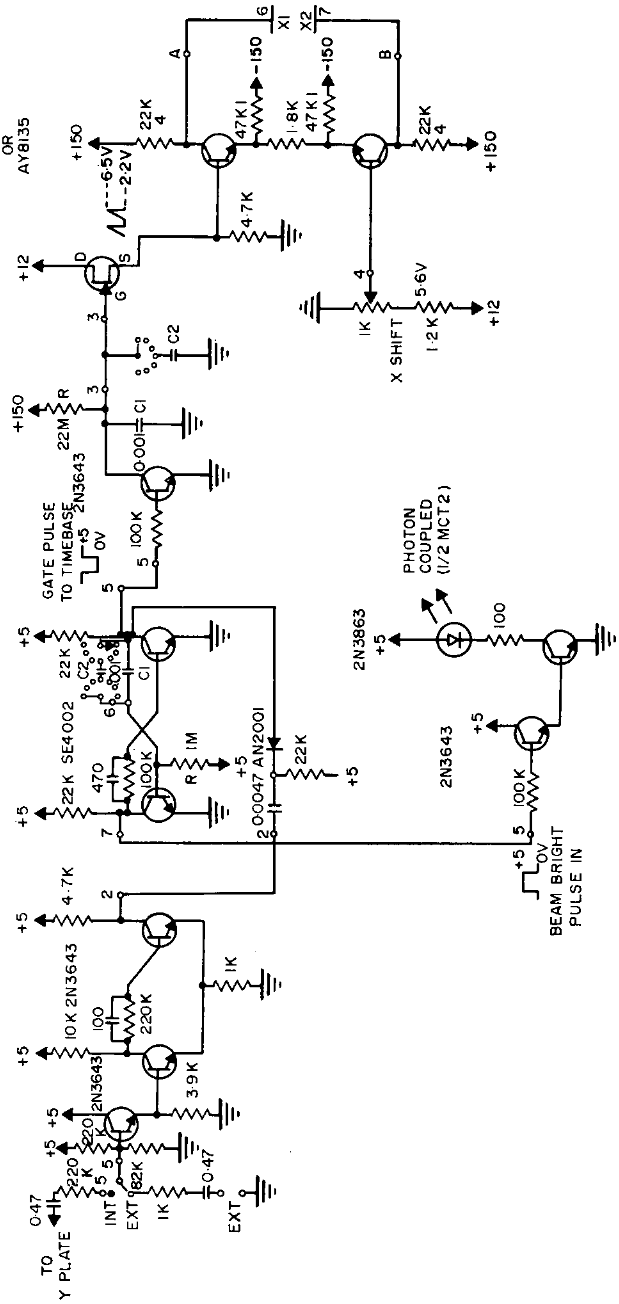


FIG. 17.9. Complete time base circuit.

CHAPTER 18

OPTOELECTRONICS

18.1 SCOPE OF OPTOELECTRONICS

The term optoelectronics covers the study of all electrically driven light-emitting devices, all devices which produce electrical effects when exposed to light, and all methods of conducting light through fixed pathways. Light-sensitive cells, or *photocells*, have been in use for over half a century, but in general have been bulky and comparatively slow in response to rapid changes in illumination. Light sources have had many disadvantages; in particular, most are bulky, inefficient, and incapable of rapid variation in light emission in response to a variation in input current.

In the past few years much work has been carried out on the development of *PN* diodes which are light-sensitive, and on others which are capable of emitting light when a current is passed through them; recently the price of such diodes has come down to a point where it is possible to use them in commercial equipment.

18.2 UNITS USED IN PHOTOMETRY

A wide variety of units for measurement of light and illumination is found in manufacturers' specifications for light sources and photocells. In this chapter we will consider only the relevant units of the Système International (S.I.).

When a current is passed through a light source, a portion of the power dissipated appears in the form of light radiated outward from the source; the frequencies radiated will depend on the source used, and usually on the current passed. The total radiated energy, or the energy within a specified band of frequencies, can be expressed in watts, and sometimes this is done. For visible light, it is more convenient to use a special unit of *luminous flux*, the *lumen* (lm). For example, a standard 6 V 300 mA globe emits a total of about 12 lm, and a 100 W domestic globe about 1200 lm.

When light falls on a surface to be illuminated, the *illumination* is measured in lumen per square metre of surface; 1 lm/m^2 is 1 *lux*. The illumination of a surface by a source of light will of course depend on what fraction of the number of lumens emitted the surface actually receives. If we require the illumination on a surface 3 m distant from a source radiating 1000 lm uniformly in all directions, we proceed by calculating the surface area of a sphere of radius 3 m with the source at its centre:

$$\begin{aligned} A &= 4\pi r^2 \\ &= 113 \text{ m}^2 \end{aligned}$$

Since 100 lm is uniformly distributed over 113 m^2, the illumination is 1000/113 lm/m^2, or 8·85 lux. (A reasonable illumination for reading is 50–100 lux.)

The *luminance* (or brightness) of a source of light is similarly measured in lm/m^2 of emitting surface. If an opalescent 100 W domestic globe is considered to be a uniformly emitting sphere of 0·03 m radius, it will have a surface area of 0·0113 m^2; if it emits 1000 lm, its surface luminance will be 1000/0·0113, or 88500 lm/m^2.

18.3 Incandescent lamp sources

Incandescent lamps are still the predominant source of light where a steady illumination is required, either to indicate to a user that some particular circuit is operating, or to provide a beam of light across an intervening space on to a photosensitive device. Interruption to this light beam can be made to perform a switching operation, or the beam can be partially absorbed or deflected to permit the measurement of the optical density of a solution. This latter arrangement has been used in a wide variety of biochemical instruments.

To obtain long life from an incandescent lamp, two points should be carefully watched. First, the lamp should be operated at a voltage lower than that recommended by the manufacturer for normal use. This will reduce the amount of light emitted, and will move the wavelength of maximum emission further into the infra-red, but will increase the life expectancy by a very large factor; the reduction of light is usually unimportant. Secondly, the initial current surge when the lamp is switched on should be limited. Lamp filaments have a wide variation in resistance between the initial cold value and the final hot value; as the filament heats the resistance will rise by a factor of as much as 10. Accordingly, for a constant applied voltage the initial current may be ten times the final operating current, and it is usual for failure to occur during this initial surge. The surge can be limited by operating the lamp in series with a resistor, whose value is selected to set the hot lamp voltage to the desired figure. This will increase the time taken for the lamp to reach full brightness, but for most applications this is not important.

For measurement of optical density of a specimen, the amount of light produced by the lamp must remain constant between calibrations. The light output is proportional to approximately the fourth power of the applied voltage, so a very constant lamp supply is essential. For accurate work an electronically regulated supply should be used.

18.4 Light emitting diodes

There are a number of applications for which incandescent lamps are unsuitable. In particular, where it is required to switch the light source on and off very rapidly, a heated filament is incapable of responding. For this purpose light-emitting diodes are excellent; they are capable of switching on and off in a few nanoseconds, and the light produced is directly proportional to the current passed through them. In addition, they operate on low voltages and currents, are highly efficient, very small, and have a very long life expectancy.

When an ordinary *PN* diode is forward biased, the energy lost when holes and electrons combine at the junction is dissipated as heat. Certain types of semiconductor material, such as gallium arsenide, yield up some of this energy in the form of emitted light; since GaAs is a transparent material the light from such a *PN* junction is visible externally.

The wavelength of this light depends on the diode voltage in the forward conducting direction; a junction between *N*-type GaAs and a *P*-type material starts to conduct at about 0·8 V, and emits only infra-red light, at a wavelength of 900 nm. Visible light may be produced by the use of materials having a higher forward voltage; a gallium arsenide phosphide (GaAsP) junction emits red light at 670 nm, and a gallium phosphide (GaP) junction green light at 500 nm. The visible light diodes are progressively less efficient as the wavelength falls, and their cost rises rapidly; however, red-emitting diodes are quite efficient and quite cheap.

Light emitting diodes are operated by passing a current in the forward direction; a typical red-emitting diode, the Monsanto MV10, has a brightness of about 5000 lm/m^2 for a current of 50 mA, and requires 1·65 V across it at this current. It takes a nanosecond to turn on or off. The type of diode most commonly available has a domed lens to increase visibility from the side; flat-faced diodes are obtainable, and these are intended for coupling to some form of "light pipe" through which the light is to be transmitted.

18.5 Photocells

The three commonest varieties of light-sensitive device in current use are the *photo-transistor*, the *photo-diode*, and the *photomultiplier tube*. The first two are solid state devices, and depend on the principle that light falling on a *PN* junction causes the formation of electron-hole pairs, and thus produces the same effect as a forward bias. The photo-multiplier tube is a vacuum tube device, depending on electron emission from a suitable material exposed to light.

18.6 The photo-transistor

The photo-transistor consists of a *NPN* silicon transistor with a window to permit light to strike its base-to-emitter junction; this window is usually in the form of a lens to focus an incoming parallel light beam on to the junction. The base connection is usually left open, so the transistor will be non-conducting from collector to emitter in the absence of illumination (Fig. 18.1). When incident light strikes the base-to-emitter junction, electron-hole pairs are formed, and the collector commences to draw current; for the circuit shown a typical photo-transistor will saturate at an illumination of about 1000 lux.

Silicon photo-transistors have a maximum sensitivity to light in the near infra-red region, at about 830 nm. The spectral peak is quite broad, and the transistors respond

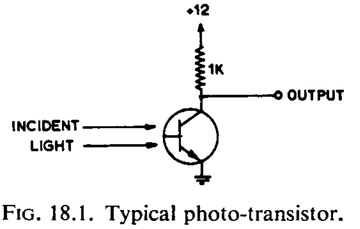


Fig. 18.1. Typical photo-transistor.

well to both red GaAsP and infra-red GaAs sources, and to light from incandescent lamps. Their response time is adequate for most purposes; rise time is typically 1 to 3 μsec.

18.7 The photo-diode

The photo-diode is a silicon diode with a window opening to the *PN* junction. In modern practice these diodes are operated with a reverse bias of about 20 V, and act as sources delivering a current dependent on the incident light. A typical diode will deliver a current of 30 μA for an illumination of 2000 lux; in the absence of illumination its *dark current* will be 0·5 μA. There is a wide individual variation between individual diodes, and this must be allowed for in circuit design.

Photo-diodes are much faster in response than photo-transistors, and so may be used where rise time is a consideration; they can also be made much smaller than photo-transistors. However, they are suitable for use only at high levels of illumination. They are generally found in such applications as punched tape or card readers for use with computers.

18.8 The photomultiplier

The photomultiplier is an extremely sensitive and precise light-detecting device, and in modern practice is used for nearly all quantitative measurements of light.

The tube, which is highly evacuated, contains a photocathode, which is coated with a material that readily emits electrons when illuminated. This is followed by a chain of electrodes (*dynodes*) at progressively higher voltages, and at the far end is an anode, at the highest voltage of all. The geometrical arrangement of the electrodes varies, but the principle is shown diagrammatically in Fig. 18.2.

Any electron emitted for the cathode is attracted to the first dynode, which it hits with sufficient energy to expel several secondary electrons (up to about 10). These in turn are attracted to the second dynode, where each expels several electrons, and so on. The anode may finally collect many million electrons for each primary electron emitted from the photo-cathode. A total working voltage of about 1000 V is usual; the tube may

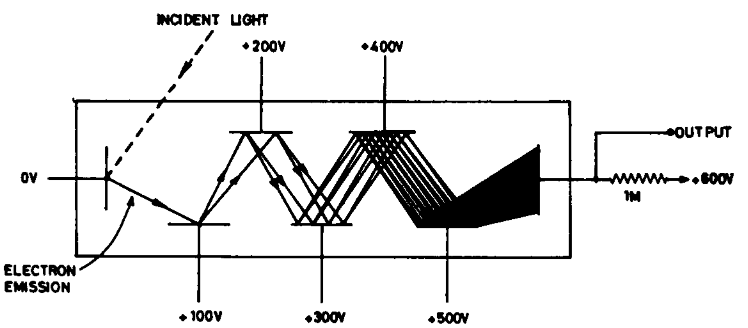


FIG. 18.2. Principle of photomultiplier tube.

contain up to about 14 dynodes. The electrons arriving at the anode represent a current flow through the anode load resistor; the anode voltage, and hence the output voltage, falls when the cathode is illuminated. The sensitivity is governed directly by the applied voltage, so highly stable supplies are required for quantitative measurement.

Photomultiplier tubes are extremely sensitive, and are operated in total darkness except for the admission to the photocathode of the small amounts of light to be measured. *A tube will be instantly destroyed if room lighting is permitted to enter while the e.h.t. is on.* Good photomultiplier tubes are very expensive.

18.9 Light pipes

Light may be transmitted for considerable distances, even round curves, by means of a transparent rod. The ends of the rod are polished, and the light fed into one end from a suitable source; a considerable proportion of the light will then be delivered from the other end. This transmission is due to the fact that if light travelling in a substance of higher refractive index strikes obliquely on the interface with a substance of lower refractive index, it will tend to be refracted towards the interface; if the angle is sufficiently oblique (greater than a certain *critical angle*) the light will be unable to leave the first medium at all, and will be totally reflected internally, as in Fig. 18.3. The higher the refractive index of the transmitting rod, the smaller will be the critical angle. A substance such as Perspex has quite a small critical angle, and is very suitable as a light pipe.

If the light pipe must be flexible, it may be made of a bundle of very thin uniform rods, usually of glass of high refractive index. Such *fibre optic bundles* are made commercially in a wide range of sizes and shapes. Typically they consist of some thousands of fibres, each of about 20 μm in diameter, enclosed in an external flexible sheath.

There are two types of bundle, *non-coherent* and *coherent*. In the non-coherent bundle, no attempt is made to keep each fibre in the same position in the bundle throughout its length; light is transmitted quite adequately, and cost is minimised. In the coherent bundle, the relationship between fibres is rigorously maintained; not only is light transmitted, but an image projected on one end is transmitted to the other. It is thus possible to use a dual pipe; one non-coherent bundle illuminates an object, and a second coherent bundle with a suitable lens system allows the object to be viewed, even round several

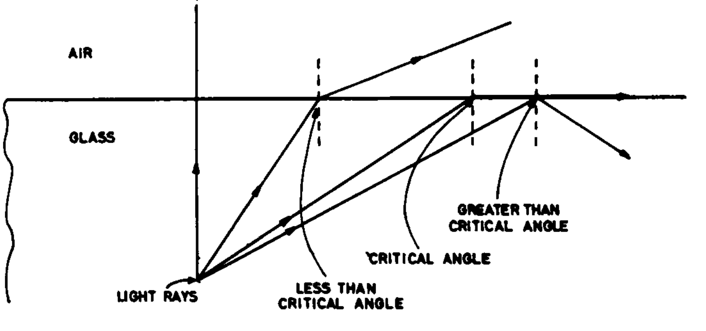


Fig. 18.3. Critical angle of incidence.

corners. Unfortunately the cost of light pipes, particularly coherent ones, is still very high (up to hundreds of dollars), but it is to be expected that reductions will occur with new manufacturing techniques.

18.10 Isolators

The use of a light-emitting diode optically coupled to a photo-transistor, either directly or through a light pipe, opens up a method of meeting the common circuit requirement of switching without direct electrical connection. This can of course be done by means of relays, but often these are too slow, too unreliable, or too poorly insulated to give the isolation required. Optoelectronic isolation is now very common. A number of manufacturers produce a range of isolators of this type, usually in a single standard transistor package.

18.11 Isolator for oscilloscope beam switching

One irritating feature of the student oscilloscope as constructed so far is the fact that except during a sweep there is always a bright spot at the left hand end of the screen, where the beam is resting. Such a spot is at best annoying; if a P7 phosphor is used, it will cause permanent damage, and if the screen is photographed it will produce halation on the film. It is clearly desirable to turn off the electron beam in the CRT except when it is wanted, and this may be done by increasing the bias on the CRT grid except during a sweep. Since the grid circuit of the CRT is at about -850 V with respect to earth, a direct connection from the time-base monostable multivibrator is not possible.

By the use of an isolator consisting of a light-emitting diode and a phototransistor this problem may be overcome, as shown in Fig. 18.4. The original EHT divider chain in the oscilloscope is replaced with one containing the isolator unit. The light-emitting diode is driven by an amplifier with an emitter follower input; the base of this is switched by the monostable multivibrator (refer also to Fig. 17.9). When light falls on the phototransistor in the MCT2 photon coupled isolator the 47 K resistor is effectively shorted, and the beam of the CRT is turned on. Otherwise the resistor is in circuit, and the extra drop across it provides sufficient bias to cut the beam off.

Further reading

Segallis, W., Solid state optoelectronics, *Electronic Products*, **28**, 1969.
Ballmer, J. E., Fiber optics, *Industrial Research*, October 1963.
RCA Phototubes and Photocells, Tech Manual PT-60, RCA, Lancaster, Pa., 1963.
Motorola Application Note AN440, Theory and Characteristics of Phototransistors.
Motorola Application Note AN508, Applications of Phototransistors in Electro-optic Systems.
The commercial literature is at present the best source of information on these topics; for light emitting diodes and photocells, refer for example to that by RCA, Fairchild, Texas Instruments, Hewlett-Packard, Monsanto, Philips, and for fibre optics, refer for example to that by Bausch and Lomb.

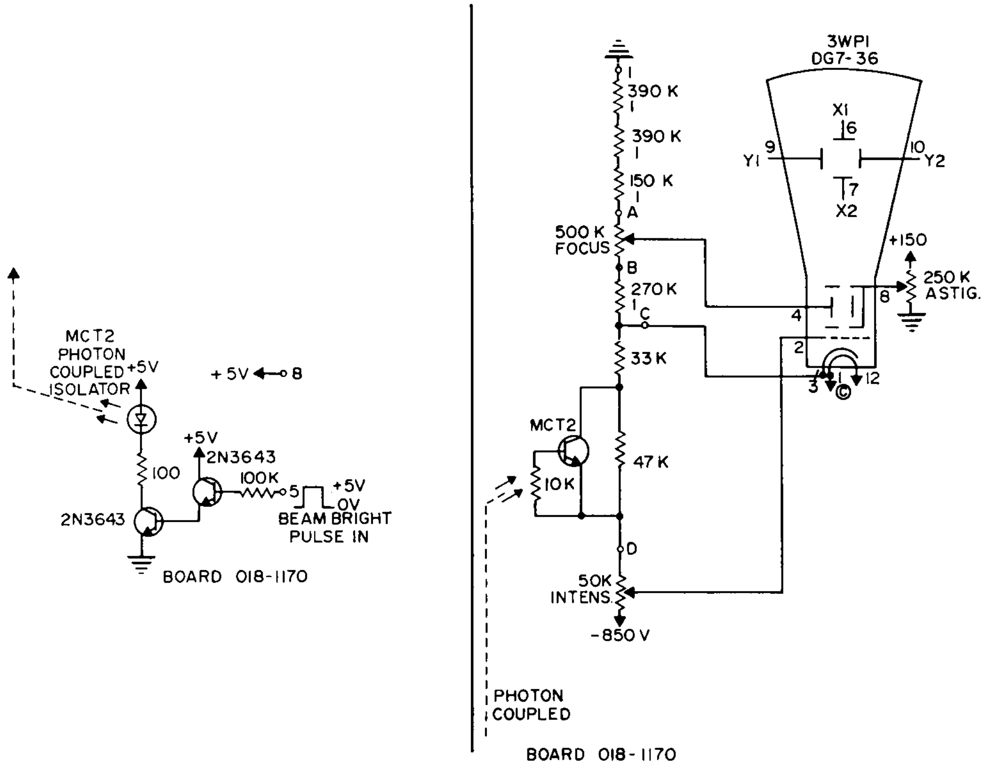


FIG. 18.4. Optoelectronic beam switch.

PRACTICAL

18.1 Connect up the optoelectronic beam switch of Fig. 18.4, using the card layout shown in Fig. 18.5, and install it in the oscilloscope. Switch on, and in the absence of a trigger pulse into the oscilloscope set the intensity control so that the spot at the left hand side of the screen just disappears. Set the time base to its slowest range, and trigger the oscilloscope externally with a battery. Ensure that the spot turns on and sweeps, vanishing at the end of the sweep. Test the oscilloscope with an applied signal from an oscillator, over the whole range of sweep speeds.

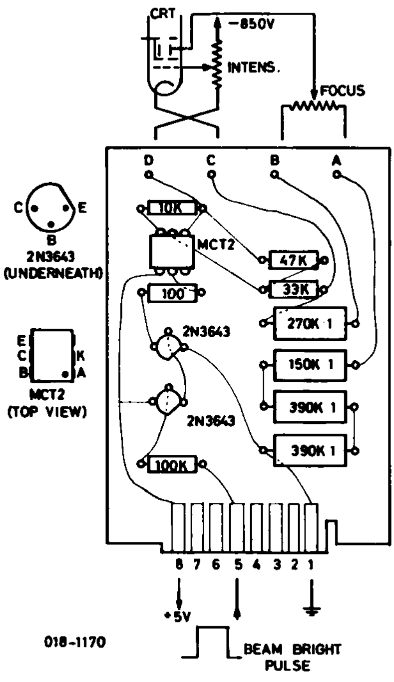


FIG. 18.5. Card for optoelectronic beam switch.

CHAPTER 19

BINARY LOGIC—GATES

19.1 Binary logic

The simplest form of decision or choice is that between two possible states; a switch may be closed or open, a light on or off, a man adjudged guilty or not guilty. This type of binary or two-valued situation was first studied by Aristotle; the rules governing combinations of binary choices and their logical consequences make up the set of *binary logic*. Binary logic is of immense importance in the field of electronics, since switches, lamps, and switching transistors all have two and only two possible states. Every modern computer operates as a binary system.

The rules of binary logic were formalised into a system of algebra by Boole, in the nineteenth century, and it is usual to employ Boolean algebra when discussing the behaviour of a switched system.

The two possible states in a binary system may be distinguished in several ways; in the case of a switch or lamp "on" and "off" are unambiguous. In the case of a network of switching transistors it is usual to standardise on two different voltage levels throughout the circuit; in modern practice these are 0 V and +5 V. There is no *a priori* reason to regard either of these as the "on" level, but generally the +5 V level is regarded as "on". A second term sometimes found for the "on" level is the *assertion level*. Another way of stating the two levels is in terms of the binary number system. Our familiar decimal number system, to the base 10, contains the *digits* 0 to 9; the binary system, to the base 2, contains only the *bits* 0 and 1. Accordingly we can call the "on" or assertion level the "logical 1" level, and the off level the "logical 0" level. It is less confusing to avoid the binary number system in discussing details of transistor switching, and to use the actual voltage levels.

19.2 Elements of Boolean algebra

In Boolean algebra, as in ordinary algebra, letter symbols are used to represent unknown or variable quantities; in the Boolean system, however, a variable can have only two possible values, 0 or 1. The symbol = has the usual meaning of "is equivalent to"; if the right hand side of a logical equation is 1, then the left hand side is also 1, and similarly for 0. Thus the statement

$$C = A, \tag{19.1}$$

means that C is 1 if A is 1, C is 0 if A is 0. Using a closed switch to give $A = 1$, and an on lamp to indicate $C = 1$ a circuit for which eqn. 19.1 applies is shown in Fig. 19.1.

The Boolean symbol of a bar over a variable negates that variable; thus $\bar{A}$ is read as

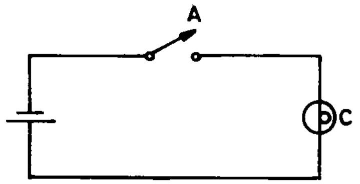


FIG. 19.1. Circuit for which $C = A$ applies.

"not-A". If $A = 1$ then $\bar{A} = 0$, and vice versa. More than one bar may be used; $\bar{\bar{A}}$ represents "not not-A", and clearly

$$\bar{\bar{A}} = A \tag{19.2}$$

With these symbols two or more mechanically coupled switches or relay contacts may be represented; in Fig. 19.2, for the two switches A and B, the statement $A = B$ means that they close and open together; the statement $A = \bar{B}$ means that when one is closed the other is open, and vice versa.

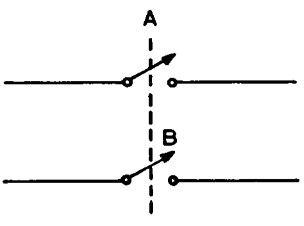


FIG. 19.2. Combination of two switches.

Where switches are joined in series or in parallel, it is necessary to have *logical connectives* in the corresponding algebraic expression. The symbol . is read as "and", and the symbol + as "or". For two switches A and B *in series*, the output C is given by

$$C = A.B \tag{19.3}$$

i.e., C is 1 if and only if both A and B are 1. For switches in parallel, the corresponding equation is

$$C = A + B \tag{19.4}$$

i.e. C is 1 if *either* A *or* B or *both* are 1.

The effect of these two logical connectives can be shown by means of *truth tables*, as follows: these express the results of all possible combinations of A and B.

TABLE 19.1

AND			OR		
A	B	C	A	B	C
0	0	0	0	0	0
0	1	0	0	1	1
1	0	0	1	0	1
1	1	1	1	1	1

From the foregoing some interesting relationships arise; the reader should satisfy his own mind as to their truth.

TABLE 19.2

		Starting from $A.B = C$	Starting from $A + B = C$
If	$B = A$, then	$A.A = A$	$A + A = A$
	$B = \bar{A}$	$A.\bar{A} = 0$	$A + \bar{A} = 1$
	$B = 1$	$A.1 = A$	$A + 1 = 1$
	$B = 0$	$A.0 = 0$	$A + 0 = A$
	$B = A = 0$	$0.0 = 0$	$0 + 0 = 0$
	$B = A = 1$	$1.1 = 1$	$1 + 1 = 1$
	$B = 0, A = 1$	$0.1 = 0$	$0 + 1 = 1$

Using these results, it is possible to simplify a complex switching network, reducing it to its simplest form before attempting to draw it and connect it up. This process is aided by realising that, fortunately, the commutative, distributive, and associative laws of ordinary algebra apply in Boolean algebra (Table 19.3). These should again be

TABLE 19.3

Commutative	Distributive	Associative
$A.B = B.A$	$A.(B + C) = A.B + A.C$	$(A.B).C = A.(B.C)$
$A + B = B + A$		$(A + B) + C = A + (B + C)$

verified by the reader by considering what they mean in terms of combinations of switches. Finally, a very considerable simplification of a more complex system can often be effected by the use of de Morgan's theorems:

$$\overline{A.B} = \bar{A} + \bar{B}, \tag{19.5}$$

$$\overline{A + B} = \bar{A}.\bar{B}. \tag{19.6}$$

If these are read saying "butter" for A, and "jam" for B, their truth will be immediately obvious.

There are sixteen possible truth tables that can be drawn up for combinations of A, B and C; the logical connectives AND and OR represent two of these. Three others are given names; these are NAND, NOR, and EXCLUSIVE OR. NAND and NOR are the combinations NOT AND ($\overline{A.B}$) and NOT OR ($\overline{A+B}$) used in de Morgan's theorems, and occur commonly in transistor switching circuits. EXCLUSIVE OR is given by

$$C = A.\bar{B} + \bar{A}.B, \tag{19.7}$$

and differs from the ordinary OR, which is

$$C = A.\bar{B} + \bar{A}.B + A.B \tag{19.8}$$
$$= A + B.$$

The ordinary or inclusive OR includes the case of $C = 1$ when $A = 1$ *and* $B = 1$; the exclusive OR does not. In ordinary language "or" is used for both, and the meaning is understood conventionally; compare

(i) "Would you like tea or coffee?" (EXCLUSIVE)

(ii) "Would you like milk or sugar?" (INCLUSIVE)

19.3 Diode logic

Quite complex logical functions are often called for in equipment for industry, in the computer field, and in the medical electronic field; these usually consist of the requirement that something shall occur if and only if certain combinations or circumstances prevail. After reducing the requirements to the simplest form by means of Boolean algebra, it is necessary to implement them by means of logical switching.

The simplest form of logical switching is by the use of diodes; a two-input diode *gate* is shown in Fig. 19.3. If initially both inputs are at 0 V, neither diode is conducting, and the output is at 0 V. If now *either* input rises to +5 V, or *both* inputs rise to +5 V, the corresponding diodes will conduct, and the output will rise to +5(−0·6) V; that is, the circuit has performed an OR function for a +5 V assertion level. If on the other hand both inputs are initially at +5 V, both diodes are conducting, and the output is at 4·4 V.

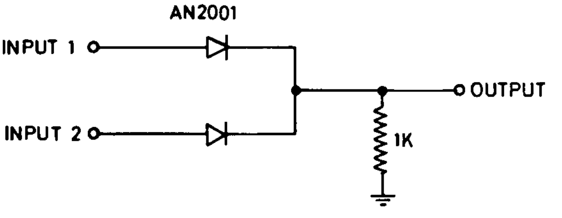


Fig. 19.3. Two-input diode gate.

Driving either input separately to 0 V will not affect the 4·4 V at the output, since the other diode is still conducting; driving *both* inputs to 0 V will take the output to 0 V. That is, the circuit has performed an AND function for a 0 V assertion level. This principle can be extended to a diode gate with any desired number of inputs. Moreover, it will work with either steady voltage levels or pulses, so is commonly used where it is desired to switch a pulse on or off by means of a voltage level. If input 2 is held at +5 V, and input 1 is initially at 0 V, 5 V positive pulses applied to input 1 will do nothing to the output, which is already at 4·4 V. If however input 2 is switched to 0 V, subsequent 5 V positive pulses applied to input 1 will appear at the output as 4·4 V pulses.

19.4 Diode transistor logic (DTL)

Because of the loss of voltage over a simple diode gate, it is not possible to interconnect a chain of such gates to perform complex logical functions. By following the diode gate with a transistor, as in Fig. 19.4, amplification is obtained; the two output voltage levels are the same as the input voltage levels, and a considerable number of gates may be interconnected. Since a transistor is used, the output is always inverted with respect to the input. For positive (+5 V) assertion at the input, the diode OR thus becomes NOR; for negative assertion (0 V) at the input, the diode AND becomes NAND. The operation is otherwise identical with the simple diode gate.

The circuit of Fig. 19.4 is described as a diode-transistor logic gate, or DTL gate for short.

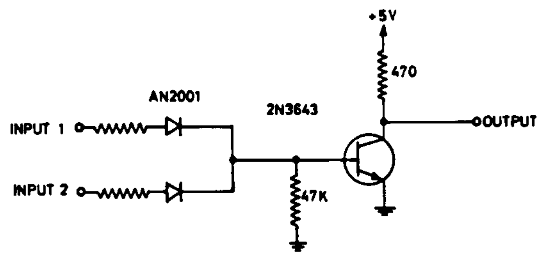


Fig. 19.4. Simple diode-transistor NOR gate.

19.5 Transistor–transistor logic (TTL)

DTL gates are quite suitable for switching operations at speeds up to about 2 MHz, but modern requirements call for higher rates, and also for reduced power consumption. To meet these needs manufacturers have replaced DTL with TTL integrated packages. The circuits used are quite involved, and will not be discussed here. They carry out their logical functions in precisely the same fashion as DTL, and are invariably used in new equipment. Probably the most widely used units are those of series 74, originally produced by Texas Instruments, and now manufactured by a number of companies. The SN7400 quad NAND gate, incorporating 4 positive assertion 2-input gates in the one package, is incorporated in the logic laboratory unit described below.

19.6 Standard symbols

It is usual to represent complete gates by standard conventional symbols in circuit drafting. It is possible to retain positive assertion throughout a circuit, and to use NAND or NOR gates as required; more commonly the one type of gate (usually NAND) is used throughout, and a negative assertion level used whenever the NOR function is required. Since positive assertion *into* a gate always produces negative assertion *out*, it may be necessary to use a gate with a single input to invert the assertion level where required. By the use of a little ingenuity the use of inverting gates may be minimised.

Fig. 19.5 shows the standard symbols used to represent positive assertion 2-input gates; any number of inputs to a gate is of course possible. The behaviour of a 2-input NAND gate for positive and negative assertion levels is shown in Fig. 19.6.

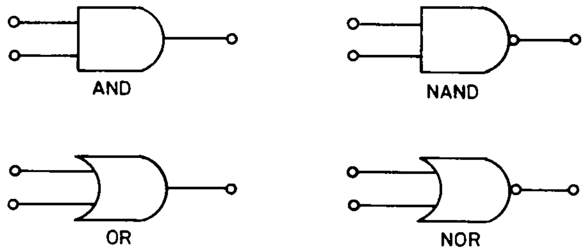


FIG. 19.5. Standard 2-input gate symbols.

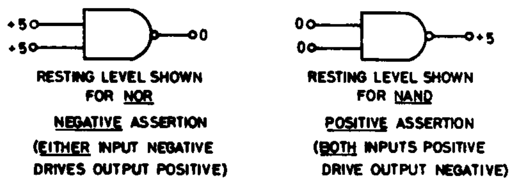


FIG. 19.6. Use of NAND gate for NAND and NOR functions.

19.7 LOGIC LABORATORY UNIT

It is possible to construct a very simple training unit to obtain practical experience in the use of gates; the design of such a unit is shown in Fig. 19.7. The unit operates on an external 5 V supply, and incorporates four indicator lamps with their transistor drivers, four 2-input TTL gates, and four bistable multivibrators (to be discussed in Chapter 20). Interconnections are made by means of leads terminated in banana plugs, as required for each experiment. Details of construction are given at the end of Chapter 20.

FURTHER READING

GE Transistor Manual (8th ed.), General Electric Company, Syracuse, N.Y., 1966.
Logic Handbook (1969 ed.), Digital Equipment Corporation, Maynard, Mass.
HERBST, L. J., *Discrete and Integrated Semiconductor Circuitry*, Chapman & Hall, London, 1969.
Texas Instruments Inc., *Designing with TTL Integrated Circuits*, McGraw-Hill, New York, 1971.
TTL Integrated Circuit Catalog CL401, Texas Instruments, Dallas, Texas, 1971.
Application notes by Texas, Fairchild, Signetics, Motorola, RCA, etc.

PRACTICAL

19.1 (a) By sketching the arrangement of switches represented and (b) by constructing truth tables, verify the following Boolean identities:

$$A + A.B = A$$

$$A + \bar{A}.B = A + B$$

$$A.(A + B) = A$$

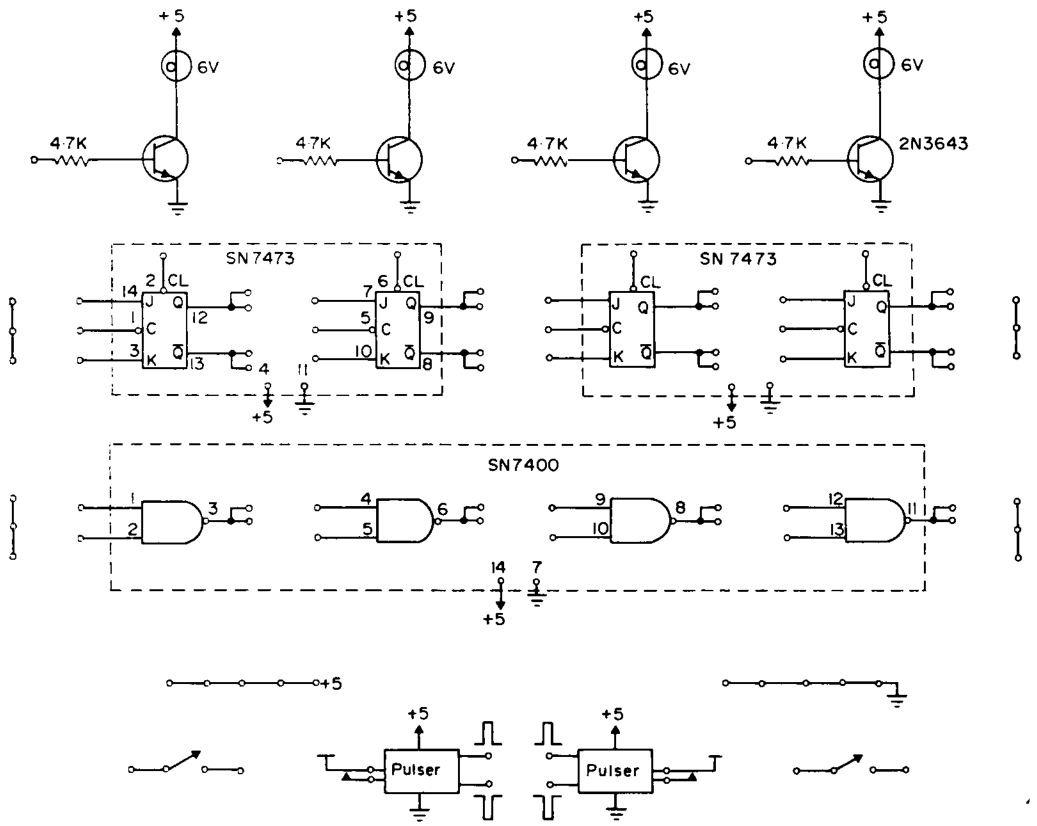


FIG. 19.7. Design of logic laboratory unit.

19.2 Using the results of section 1 above, prove by Boolean algebra that

$$\bar{A}.B + A.\bar{B} + A.B = A + B$$
$$\bar{A}.\bar{B}.\bar{C} + \bar{A}.\bar{B}.C = \overline{A + B}$$

19.3 Sketch an arrangement of TTL gates to calculate C from the following equation, given A and B. Do not simplify.

$$C = \bar{A}.B + A.\bar{B} + A.B$$

19.4 When Rabbit said, "Honey or condensed milk with your bread?" Pooh was so excited that he said, "Both", and then, so as not to seem greedy, he added, "But don't bother about the bread, please."

Winnie-the-Pooh A. A. Milne

Using the symbols H, C, and B for honey, condensed milk, and bread, write in Boolean form Rabbit's offer as intended by Rabbit, and Rabbit's offer as understood by Pooh.

19.5 Using the logic laboratory unit, and sketching each circuit before wiring it

(i) verify that the lamps function for positive assertion

(ii) connect a gate and lamp to perform the functions

$$C = \overline{A}$$
$$C = \overline{A.B}$$
$$C = \overline{A + B}$$

(iii) connect a gate to be operated by a voltage level, to determine whether a positive assertion pulse will or will not pass.

(iv) connect a combination of gates to perform the functions

$$C = A.B$$
$$C = A + B$$
$$C = A.\overline{B}$$

providing positive assertion at both inputs and output.

CHAPTER 20

BINARY LOGIC—DATA STORAGE AND COUNTING

20.1 Need for data storage

It is often required to use a pulse to start some particular operation in a piece of equipment, and for this operation to be continued indefinitely until a further pulse terminates it. Such a requirement can be met only by providing some type of memory device in the equipment, and the bistable multivibrator is ideal for this purpose.

20.2 The bistable multivibrator

The bistable multivibrator (also called the flip-flop circuit) differs from the monostable and astable versions in the use of two direct couplings, from each collector to the opposite base, as shown in Fig. 20.1. The circuit is completely symmetrical; however, on switching on there will always be sufficient asymmetry to cause one transistor to go to the saturated state, and the other to the cut off state.

The circuit has two trigger inputs, one conventionally called "*direct set*", and the other "*direct clear*". It has two outputs, one from each collector, conventionally marked Q and $\bar{Q}$.

Suppose that, on switching on, the circuit comes to the state where Q1 is cut off and Q2 is saturated. The Q output will be at +0·1 V (approximately 0 V, logical 0 for positive assertion) and the $\bar{Q}$ output at +4·2 V (approximately +5 V, logical 1 for positive assertion). This is conventionally the *clear* state of the flip-flop, the state where Q is not asserted.

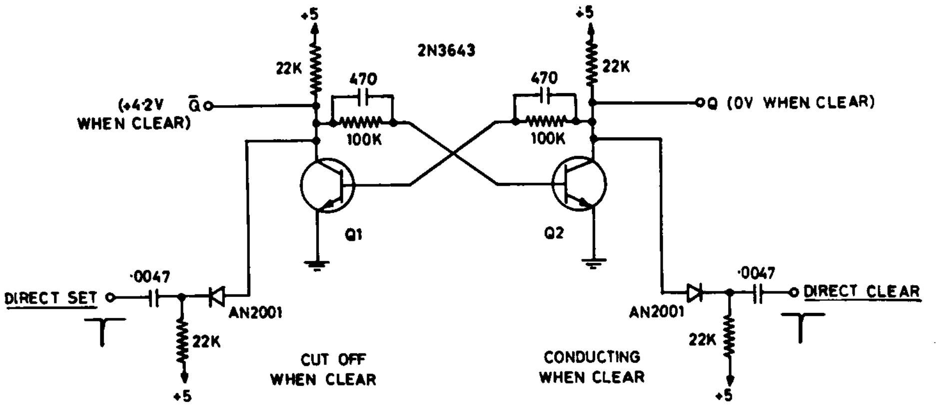


FIG. 20.1. Basic bistable multivibrator.

Under these conditions the base of Q1 will be held at +0·1 V by the collector of Q2, holding Q1 cut off. The collector of Q1 will draw no current, and the base to emitter diode of Q2 will be forward biased by the current from the +5 supply through the 22 K and 100 K resistors in series: Q2 will thus be held saturated. All four voltages will be quite stable, and the circuit will remain in this state indefinitely.

If now a negative-going trigger pulse is applied to the "direct set" input, it will start to cut Q2 off; the collector voltage of Q2 will rise, this will be transferred to the base of Q1 to turn it on, and a cumulative action will result. Q1 will very rapidly saturate, and Q2 turn off, and a second stable state will be reached, in which the Q output is at +4·2 (logical 1) and the $\overline{Q}$ output at +0·1 (logical 0). The flip-flop is now conventionally said to be *set*, and will remain in this condition indefinitely.

By the application of a negative-going trigger pulse to the "direct" clear input, the circuit may be changed back to the clear state.

A flip-flop may be regarded as a binary device, capable of "remembering" the value of a single bit in the binary number system. In various modifications it is the basis of all fast data storage in computers and many similar devices.

20.3 Gated flip-flops

As provided in the form of an integrated circuit, a flip-flop is usually supplemented by the addition of a logic gate on each input. This greatly increases its versatility, for it means that pulse inputs to a flip-flop can be enabled or disabled remotely, very often by the outputs of other flip-flops in the circuit. This gives the arrangement shown schematically in Fig. 20.2. If the gates give the NAND function for positive assertion, no trigger pulse at the input marked "gated set" or "gated clear" can pass unless the corresponding "S" or "R" gate input is at the positive assertion level. It is not uncommon to provide a continuous stream of pulses (usually called "*clock*" pulses) to both "gated set" and "gated clear" terminals; this means that the flip-flop will respond by being set or cleared, depending on whether the "S" or the "R" terminal is enabled. This whole device is commonly called an *RS flip-flop*.

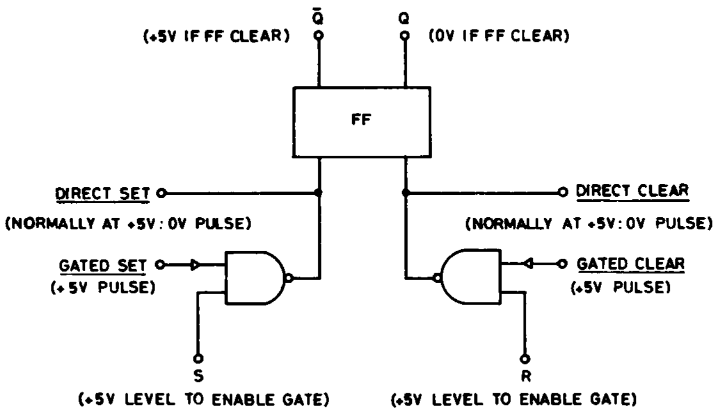


Fig. 20.2. Gated RS flip-flop.

The RS flip-flop has one major defect: what happens if clock pulses arrive while *both* S and R are enabled simultaneously? The result in this case will be indeterminate. This difficulty is overcome by the addition of a third input to each of the NAND gates, these inputs being supplied from $\overline{Q}$ and Q, as in Fig. 20.3. If J and K are both asserted, and the flip-flop is in the clear state, Q is at 0 V, and the "clear" NAND gate is disabled. $\overline{Q}$ is at +5 V, and the "set" NAND gate is enabled. If clock pulses now arrive simultaneously at both gates, only that into the "set" gate is effective, and the flip-flop switches over. Now the "clear" gate is enabled, and the "set" gate disabled, and the reverse condition holds. The ambiguity of the simple RS flip-flop is removed. A flip-flop provided with these *conditioning* gates is conventionally called a *JK flip-flop*, and is the type

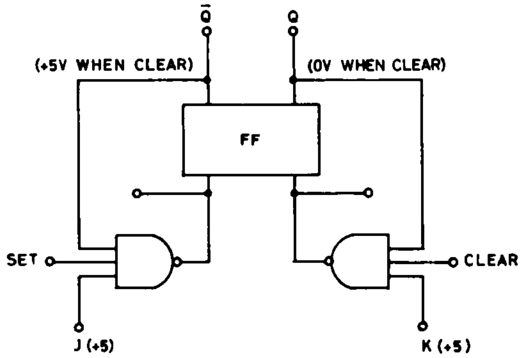


FIG. 20.3. Basic JK flip-flop.

normally used. The conditioning connections are not usually drawn; the "gated set" and "gated clear" inputs are joined internally, and the resulting circuit appears as in Fig. 20.4. A direct clear and often a direct set input are also commonly provided, and two JK flip-flops may be supplied in the one integrated circuit package. The Texas Instrument SN7473 is a typical TTL package, and is used in the logic laboratory unit described in Chapter 19. Its internal connections are somewhat more elaborate than those discussed above, to give reliable triggering on clock pulses of any shape whatsoever. It triggers on the *trailing* edge of a clock pulse.

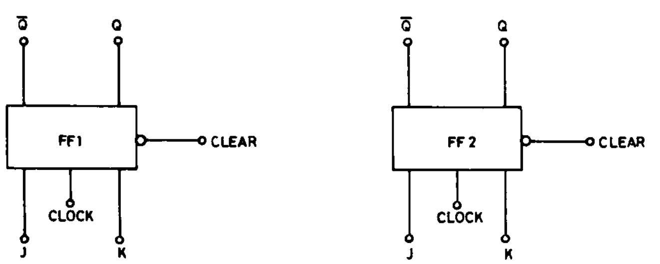


FIG. 20.4. Texas Instruments SN 7473 dual JK flip-flop.

20.4 Flip-flop as a frequency divider

If the J and K connections are both asserted by connection to a +5 V supply (J = K = 1), and a train of pulses is fed into the clock input, a JK flip-flop will turn on and off with alternate pulses; it is said to *toggle*. It will produce at its Q output a train of rectangular pulses at a frequency half that of the incoming pulse train.

20.5 Binary number system

It is conventional to write a decimal number from left to right, the position of each digit implying its power of 10; thus

$$563$$

is short for

$$5 \times 100 \quad + \quad 6 \times 10 \quad + \quad 3 \times 1,$$

or

$$5 \times 10^2 \quad + \quad 6 \times 10^1 \quad + \quad 3 \times 10^0.$$

In the same way it is conventional to write a binary number from left to right in descending powers of 2; thus

$$101$$

is short for

$$1 \times 2^2 \quad + \quad 0 \times 2^1 \quad + \quad 1 \times 2^0,$$

or

$$1 \times 4 \quad + \quad 0 \times 2 \quad + \quad 1 \times 1.$$

In decimal form this of course would be 5.

If there is any ambiguity as to the base which is actually in use when writing a number, the base may be subscripted after the number:

$$563_{10} \quad \text{but} \quad 101_2.$$

20.6 Binary counter

A chain of JK flip-flops may be set up to form a binary pulse counter; each flip-flop in the chain effectively divides by 2, and thus the state of the chain at any time represents the number of 1s, 2s, 4s, 8s, etc., that have occurred since the start, when it is assumed that all flip-flops were cleared. Fig. 20.5 shows a chain of four such flip-flops. The lamp indicators show the status of the corresponding flip-flops. The CLEAR line is normally connected to +5 V; a negative-going pulse on this line will reset all the flip-flops to zero, and extinguish all the lamps.

To understand the operation of the counter, it must be remembered that an SN7473 flip-flop is triggered by the *trailing* edge of an applied clock pulse. The first pulse to be counted sets FF1; the Q output of FF1 asserts, but does not trigger the clock input of FF2. The second incoming pulse clears FF1. In clearing its Q output goes to zero, forming a trailing edge, which sets FF2. The lamps now show 0010, or 2_{10}. The next pulse

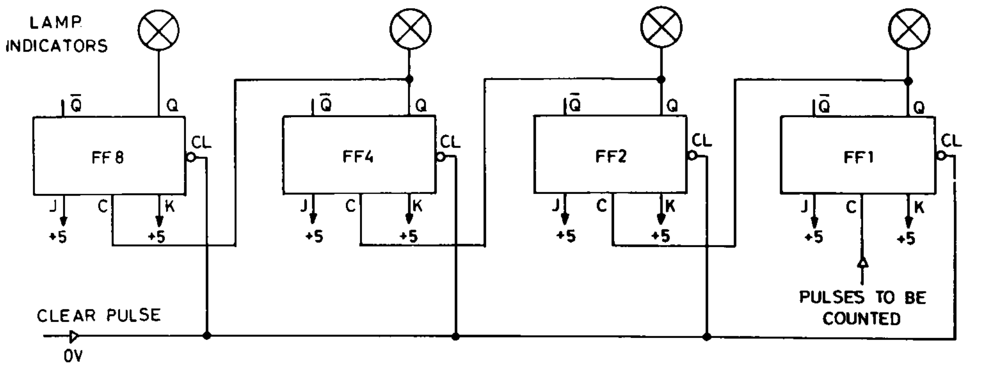


FIG. 20.5. Binary counter.

sets FF1, giving 0011, or 3_{10}. The next pulse resets FF1; the trailing edge of its Q output resets FF2. The trailing edge of the Q output of FF2 in turn sets FF4, and the lamps now read 0100, or 4_{10}. This process continues up to 1111, or 15_{10}; the sixteenth incoming count then resets all four counters, and the whole sequence recommences. As many binary places as may be required can be accommodated by adding more flip-flops.

Such a set of flip-flops is referred to as a *register*. Sets of flip-flops already connected together internally (usually four to a package) are available commercially.

20.7 BINARY CODED DECIMAL (BCD) COUNTING

It requires considerable practice to read a large binary number accurately from a set of lamps; moreover, the answer is usually required in decimal form. For this reason it is usual to arrange a counter in groups of four flip-flops. By means of internal connections within each group, the group is forced to reset to 0 and carry a pulse to the next group after a count of 9 rather than of 15. The resulting output on lamp indicators, four for each decimal digit, is much easier to read and interpret, and is also much more simply converted to a true decimal readout. It is said to be in *binary coded decimal* (BCD) form. A binary coded decimal decade of four flip-flops may be set up in a variety of ways; the simple arrangement shown in Fig. 20.6 is known as a ripple counter. The clear line to all flip-flops is not shown in Fig. 20.6 for simplicity.

Initially, assume that all flip-flops are clear		0000
On the first count,	FF1 is set	0001
On the second count,	FF1 is cleared; in clearing, its Q output sets FF2. (J of FF2 is enabled, since the $\bar{Q}$ output of FF8 is asserted.) The Q output of FF1 does nothing to FF8, since J of FF8 is disabled. (K of FF8 is enabled, but FF8 is already clear)	0010
On the third count,	FF1 is set	0011
On the fourth count,	FF1 is cleared; in clearing, its Q output clears FF2. As FF2 clears, its Q output sets FF4	0100
On the fifth count,	FF1 is set	0101

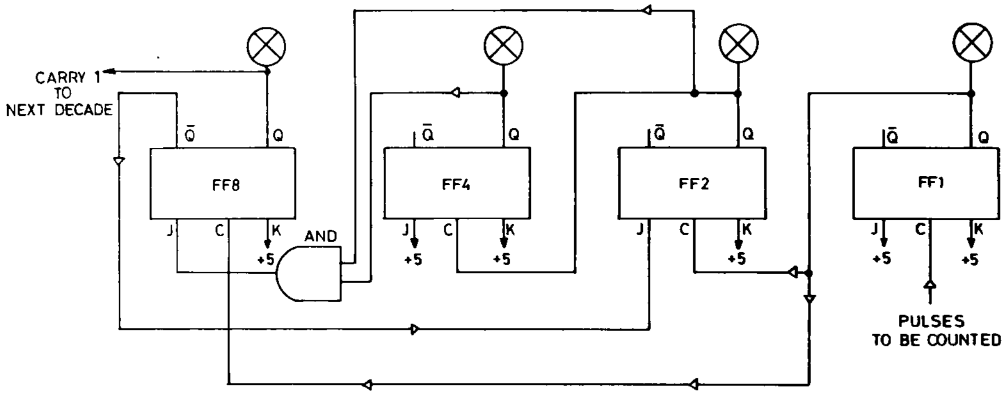


FIG. 20.6. BCD ripple counter.

On the sixth count,	FF1 is cleared; in clearing, its Q output sets FF2. FF2 and FF4 Q outputs are now both asserted, and the J input of FF8 is thus enabled through the AND gate	0110
On the seventh count,	FF1 is set	0111
On the eighth count,	FF1 is cleared; in clearing, its Q output sets FF8 and clears FF2. In clearing, the Q output of FF2 clears FF4. The $\bar{Q}$ output of FF8 goes to 0, which disables the J input of FF2	1000
On the ninth count,	FF1 is set	1001
On the tenth count,	FF1 is cleared. In clearing, its Q output cannot set FF2, because the FF2 J input is disabled. Since FF8 has previously been set, the Q output from FF1 clears it. In clearing, the $\bar{Q}$ output of FF8 enables the FF2 J input, and the Q output of FF8 provides a carry pulse to the next decade	0000

The cycle is then ready to recommence.

TTL BCD decade counters are available as single integrated circuit packages from most manufacturers.

20.8 BDC TO DECIMAL CONVERSION

While a BCD lamp display is considerably easier to read than a pure binary display, it is still quite unsuitable for the ordinary user. It is however relatively easy to convert, or *decode*, the four Q and four $\bar{Q}$ outputs from a BCD decade counter to produce a true decimal display. This is usually done by the use of ten 4-input AND gates, one for each of the digits 0 to 9. In Fig. 20.7, which shows the *diode matrix* required for decoding, each dot on a vertical line represents a diode input to the corresponding AND gate. For example, if the BCD output is 5, the BCD lines 1, 2, $\bar{4}$ and $\bar{8}$ will all be enabled. Only the "5" output AND gate will then be selected.

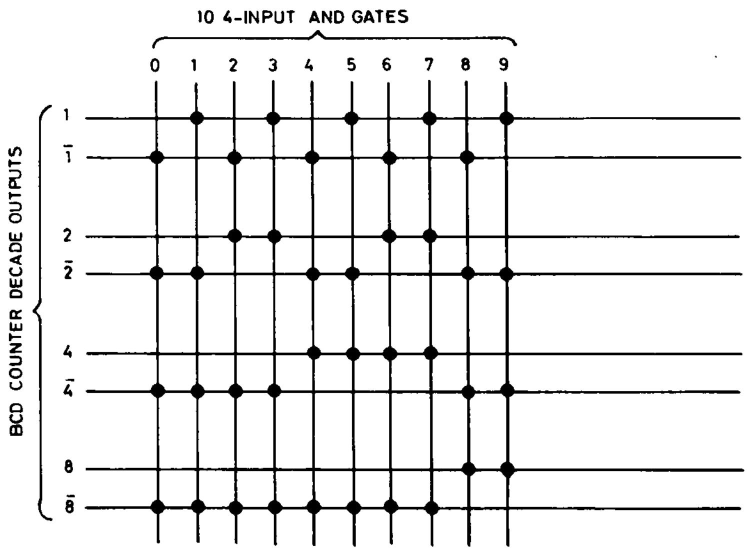


FIG. 20.7. BCD to decimal diode matrix.

The outputs of the ten AND gates may then be used to drive the digits 0 to 9 of a suitable indicator.

BCD-to-decimal decoders of this type are available as single integrated packages.

20.9 DECIMAL DISPLAYS

The most common forms of display in current use are all based on light-emitting diodes. The simplest is the seven-segment display, in which all the digits can be made up by selecting suitable combinations from a set of seven possible lines, as shown in Fig. 20.8. The lines in turn are formed from very small light-emitting diodes. The combinations are selected by a decimal-to-seven-segment diode matrix, which is often incorporated in the package containing the number display. It is also possible to obtain a simple package containing a complete BCD decade counter, a count-holding flip-flop register, the decoding matrices, and the display itself.

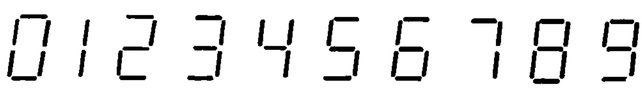


FIG. 20.8. Seven-segment digits.

A somewhat more realistic set of numerals, and also all the letters of the alphabet and punctuation marks, can be generated by the use of a *dot matrix* display. Each number or letter is represented by selecting a suitable combination from a rectangular array of dots, as shown in Fig. 20.9. Arrays of 6×4 and 7×5 dots are common. Even a 6×4 matrix limited to a numerical display obviously requires a considerably more elaborate decoding

FIG. 20.9. Dot matrix display.

matrix for each digit than does the seven-segment display, and consequently costs tend to be somewhat higher; suitable decoding systems are however readily available.

20.10 Data transfer between registers

If a binary number is stored in one flip-flop register, and it is desired to transfer it to a second register (which may be some distance away), the transfer is usually effected by a parallel connection from each flip-flop Q and $\bar{Q}$ outputs in the first register to the corresponding J and K inputs in the second, as shown in Fig. 20.10. A single clock pulse into all the second register clock inputs will then result in a transfer of the data. If a particular flip-flop in register 1 is *set* (contains a logical 1) its Q output is at +5 V, and its $\bar{Q}$ output at 0 V. The Q output will enable the J input of the corresponding flip-flop in register 2, and the $\bar{Q}$ output will disable the K input. A transfer pulse will then cause the register 2 flip-flop to *set* to a logical 1 also. Similarly a flip-flop in register 1 set to 0 will cause the transfer pulse to set the corresponding flip-flop in register 2 to 0. The process is described as a *jam transfer*, since the contents of register 1 are forced into register 2. The transfer requires one pulse only to carry it out, but a pair of lines for each bit to be transferred. Since each line will probably need to be a coaxial cable, transfer over long distances is very expensive. It is possible to use one line only for each bit to be transferred. The simplest way to do this is to use two successive pulses on register 2, as shown in Fig. 20.11.

The first pulse is applied to the direct clear line, resetting all the flip-flops in register 2 to 0. The transfer pulse then sets only those flip-flops connected to $Q = 1$ outputs of register 1.

To retain the advantages of a jam transfer with single line connections between flip-flops, it is possible to add inverting gates to the register 2 inputs; these reconstruct the

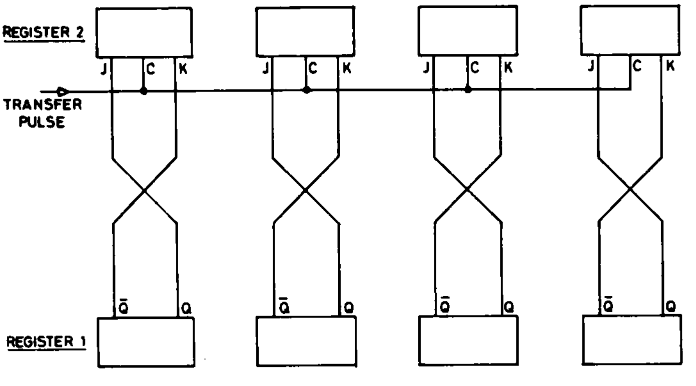


FIG. 20.10. Jam transfer between registers.

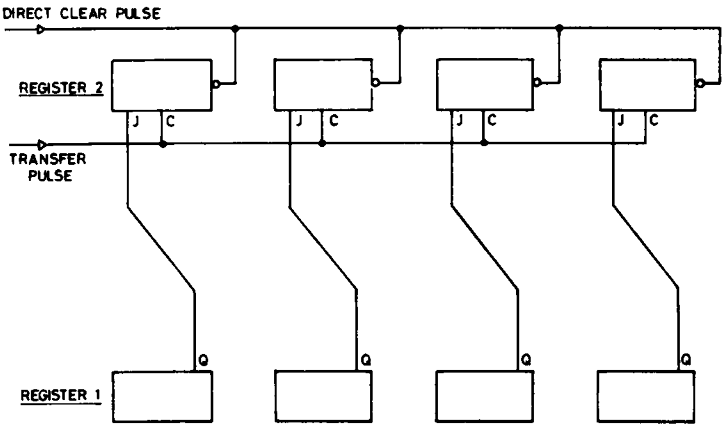


FIG. 20.11. Single line transfer between two registers.

$\bar{Q}$ level right at register 2, and are connected to its K inputs, as shown in Fig. 20.12. Notice the symbol for a simple inverting gate.

This requirement is so common that JK flip-flops including the inverter are available. If both inputs are brought out, as in Fig. 20.13(a), the arrangement is known as a J$\bar{K}$ flip-flop; if the inputs are connected internally, as in Fig. 20.13(b), as a D (for data) flip-flop. Obviously a J$\bar{K}$ becomes a D flip-flop if its J and $\bar{K}$ inputs are joined together.

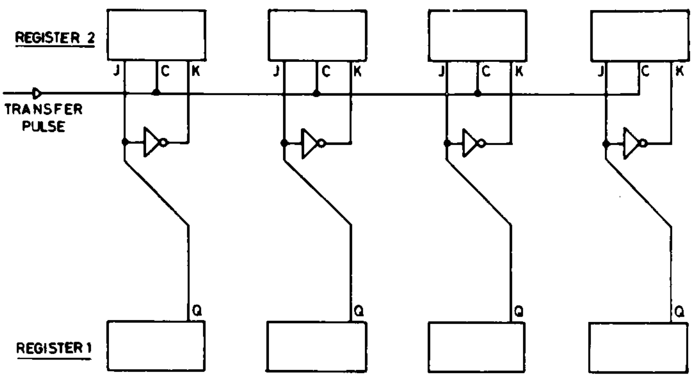


FIG. 20.12. Jam transfer with single line connections.

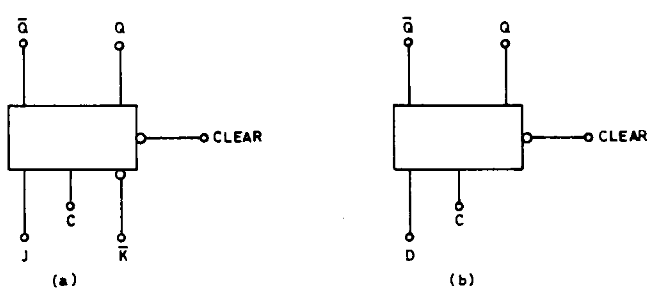


FIG. 20.13. Type J$\bar{K}$ and D flip-flops.

20.11 TTL astable and monostable multivibrators and Schmitt triggers

A number of manufacturers produce TTL integrated circuit versions of monostable multivibrators and Schmitt triggers; these are highly accurate and versatile, and in modern practice are always used in preference to discrete circuitry.

A typical dual Schmitt trigger unit is the Texas Instrument SN7413; one half of this unit is shown in Fig. 20.14. This unit operates on a supply voltage of 5 V.

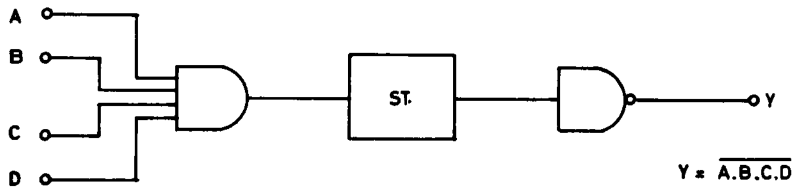


Fig. 20.14. Half of Texas Instruments SN 7413.

The actual trigger circuit is preceded by a four-input AND gate, and followed by an inverting gate, which renders its operation independent of external circuitry. (Unwanted inputs should be connected to +5 V.) The upper threshold is +1·7 V, and the lower +0·9 V, and these values are almost independent of temperature or supply voltage. It should be noted that the inputs must be driven from low-impedance sources, either other logic units or the outputs of operational amplifiers, to maintain accuracy of triggering.

A typical monostable multivibrator unit is the Texas Instruments SN74121; this is shown in Fig. 20.15. This device consists not only of a monostable, but its input incorporates a Schmitt trigger and OR gate; these features greatly extend its usefulness. The A1 and A2 inputs go into a negative-assertion OR gate; the monostable will trigger if either goes to zero, provided that the B input is held at +5 V. The B input is a positive-assertion Schmitt trigger input with a positive-going threshold of 1·55 V and a negative-going threshold of 1·35 V; the monostable will trigger if B goes more positive than 1·55 V after having been below 1·35 V, provided that at least one of A1 and A2 is held at zero. The B input must be driven from a low-impedance source, such as an operational amplifier.

Once the monostable is triggered, Q will go to +5 V and $\overline{Q}$ to 0 V, and they will remain so for a time determined by C and R, no matter what subsequently happens at

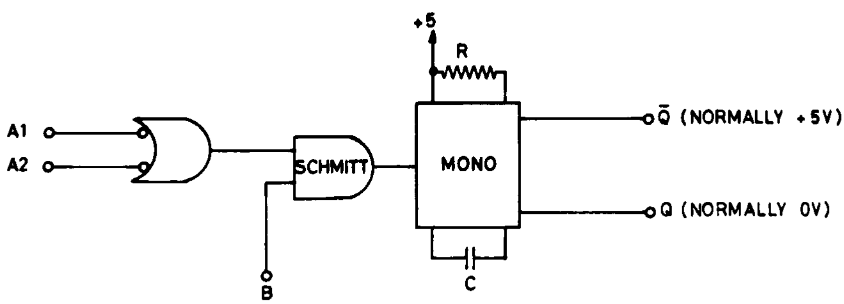


Fig. 20.15. Texas Instruments SN 74121.

the A1, A2, or B inputs. C and R are provided externally to set the output duration. R may be varied from 1·5 K to 40 K, and C from 10 pF to 1000 μF; this will give a delay ranging from 40 nsec up to 30 sec.

By using two SN74121 units, and connecting the Q output of each to the two A inputs of the other (and connecting the B output of each to +5) an astable multivibrator will be produced; its timing will be determined by the values of R and C used in each SN74121.

20.12 Medium and large scale integration (MSI and LSI)

Single integrated circuit packages containing very large numbers of interconnected flip-flops and gates are now being manufactured, to serve as data stores, counters, decoders, and arithmetic elements. It is quite usual to get 256 (2^8) bits of information on a single chip, and units of up to several thousand bits are available.

20.13 Details of logic laboratory

Fig. 20.16 shows the design of the main printed card for the logic laboratory, and Figs. 20.17 and 20.18 the circuit and printed card for the dual pulser unit, which converts the relatively slow and rather noisy pulses from the push buttons into fast clean pulses suitable to drive TTL logic. Fig. 20.19 shows a photograph of the complete unit.

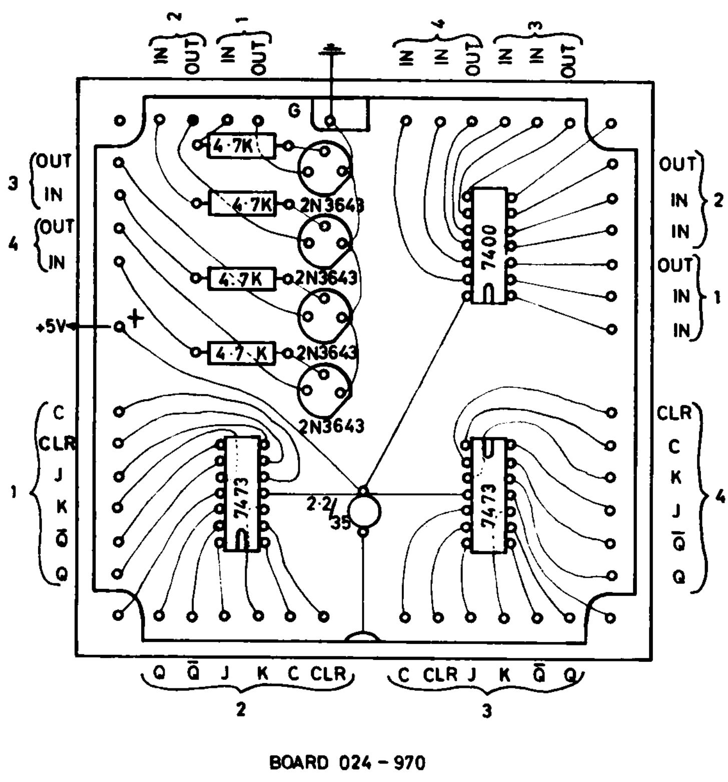


Fig. 20.16. Printed card for logic board.

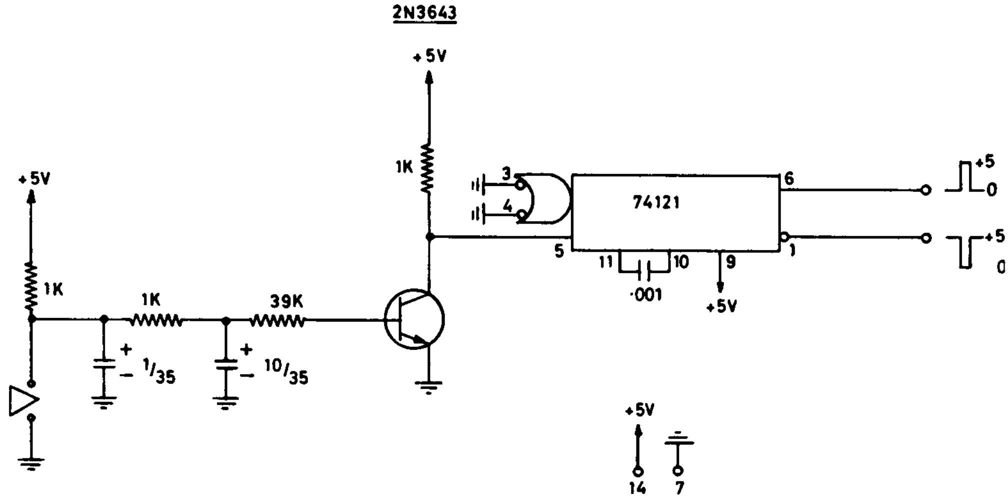
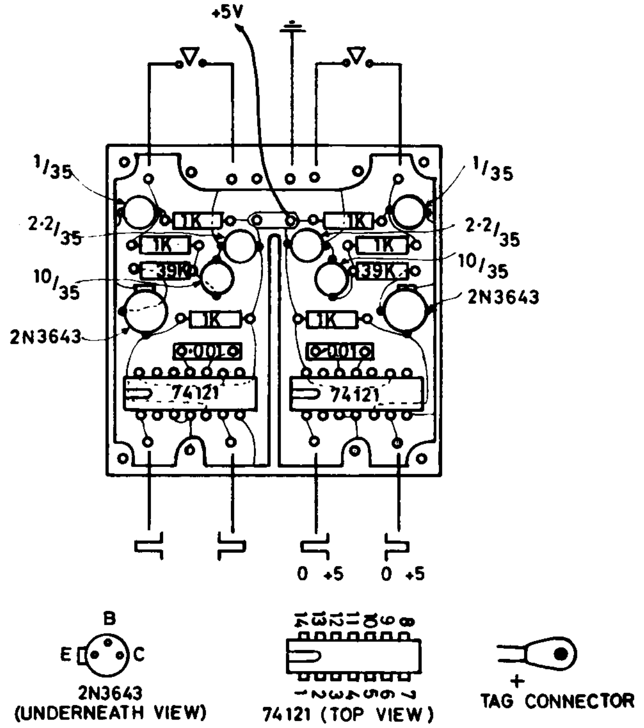


FIG. 20.17. Circuit of one pulser unit.



BOARD 061-772

FIG. 20.18. Printed card for dual pulser unit.

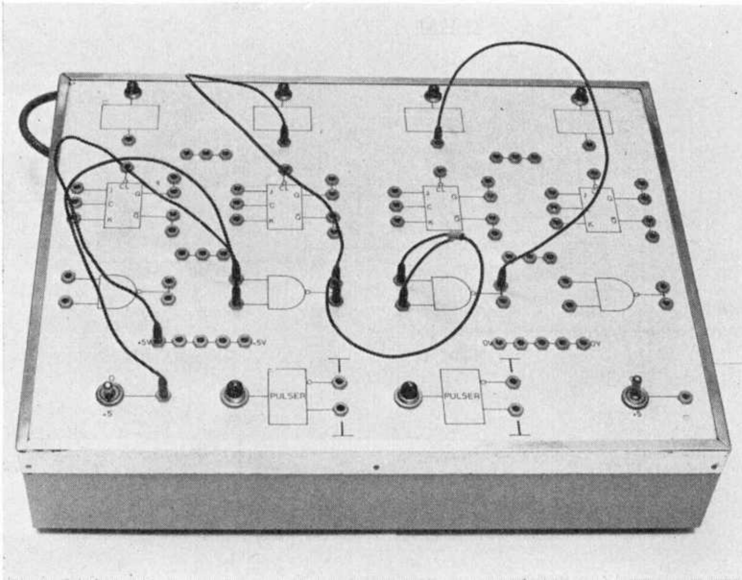


FIG. 20.19. Complete logic laboratory.

FURTHER READING

GE Transistor Manual (8th ed.), General Electric Company, Syracuse, N.Y., 1967.
HERBST, L. J., *Discrete and Integrated Semiconductor Circuitry*, Chapman & Hall, London, 1969.
TTL Integrated Circuit Catalog CC401, Texas Instruments, Dallas, Texas, 1971.
HUGHES, J. L., *Computer Lab Workbook*, Digital Equipment Corporation, Maynard, Mass., 1969.
Texas Instruments Inc., *Designing with TTL Integrated Circuits*, McGraw-Hill, New York, 1971.
Information on integrated circuits generally is available from a wide variety of manufacturers.

PRACTICAL

20.1 Using one of the JK flip-flops in the logic laboratory unit, connect indicator lamps to show the states of Q and $\bar{Q}$; connect J and K to +5 V, and drive its clock input with a series of pulses from a push button. Use the second push button to generate a clear pulse, and use it to reset the flip-flop. Test the effects of disabling either J or K or both.

20.2 Set up the four JK flip-flops as in Fig. 20.5, to form a 4-bit binary counter with facilities for clearing at any time.

20.3 Set up a BCD decade, using the circuit of Fig. 20.6, and verifying that it counts correctly. Include facilities for clearing at any time.

20.4 (For good students: this is a challenge!) A BCD decade counter is sometimes described as a modulo-10 counter. Design and test on the logic laboratory unit a modulo-7 counter.

20.5 Connect together two logic laboratory units so that one acts as a 4-bit counter, and the other can copy its contents. Use (a) simple jam transfer (b) clear and transfer (c) J$\bar{K}$ transfer.

CHAPTER 21

MEASUREMENT OF RADIOACTIVITY

21.1 Detection of radiation

The presence of a radioactive source may be detected by means of a device capable of responding to high-energy electrons. The two types of radiation normally encountered in biological radiochemistry are beta and gamma. Beta radiation is simply a stream of high-energy electrons, and is detected directly. It is characterised by a relatively short range in air, and is readily stopped by quite thin sheets of solid material. Gamma radiation is an electromagnetic wave of the same nature as light or X-radiation, but of very short wavelength. It is detected by the fact that it can produce secondary electrons if it impinges on any solid material. On passing through a detector, it causes the emission of electrons from the material of which the detector is made, and these are then measured. Gamma radiation is characterised by long range, and high penetrating ability through solid materials.

21.2 Rules for a radiochemical laboratory

Radioactive materials in the small amounts used for biological estimations may be handled with perfect safety so long as they are treated with respect. In ordinary chemical practice minor spills often go unnoticed; in the handling of radioisotopes a laboratory may easily become sufficiently contaminated with spilled radioactive material to materially reduce the accuracy of counting procedures, or even to render them impossible. Further, careless work will certainly produce hazards to health, particularly if radioactive material is ingested into the body.

The isotopes normally used for biological estimations have either a short half-life, or produce low energy or nonpenetrating radiation. With modern counting equipment of high sensitivity, they are also used in very small quantities. Nevertheless standard precautions should always be taken, since sooner or later more dangerous isotopes may be handled. Tables indicating the hazards involved in handling any type of isotope are readily available, and should be thoughtfully consulted before an unfamiliar isotope is introduced into the laboratory.

Apart from scrupulously following the general regulations laid down by law in every country, specific local rules should be enforced. Typical rules for a radiochemical laboratory are as follows:

1. No person untrained in the use of radioactive materials may handle any such material except under direct supervision and in the presence of an approved operator.

2. Specific areas are set aside for manipulation of radioisotopes, and clearly marked. Except in approved sealed containers, radioactive materials must not be handled elsewhere.
3. A film badge will be issued to each person entering the radioisotope area; it must be worn at all times while in the area, and returned for checking at the scheduled time.
4. Smoking, eating and drinking are not permitted in the radioisotope area.
5. No glassware or other material may be removed from the radioisotope area unless it has been monitored for radioactivity by a responsible person. No operator may leave the area without being checked for radioactivity by a responsible person; the hands should be most carefully checked, and if traces of contamination are apparent, must be repeatedly scrubbed until they are clear.
6. Fluids must never be pipetted by mouth.
7. All manipulations must be carried out on several thicknesses of absorbent paper, which must afterwards be monitored for radioactivity.
8. At the end of a manipulation all active materials must be disposed of as directed for the isotope concerned, and the whole working area monitored for possible contamination.
9. Radioisotopes not actually being manipulated must be kept in an approved storage; gamma radiators should be kept behind a barrier of lead bricks even on the workbench.
10. Accidents in handling radioisotopes must be promptly reported and decontamination procedures carried out at once.

21.3 The Geiger–Müller counter

The Geiger counter was the first successful instrument developed for detection and estimation of radiation. Though not highly sensitive, it is compact and portable, and is still used as a monitor in the laboratory to detect contamination. By use of a suitable Geiger tube, either beta or gamma radiation may be detected. A Geiger counter can give no indication of the energy of the individual particles concerned.

The Geiger tube consists of a metal cylinder, through the axis of which passes a thin wire. This wire is insulated from the cylinder, and maintained at a high voltage with respect to it. The cylinder is filled with a mixture of argon with a trace of alcohol or halogen vapour, at a pressure of about one tenth atmospheric. For gamma counting, the cylinder is closed off with a metal end plate; for beta counting, the end is made of thin mica or aluminium foil. A typical arrangement is shown in Fig. 21.1.

A high energy electron passing through the tube collides with atoms of the gas filling, releasing electrons from them. These electrons are accelerated towards the central wire, and in turn ionise further gas atoms by collision. On collection by the central wire, the electrons cause a flow of current through the load resistor, and a change in potential across it; this change is transferred to a subsequent pulse detector. The rise in current through the tube would become cumulative if the tube were not *quenched*; the alcohol or halogen filling absorbs electrons, and produces a rapid decay in current once the original high energy electron has passed. As a result, a pulse of current flows for each electron that enters the

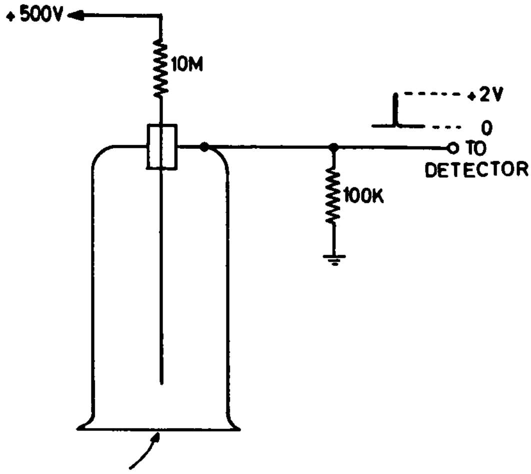


FIG. 21.1. Geiger tube and output circuit.

tube. In a correctly operated tube, the height of this pulse is independent of the energy of the original electron. After the production of a pulse, a period of 100 to 200 μsec is required for recovery before the tube can detect a further pulse; this is the *paralysis time*, and is one of the major disadvantages of the Geiger tube.

If the average pulse rate produced by the tube in the presence of a radio-active source is measured, while the e.h.t. voltage is slowly raised from zero, a curve of the form shown in Fig. 21.2 will be obtained. The tube is operated in the plateau region, where the count rate is least sensitive to minor changes in e.h.t. voltage. The position and shape of the plateau depend on the type of tube used, and the recommended operating voltage is usually indicated on each individual tube. If the plateau voltage is exceeded the tube will be destroyed.

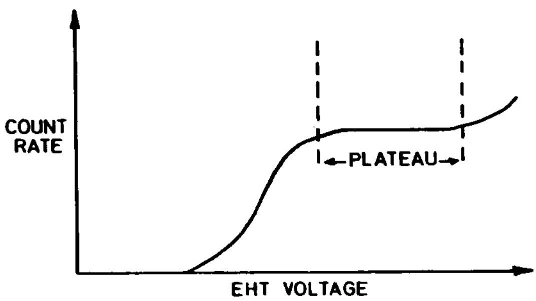


FIG. 21.2. Characteristic curve of Geiger tube.

21.4 BACKGROUND

In the absence of a deliberately introduced source of radioactivity, a counter will always indicate a low but definite count rate. This is due in part to cosmic radiation, in part due to radiation from natural radioactive materials and traces of nuclear fallout, and possibly also in part to contamination of the counting equipment or the adjacent bench. Regular checks should be made of this background level, which is subtracted from each actual count made. Background may be minimised by lead shielding about the Geiger tube, and by scrupulous care to avoid any spillage of radioactive material.

21.5 Equipment for Geiger counting

A simple unit for the detection of *gamma* radioactive contamination will be constructed in the practical session; this uses an earphone to detect individual pulses. For more quantitative work an estimate of the count rate is desired; *count rate* is proportional to *amount* of isotope present in the vicinity of the counter. A typical arrangement in a portable counter is shown in Fig. 21.3. Pulses from the Geiger tube are a few volts high. They are passed through a suitable amplifier, and then used to drive a *ratemeter*.

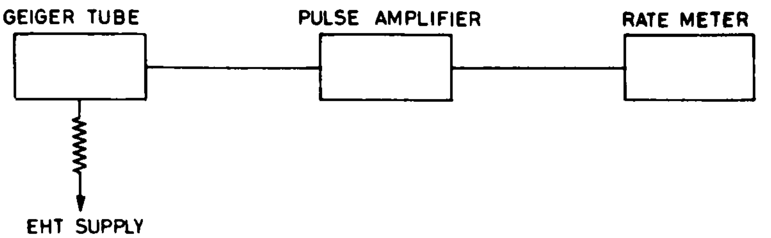


Fig. 21.3. Portable Geiger counter.

A typical ratemeter is shown in Fig. 21.4. This consists of a monostable multivibrator, an emitter follower driver stage, and a *diode pump* circuit, with values of C1, C2 and C3 suitable for a range of 0–100 pulses per second on the meter. On the arrival of a pulse, the monostable multivibrator is triggered; Q2, which is normally conducting, is cut off

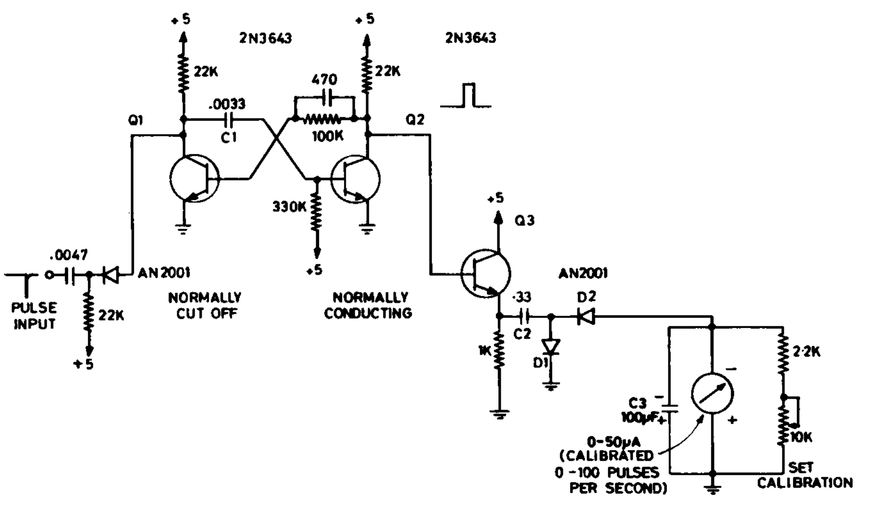


Fig. 21.4. Diode pump ratemeter.

and its collector rises to +4·2 volts. The emitter of the emitter follower Q3, which is normally cut off, rises to 3·6 V. D1 conducts, and C2 charges towards 3·0 V (3·6 V–0·6 V). D2 is reverse biased and cannot conduct. By the end of the delay time of the multivibrator (about 800 μsec), C2 is fully charged to 3·0 V. Now if a capacitor of C farad is charged to a voltage e volt, it will contain a charge q coulomb, given by

$$q = Ce. \tag{21.1}$$

The multivibrator now resets and the collector of Q2 goes to 0·1 V, and cuts off Q3. C3 and the meter shunt may be omitted for the moment, and the situation is as shown in Fig. 21.5. D1 is clearly cut off and may be neglected; D2 will conduct, and C2 will discharge through the 1 K resistor, the meter, and D2, giving an upward impulse to the meter pointer.

If the multivibrator receives f incoming pulses per second and C2 discharges completely each time, C2 will discharge fq (or fCe) coulombs per second through the meter. Since coulombs per second are amperes, the *average current* through the meter will thus be

$$I = fCe \tag{21.2}$$

Since both C and e are constant, I is proportional to f.

In fact, C2 will not discharge completely, because of the voltage required to make D2 conduct, but this will only lower the effective value of e somewhat.

For the values used, at 100 pulses per second

$$I = 100 \times 0{\cdot}33 \times 2{\cdot}4$$
$$= 79\ \mu\text{A}$$

By the use of a 50 μA meter and a suitable shunt, a full scale reading for 100 pulses per second may be obtained.

The meter will of course be receiving a series of impulses rather than a steady current. At 100 pulses per second the pointer reading will be quite steady, since the mechanical constants of the movement will constitute an effective low-pass filter. For lower pulse rates the pointer movement may become excessive; the addition of C3 will reduce this movement to an acceptable amount.

If a permanent record is required, a suitable recorder may be substituted for the microammeter.

Diode pump ratemeters are frequently used for purposes other than nuclear counting; frequency meters, cardiotachometers and respiration rate meters are common.

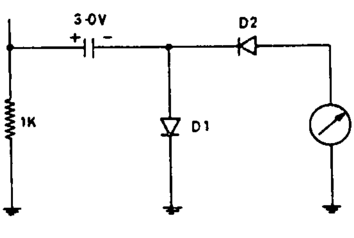


FIG. 21.5. Diode pump after end of delay period.

21.6 Simple Geiger counter

A very simple but quite effective Geiger counter may be constructed using the circuit of Fig. 21.6. Fig. 21.10 (at the end of chapter) shows the layout of the counter, and Fig. 21.11 the complete unit.

The Geiger tube used is a miniature gamma detector (Philips 18503 or its equivalent). Its output is detected by means of a crystal earphone connected directly across the 82 K tube load. The 500 V e.h.t. supply is obtained from a simple transistor oscillator, which converts the 3 V from a battery of two pencil cells to a series of pulses. When the 3 V supply is first turned on, the 2N3645 transistor is forward biased, and its collector current rises, flowing through the transformer. This in turn induces a rising voltage in the transistor base winding in a direction such as to cut the transistor off. The collector current now diminishes faster and faster, and presently cuts off abruptly, inducing a very large pulse (about 800 V high) in the transformer secondary. This charges the filter capacitor through the diode. With the collector current cut off, the transistor base is again forward biased, and the cycle repeats. The pulses generated are about 30 μsec long, and are at a frequency of about 3 kHz. Since operation anywhere in the plateau region is satisfactory, high stability of the e.h.t. supply is not required for operation of a Geiger tube.

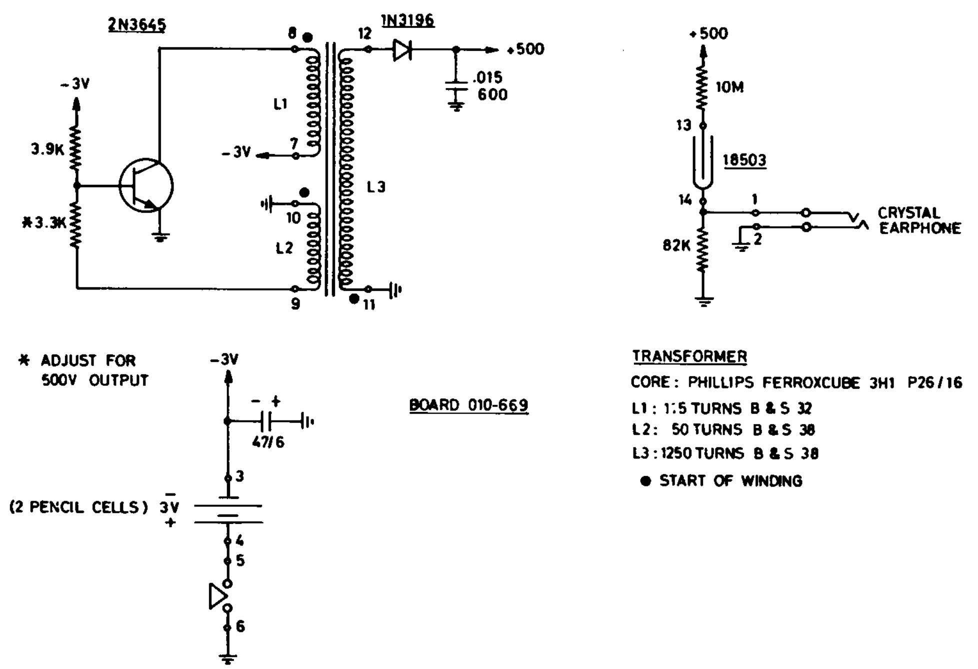


Fig. 21.6. Simple Geiger counter.

21.7 SCINTILLATION COUNTER

The scintillation counter depends on the fact that a flash of visible light is produced when an electron or burst of gamma radiation passes into a suitable phosphor; this phenomenon has already been discussed in connection with cathode ray tube screens. Solid phosphors for scintillation counting usually consist of large crystals of activated sodium iodide; liquid phosphors are made by dissolving substances like anthracene or stilbene in toluene or a similar solvent. Under suitable conditions the liquid phosphor may be added directly to the radioactive liquid to be studied.

Each flash of light from the phosphor is "seen" by a photomultiplier tube, giving a pulse at its output which has a duration of about 100 nsec, and a height proportional to the light emitted. This in turn is a measure of the energy of the original particle entering the crystal, provided that the crystal is large enough to bring the particle completely to rest, and so absorb all its energy. Unlike the Geiger counter, the scintillation counter produces a whole range of pulse heights from a typical radioisotope, and the relative frequencies of occurrence of these various heights are characteristic for any particular isotope. By drawing a graph of the relative abundance of the different pulse heights produced, an energy *spectrum* for the isotope can be drawn, as in Fig. 21.7. In Fig. 21.7(a) it will be seen that the most frequently occurring pulse heights are a 4·0 V and 9·5 V, and this is reflected by the peaks in Fig. 21.7(b).

To generate a curve such as Fig. 21.7(b) from the incoming pulses, it is necessary to amplify the pulses from the photomultiplier suitably, and then to *sort* them according to height. To do this successfully, the equipment must have a highly stable e.h.t. supply on the photomultiplier, an amplifier of constant gain, and an accurate *pulse height analyser*.

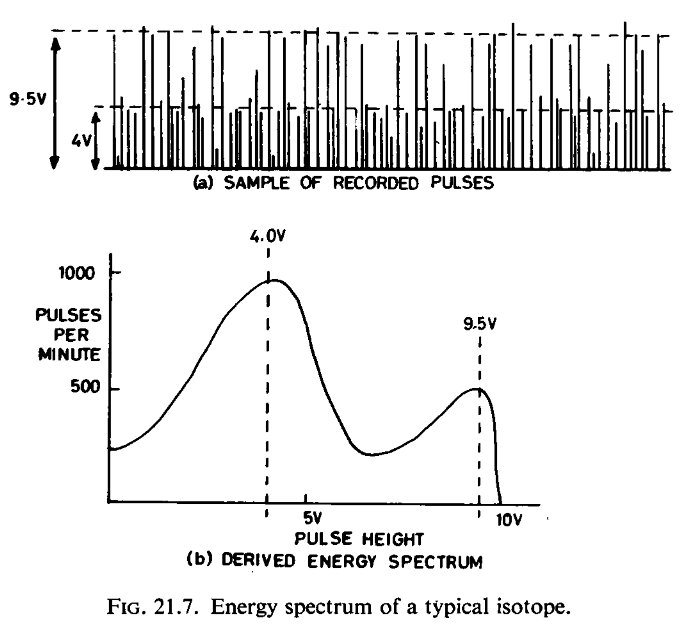


FIG. 21.7. Energy spectrum of a typical isotope.

21.8 Discriminator and pulse height analyser

It is possible to perform a radiochemical estimation by counting all the pulses detected by the photomultiplier, irrespective of height. Many estimations, and particularly those in the biological field, are performed with very small quantities of radioactive materials, and the count rate due to the isotope used may well be less than the background count rate unless some preliminary pulse sorting is carried out. The simplest form of sorting is by means of a discriminator, which rejects all pulses less than a known voltage. Since background pulses are predominantly of low voltage, a considerable improvement in isotope-to-background count ratio may be effected in this way. The circuit most commonly used is the Schmitt trigger, with its input arranged as in Fig. 21.8.

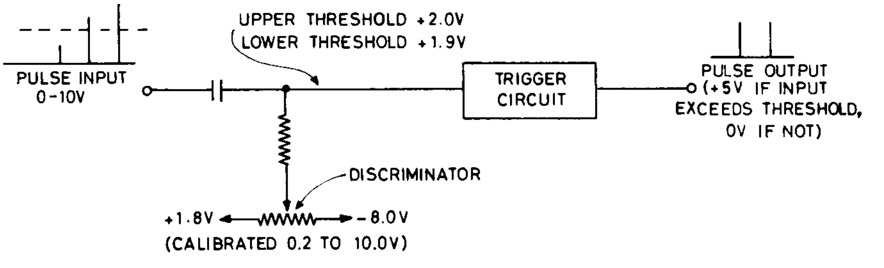


Fig. 21.8. Principle of discriminator.

Suppose that the trigger circuit has an upper threshold of +2·0 V, and a lower threshold of +1·9 V, as shown in Fig. 21.8. In the absence of a pulse input, its output will be at 0 V for any setting of the discriminator control. Suppose the discriminator control is set to −1·0 V actual, which would correspond to 3·0 V on its calibrated dial. An incoming pulse of 3·0 V or greater will take the trigger input to +2·0 V or greater, and the trigger will fire, resetting as the pulse falls back to zero. A pulse less than 3·0 V will produce no output. It will be seen that for any other setting of the discriminator control a similar sorting process will occur. The pulses produced by the discriminator are all of constant height (usually the standard +5 V logic assertion level); they are then passed to a counter or ratemeter for registration.

By the use of two discriminators, an inverter and a logic gate, it is possible to count only those pulses between an upper and a lower threshold; the difference between these levels is called the *gate width*, or *window*. Such a device is known as a pulse height analyser, and is used to further improve the isotope-to-background ratio, to count one isotope in the presence of another, or with a narrow window to generate the energy spectrum curve for an isotope. The principle of the analyser is shown in Fig. 21.9.

The lower discriminator passes all pulses above its threshold, and since the gate has its other input normally at +5 V these pulses pass the gate also, *except when the upper discriminator is also triggered.* If this occurs the gate is disabled. Accordingly only pulses between the upper and lower discriminator thresholds are passed on to be counted. It is usual to have two controls on a pulse height analyser; one sets the lower threshold, and the other the window.

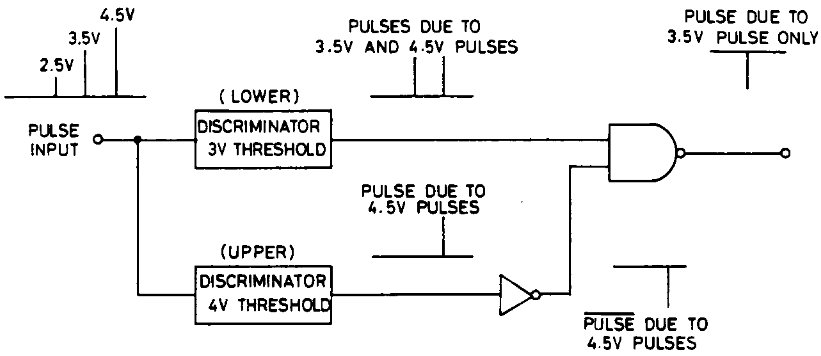


FIG. 21.9. Principle of pulse height analyser.

A modern counter usually incorporates two or more of these analyser channels, thus allowing two or more isotopes to be counted at once, provided that their energy peaks are sufficiently separated.

To derive the energy spectrum of an isotope, a narrow window width is selected, and this is slowly moved up the spectrum by rotating the threshold control, while the count rate is being observed. More elaborate (and expensive) methods of pulse sorting are available, based on computer techniques.

21.9 Statistics of counting

Radioactive decay is in statistical terms a rare and random event, with a *Poisson distribution.* This implies that if the same sample is counted a number of times under identical conditions, the number of counts obtained may be expected to vary, and the *standard deviation of the number of counts obtained in a series of experiments will closely approximate the square root of the mean number of counts.* That is, the actual number of counts obtained in each trial of a long series should depart from the mean number of counts M by less than $\sqrt{M}$ in 68% of trials, by less than $2\sqrt{M}$ in 95% of trials, and by less than $3\sqrt{M}$ in 99·7% of trials. Suppose that in ten identical trials the number of counts recorded were

1445	1468
1536	1280
1557	1541
1446	1452
1517	1520

Are these variations to be expected, or is the equipment behaving erratically? Let us check. The mean M is 1486, and $\sqrt{M}$ is 39.

1445 − 1486 =	41
1536	50
1557	71
1446	40
1517	31

1468	18
1380	106
1541	55
1452	34
1520	34

In a sample of ten cases this is about what would be expected; the equipment is said to be "counting statistically".

FURTHER READING

WANG, C. H. and WILLIS, D. L., *Radiotracer Methodology in Biological Science*, Prentice-Hall, New Jersey, 1965.

MELOAN, C. E. and KISER, R. W., *Instrumental Analysis*, Merrill, Columbus, Ohio, 1963.

QUIMBY, E. H. and FEITELBERG, S., *Radioactive Isotopes in Medicine and Biology*, Vol. I, *Basic Physics*.

SILVER, B., *Radioactive Isotopes in Medicine and Biology*, Vol. II, *Medicine*, Lea & Febiger, Philadelphia, (2nd ed.) 1963.

PRACTICAL

21.1 Construct the Geiger counter illustrated in Figs. 21.6, 21.10 and 21.11. Check that the e.h.t. generator starts when the battery is turned on (listen for the high-pitched whistle from the transformer core, and measure the e.h.t.). Test the counter on the luminous dial of a watch, and on a source of gamma radiation. Test the effectiveness of various materials as radiation shields.

21.2 Examine any types of radiation counting equipment that may be available, noting the features described in this chapter.

21.3 Carry out a series of counts of background, and of a radioactive source, using any simple counting equipment available. Check that the equipment is counting statistically.

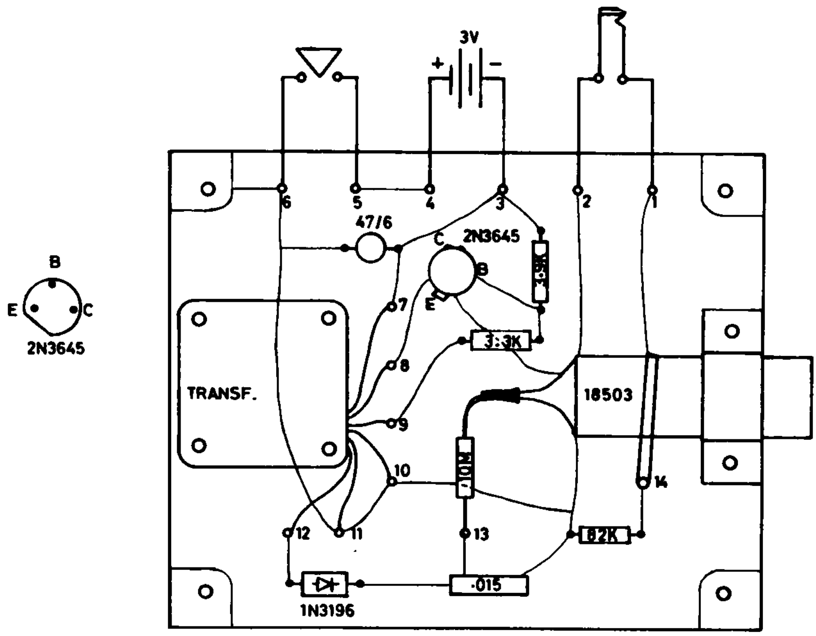


FIG. 21.10. Layout and circuit board of Geiger counter.

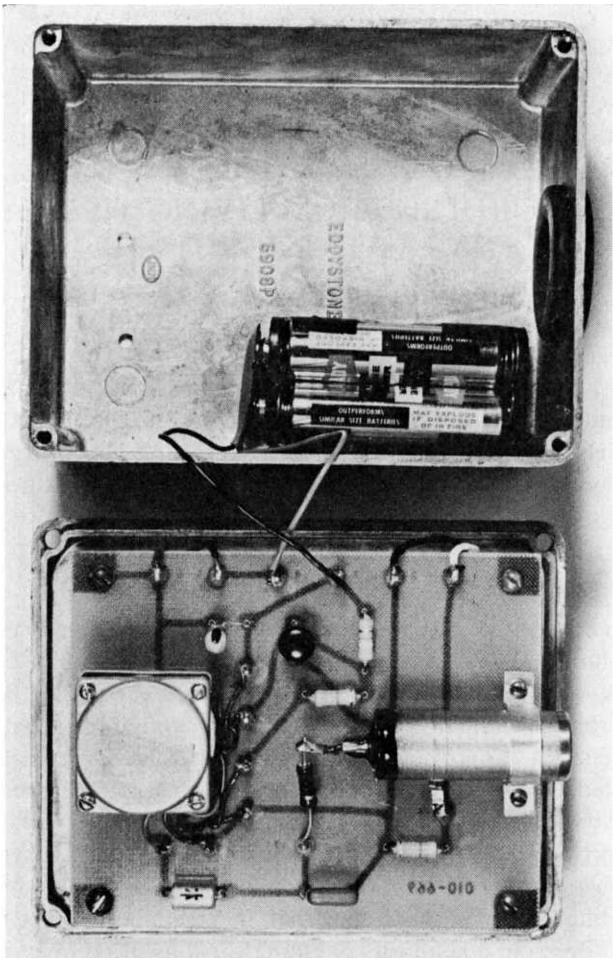


FIG. 21.11. Complete Geiger counter.

CHAPTER 22

RECORDING FROM LIVING TISSUE: STIMULATION OF TISSUE

22.1 Recording from tissue

Two problems arise in the recording of electrical activity from tissue. One is the provision of a suitable electrode system for connecting the equipment to the tissue; the other is in selectively recording the electrical activity of a particular organ in the presence of other active structures, and in the presence of voltages extraneous to the tissue.

22.2 Electrode systems

The conduction of an electric current through living tissue is purely electrolytic, as it is through a salt solution. The moving charges are not electrons, but consist of positive and negative ions, which are atoms or groups of atoms carrying either a surplus or a deficiency of electrons as compared with the number required to render them electrically neutral. (Thus a current cannot flow through an electrolytic system unless material is actually transported.) Each ion has a characteristic charge; the predominant ions in tissue are sodium (with a single positive charge) and chloride (with a single negative charge). To connect a metallic conducting system to an electrolytic system, special care must be taken over the selection of materials at the junction points, or undesirable side-effects will be produced. It is preferable that the metallic electrodes contacting the electrolytic system do so in a solution contaning the ions of the metal concerned. Such an electrode system is said to be *reversible*, since a current can pass from either electrode into the solution and out through the other electrode without producing a net voltage drop at the interfaces, as shown in Fig. 22.1. Notice that the two curves have gradients which represent the resistances of the solutions themselves, but in the reversible system a current will flow no matter how small is the voltage supplied; in the irreversible system a considerable voltage must be applied before any current at all flows. This *overvoltage* is typically 1 or 2 V.

The ions of all heavy metals, such as zinc or copper, are extremely toxic to living tissues, and cannot be brought in direct contact with them; on the other hand, the metals corresponding to the ions occurring naturally in tissue, such as sodium, react violently when brought in contact with water. To produce a suitable electrode system for stimulating or recording from tissue, there are three common expedients; they are (i) to use zinc and a dilute zinc chloride-containing paste, but only in contact with the skin (ii) to use an irreversible system, such as stainless steel in contact with tissue fluids or electrode paste, and to overcome the resulting overvoltage effects by suitable design of the electrical

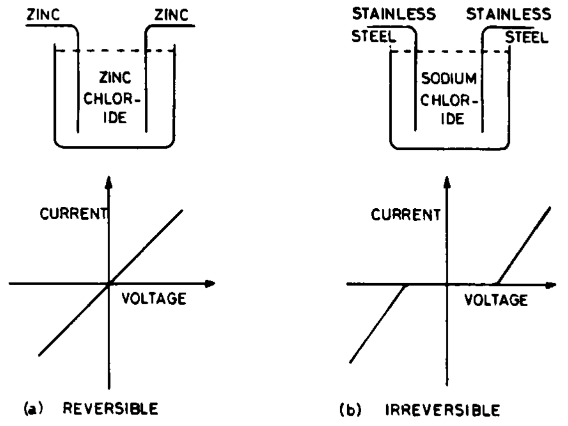


FIG. 22.1. Reversible and irreversible electrode systems.

circuitry (iii) to use composite silver–silver chloride electrodes, which are reversible to chloride ions, and which are non-toxic in contact with exposed tissue.

The silver–silver chloride electrode consists of a silver wire or disc of high purity (spectroscopic grade, at least 99·99% pure) coated with a film of solid silver chloride, which is almost, but not quite, insoluble in tissue fluid. When placed in contact with a chloride-containing solution, there is a very small but finite amount of silver ion in solution in the pores of the silver chloride, and hence the silver electrode is reversible. If the silver is made negative to the solution, as in Fig. 22.2(a), silver ions pass to the surface of the silver; each picks up an electron, and becomes a new silver atom on the surface of the electrode. At the same time chloride ions in the pores of the silver chloride pass out into the tissue; silver chloride from the coating dissolves, and a supply of silver and of chloride ions is maintained so long as any coating is left. As seen from outside the composite electrode, it appears that chloride ions are being generated by it.

If the electrode is made positive to the solution, as in Fig. 22.2(b), silver atoms from the surface of the metallic silver lose an electron each, and pass into solution as ions. Chloride ions arrive from the tissue, and precipitate out with the silver ions to form more solid silver chloride. As seen from outside the composite electrode it appears that chloride

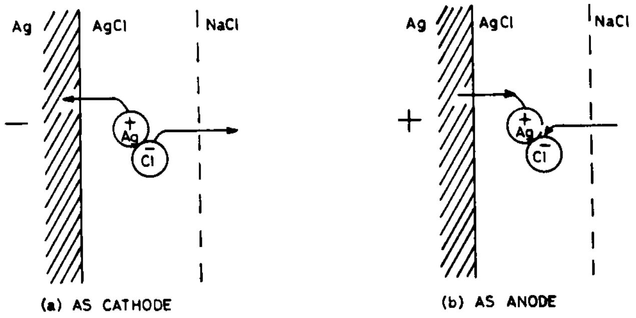


FIG. 22.2. Silver–silver chloride electrode.

ions are being absorbed by the electrode; the whole system appears to be *reversible to chloride.*

A suitable film of silver chloride may be built up on a clean silver surface by making it positive to a chloride-containing solution, and passing a current of about 1 mA/cm^2 of surface for several minutes. For some recording purposes, where only the most minute currents are allowed to flow, pure silver alone may be used; in contact with a chloride-containing solution a very thin film of silver chloride always forms.

For electrocardiographic recording, silver–silver chloride electrodes may be used in contact with the skin, or in conjunction with an electrode paste containing saline. A number of commercially made electrodes, complete with adhesive tape rings, are available, and are excellent for chronic application. For short-term recording it is more usual to employ zinc or stainless steel electrodes with an electrode paste. Stainless steel being irreversible is better avoided, since it gives rise to large drifts in the baseline of the ECG recording.

For electroencephalography, silver–silver chloride electrodes are common; they are usually held in place in the desired positions with an elastic head harness. Lead or solder electrodes, although once in general use, are now seldom employed. They are quite satisfactory electrically, but require to be cemented in place before use.

For direct recording from exposed tissues, silver–silver chloride electrodes can be used directly, although any movement of the underlying tissues may give rise to spurious recordings. They may also be used with *wick electrodes*, as shown in Fig. 22.3, to eliminate this effect.

For electromyography, and for implanted brain electrodes, stainless steel or tungsten is used. These materials are necessary for mechanical reasons, but are of course irreversible, and it is necessary to use a high-pass filter in conjunction with them to remove spurious potentials and movement artefacts.

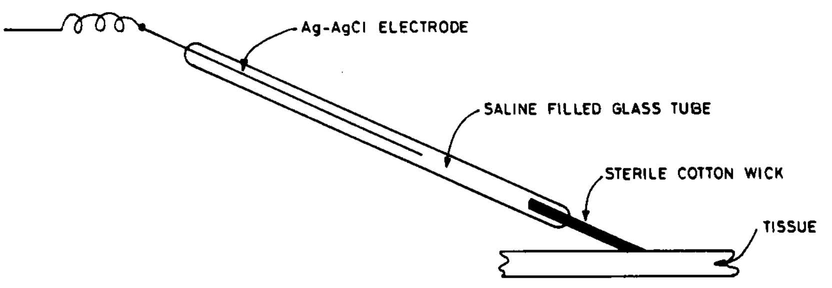


FIG. 22.3. Wick electrodes.

22.3 RECORDING FROM A VOLUME OF TISSUE

Whenever a living cell is activated to carry out its special function in an organism, it sets up a flow of current in the surrounding volume of tissue; such a current flow can be quite considerable where a mass of cells, such as those in the heart, acts in synchronism. This current flow in the surrounding medium may be detected by the use of suitable electrodes, and used diagnostically.

An active volume of tissue may at any instant be represented as an array of *dipoles*, each consisting of a paired source of current and sink of current a small distance apart. Current will leave the source in all directions, and will pass through the conducting fluid surrounding the dipole, to eventually enter the sink. In two-dimensional cross-section, the paths of current flow may be represented by lines, as in Fig. 22.4.

The maximum current flow will be in the region between source and sink, but there will be some current in all the surrounding volume.

These currents will give rise to voltage drops in the resistance of the conducting medium; tissue fluid has a resistance of 50 ohm cm (50 ohms between opposite faces of a 1 cm cube) and muscle tissue about 150 ohm cm. If the current sink is taken as the reference point, the potential of any point in the medium may be measured. By joining points of equal potential, a diagram such as Fig. 22.5 may be produced.

It will be noticed that the potentials fall off very rapidly in the vicinity of the source and sink, and quite slowly elsewhere; a pair of measuring electrodes inserted at points *A* and *B*, for example, would record only about 50 mV between them. It can be shown that the potential difference recorded by a pair of electrodes will fall off approximately inversely as the square of their distance from the dipole.

If the source and sink of current are in fact a mass of living cells, the strength and position of the equivalent dipole will vary from instant to instant, and so of course will

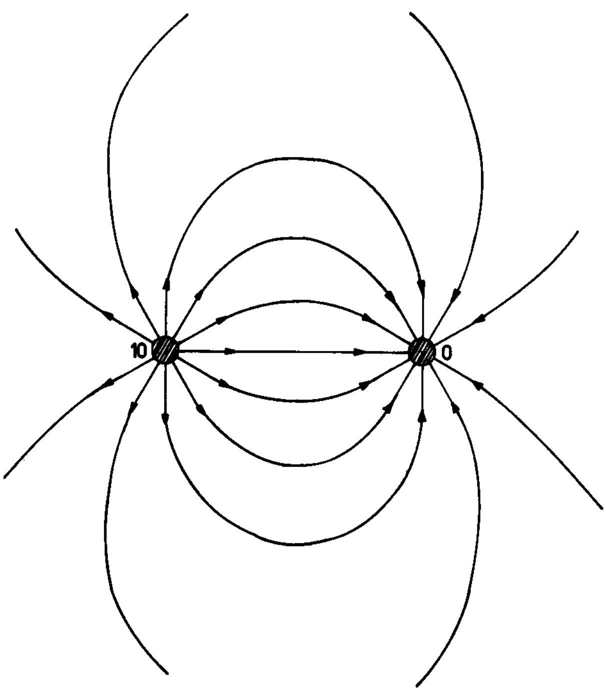


FIG. 22.4. Current flow surrounding a dipole.

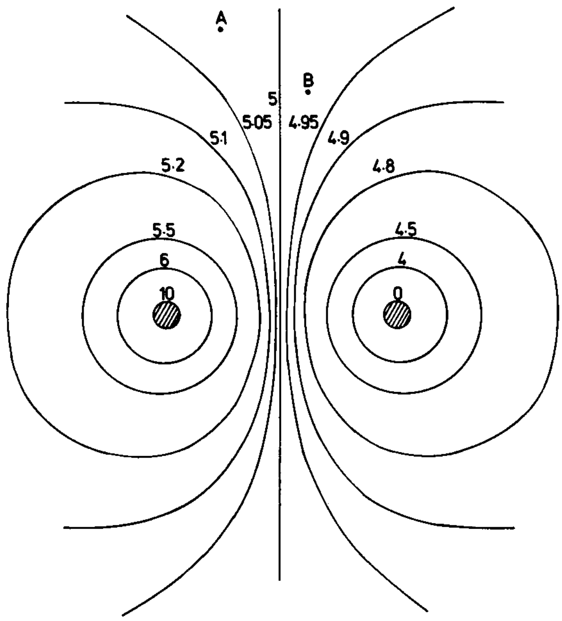


FIG. 22.5. Equipotential lines surrounding a dipole.

the potential recorded between a pair of electrodes at a fixed location in the surrounding volume. The use of a pair of electrodes in this fashion is spoken of as *bipolar* recording.

A second method of recording is *monopolar*; in this case one electrode is placed far out from the dipole, and the other in proximity to it. It will be clear from Fig. 22.5 that the remote (or *indifferent*) electrode must always have a potential very close to the mean of source and sink (here +5 V) no matter where it is placed, so long as it is far out. It then serves as a constant reference point, with respect to which the potential of the other electrode is measured. In general this gives a larger potential than bipolar recording, but this potential falls off approximately inversely with distance from the dipole. If other groups of cells are active, these are more likely to affect the record obtained if monopolar rather than bipolar recording is used.

22.4 Rejection of external interference

Consider a patient lying on a couch, as in Fig. 22.6, to allow an electrocardiogram to be taken. There will be a.c. supply wiring at various points in the floor, walls and ceiling, and in modern practice this is usually unshielded. For simplicity, consider an active 240 V r.m.s. lead in the ceiling alone. The patient is clearly between the plates of a capacitor (the wire and earth), to which an a.c. voltage is applied. If he were half-way between them, he would assume an alternating potential of 120 V r.m.s.! Of course, the capacitances from patient to wire and patient to earth are quite small, and the impedances at the supply frequency will be many megohms, so no danger is involved. However, the effect can cause intolerable interference with the recorded electrocardiogram unless

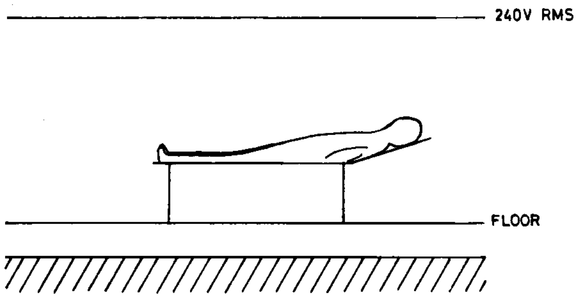


FIG. 22.6. Patient in an a.c. field.

suitable precautions are taken. The removal or shielding of all wiring is impracticable. The use of a properly screened room is desirable for precise laboratory work, but it is not suitable for clinical recording. A considerable reduction in the voltage picked up by the patient may be made by the use of a third earthed electrode elsewhere on his body, and by earthing adjacent metal objects such as a bed frame, but these measures alone are insufficient. Fortunately, it is possible to design amplifiers which will reject the supply frequency interference while responding to the desired biological signals. The general principle of rejection may be understood from Fig. 22.7, in which a pair of electrodes has

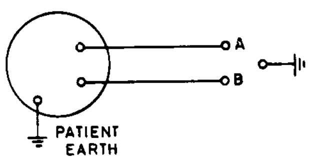


FIG. 22.7. Patient and recording electrodes.

been applied to an idealised spherical patient. The desired signal is that from *A* to *B*, but in addition both *A* and *B* have an undesired signal with respect to earth. These two signals are referred to respectively as the *out-of-phase* and the *in-phase* (or *common mode*) signal. The in-phase signal would not matter if the recording equipment had absolutely no connection to earth, but this is not possible. The equipment is usually supplied with a.c. power, and its frame is earthed for safety reasons; even if battery-operated, its frame would have considerable capacitance to earth.

If a suitable transformer were available, it could be connected as in Fig. 22.8. The out-of-phase signal would produce a primary current, which would induce a corresponding secondary voltage, while the common mode signal would not. Unfortunately it is not feasible to construct a transformer having the desired properties at the very low frequencies

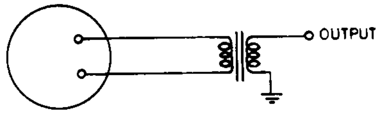


FIG. 22.8. Patient isolating transformer.

involved, although the scheme is sometimes used in conjunction with a device which first converts the signal frequencies to much higher values. Following the transformer, reconversion restores the original signal.

In modern practice biological preamplifiers for both clinical and research purposes are constructed from operational amplifiers. The principle of such a preamplifier is shown in Fig. 22.9.

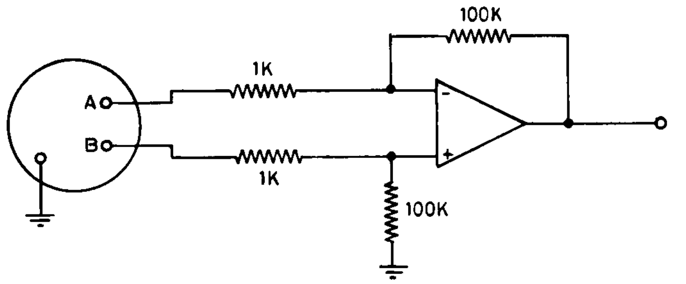


FIG. 22.9. Principle of biological preamplifier.

It will be seen that this constitutes a difference amplifier, as discussed in § 11.7; the out-of-phase signal between *A* and *B* will appear at the output amplified 100 times, while any in-phase signal will drive both *A* and *B* identically, and will not appear at the output.

Since patient electrode resistances are typically in the range 1 K to 10 K, and are almost certainly unequal, and since the electrode resistances appear in series with the 1 K input resistors, both gain and rejection of in-phase signals would suffer. This defect is overcome by the use of a unity-gain non-inverting amplifier in each patient lead; a practical amplifier is shown in Fig. 22.10. (Note that this amplifier, while perfectly adequate for animal experimentation, has no modern safeguards against accidental

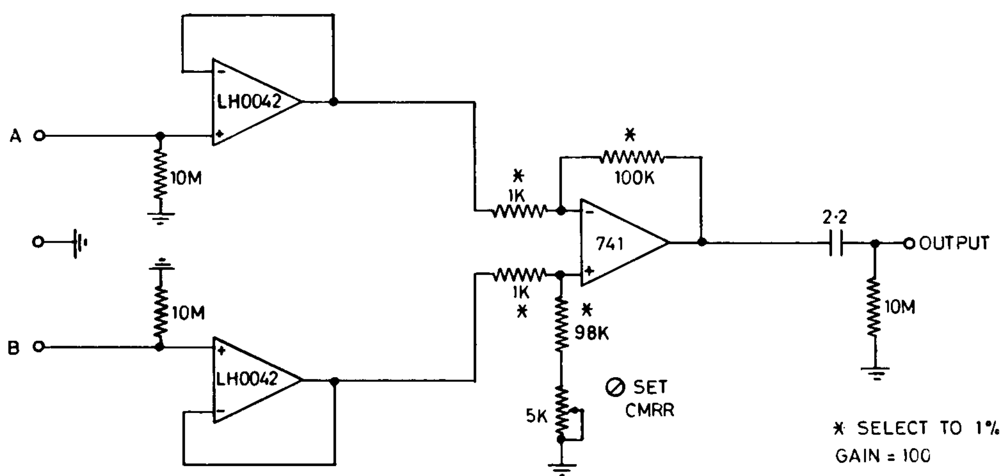
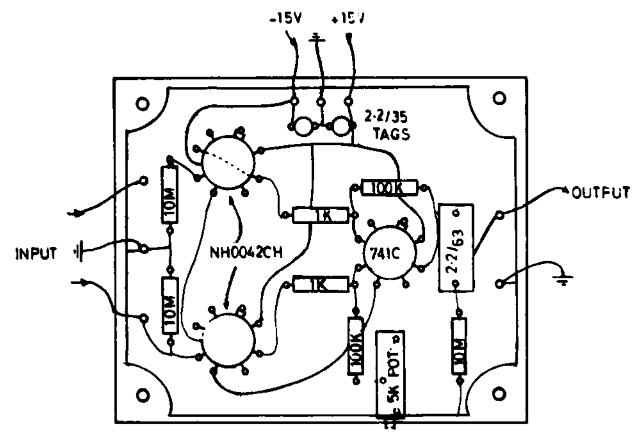


FIG. 22.10. Practical ECG amplifier.

electrocution, and should *not* be used on human patients. Safeguards are discussed in Chapter 23.) The printed card for this amplifier is shown in Fig. 22.11.

In a practical amplifier the two halves of the circuit are never quite identical, and the quality of an amplifier is measured by its *common mode rejection ratio* (CMRR). This is *the ratio of output for a given out-of-phase input to output for the same input applied in-phase*. The out-of-phase input is expressed as e volts, where e is measured between A and B. For example, if the amplifier of Fig. 22.10 gives 100 mV out for 1 mV input out-of-phase, and 100 μV out for 1 mV input in-phase (applied to A and B in parallel with respect to earth), it will have a CMRR of 1000, or 60 db.

A complete electrocardiograph or electroencephalograph amplifier should have a CMRR of at least 80 db, measured at the a.c. supply frequency. There is no advantage in a CMRR in excess of about 95 db, since there is always some *out-of-phase* voltage drop in the patient's tissues produced by stray a.c. electric fields. This will of course be treated in the same way as a desired signal, even by a perfectly balanced amplifier.



PRINTED BOARD 066-872

(N.B. DRAWN FROM COMPONENT SIDE)

FIG. 22.11. Printed card for ECG amplifier.

22.5 Tissue stimulation

Electrical stimulation of tissue, particularly of nerve or muscle, is a valuable test of its ability to perform its normal function, and is commonly used for diagnostic purposes.

Excitation of a living cell is brought about if an electric current of sufficient magnitude and duration is passed outward through the cell membrane; this may be achieved experimentally by the insertion of a very small insulated electrode inside the cell, as in Fig. 22.12(a), or by the passage of an electric current through the conducting medium in which the cell is situated, as in Fig. 22.12(b).

The use of *microelectrodes*, as in Fig. 22.12(a), is beyond the scope of this book, but details may be found in many standard texts (e.g. Kay, listed at the end of this chapter).

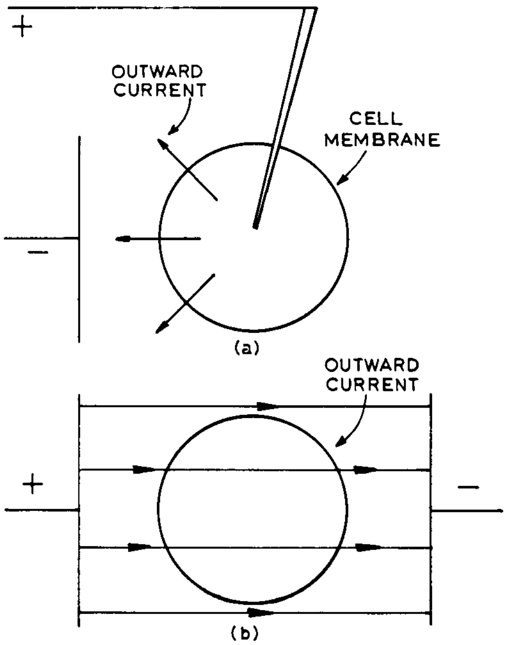


FIG. 22.12. Methods of stimulating a living cell.

The use of external stimulating electrodes is very common. The principles of selection of electrode materials already discussed also apply to stimulating electrodes.

To activate a living cell, a short pulse of current is applied, and it is found that for any given pulse duration there is a critical *threshold* amplitude, below which no activation occurs, and above which full activation occurs. As the pulse duration is increased, the threshold amplitude becomes smaller, but there is a certain minimum amplitude that any pulse, no matter how long, must reach to achieve activation. This relationship between current strength and duration is shown in Fig. 22.13. Pulse durations between 50 μsec and 10 msec are commonly used; the actual value of current required will depend on the volume of tissue to be stimulated and the arrangement of the electrodes on it. For electrodes 50 mm in diameter on the human body, 1 mA produces a barely perceptible stimulation of the sensory nerve endings.

It is frequently desired to stimulate a particular nerve tract or muscle through the skin; in this case much of the total current will flow through regions other than the desired one, and no significance can be attached to the actual electrode current flowing when threshold is reached. The *proportion* of the total current reaching the target organ will however remain constant, and it is readily possible to obtain a strength-duration curve. Such curves are often used to study the prognosis and course of recovery following accidental destruction of the nerve supply to a muscle. A second common technique is to measure the velocity of conduction of a nerve tract by stimulation at various points along its course; a region of damage can be identified in this fashion.

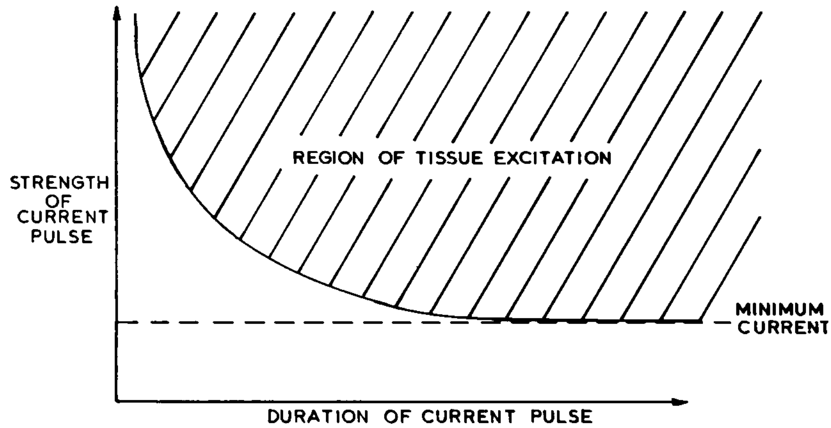


FIG. 22.13. Strength–duration curve.

22.6 STIMULATING PULSE WAVEFORMS

In modern practice a rectangular pulse of current or voltage is applied to the stimulating electrodes; this is simply generated, and occasions a minimum of discomfort to the patient. Older stimulators may employ an exponentially decaying pulse produced by discharging a capacitor of suitable size through the patient, or a pulse obtained from the secondary of a transformer when a current flowing through its primary is suddenly shut off. The obsolete terminology of "galvanic" for stimulation by a d.c. pulse, and "faradic" for stimulation through a transformer is still occasionally used. The output of a stimulator is often calibrated in volts; in this case the current which flows depends on the tissue resistance, the electrode resistance, and the internal resistance of the stimulator itself. The latter should not exceed 100 ohms.

Electrode paste is used to reduce the electrode resistance. It should be noted that surgical lubricant jelly is *not* suitable; it is a reasonably good insulator.

FURTHER READING

STRONG, P. R., *Biophysical Measurements*, Tektronix, Beaverton, Oregon, 1970.
KATZ, B., *Nerve, Muscle and Synapse*, McGraw-Hill, New York, 1966.
KAY, L., *Experimental Biology*, Chapman & Hall, London, 1964.
GEDDES, L. A., *Electrodes and the Measurement of Bioelectric Events*, Wiley, New York, 1972.

PRACTICAL

22.1 Using spectroscopic grade silver wire of about 1 mm diameter (SWG 18), prepare two silver–silver chloride electrodes as described in § 22.2. Immerse them in a physiological saline solution (sodium chloride 150 mmal/l., 0·9 gm/100 ml of water), and apply a voltage between them by means of a 3 V battery and 500 ohm potentiometer. Plot a graph of current flowing against applied voltage, as in Fig. 22.1.

Repeat for stainless steel electrodes of about the same diameter, placed the same distance apart in the saline solution.

22.2 If a standard recording electrocardiograph is available, record a human electrocardiogram from lead 1 (right arm to left arm), using (a) reversible (b) irreversible electrodes. Note any difference in baseline drift.

For each type of electrode, momentarily connect a 1·5 V battery between the electrodes during recording, and note the time taken for recovery of the trace to give normal recording.

22.3 Using an oscilloscope with a high resistance input probe (100 M or greater), measure the peak to peak voltage at the frequency of the a.c. supply lines appearing between a human subject and earth. Repeat this measurement in different parts of the room, and also with an earthed electrode attached to the subject elsewhere on his body.

22.4 Construct the biological preamplifier shown in Figs. 22.10 and 22.11, and test it by using it to record an ECG tracing on your oscilloscope. Set its CMRR to a maximum by connecting its input terminals together, and applying a 1 V r.m.s. alternating voltage at supply line frequency between them and earth; observe the output on your oscilloscope, and adjust the "set CMRR" control to give a minimum. By use of (a) a low frequency oscillator, and (b) a rectangular voltage pulse, measure its upper and lower cut-off frequencies.

CHAPTER 23

ELECTRICAL SAFETY

23.1 Need for precautions

Accidental electrocution in industry or in the home is not an uncommon occurrence, despite the exercise of great care in the design of the equipment involved. In the biomedical area, where electrodes and other conductors are deliberately applied to patients, even greater care in design is required, and the manufacturer, the maintenance engineer and the physician all carry a heavy responsibility in ensuring safe operation at all times.

Most deaths from electrocution are the result of ventricular fibrillation of the heart. In this condition the heart musculature ceases to contract sequentially to produce ejection of blood; small areas of muscle set up local "circus movements" of contraction. The condition can be corrected, if it has not persisted to the point where the heart's own blood supply has been impaired for too long, by a massive electrical stimulus which causes all the musculature to contract simultaneously for a brief period. Such a stimulus may be provided by a cardiac defibrillator; these devices will be discussed later in this chapter. Until a defibrillator can be made available, external cardiac massage may be used to maintain a minimal circulation to the heart and central nervous system; this technique must be used with care, and preferably by a qualified operator. Under some circumstances respiratory failure may occur; mouth-to-mouth resuscitation is then indicated in addition to any other treatment.

Since ventricular fibrillation is a common terminal condition, particularly in patients with cardiac disorders, it is very difficult to determine whether in any particular case death could in fact have been caused by electrocution. Statistics on the matter are obviously difficult to obtain. On the one hand, nearly all occurrences of ventricular fibrillation can reasonably be attributed to some prior condition of the patient; on the other, a good deal of electrical equipment in common use in medical practice is demonstrably capable of delivering shocks within the known lethal range, should certain combinations of circumstances arise. There is *no excuse for permitting this latter situation to continue.*

23.2 Macro- and micro-shock; safe current limits

A useful distinction can be drawn between electrocution by macroshock and electrocution by microshock.

Macroshock occurs when a current is passed through the intact body by contact with a source of e.m.f. It is the usual hazard which exists in the home and in industry, as well as in medical institutions, and it is a hazard for patients and hospital staff alike. Its causes are well understood and documented, and legislation in every country exists to enforce codes of safe practice. It will be apparent that *two* parts of the body must simultaneously be in contact with a source of e.m.f. to permit a current to flow, and the distribution of current flow in the body will depend on the location of these two parts. The fraction of this current passing through the heart will determine the probability of fibrillation. A general guide to the effect of *total* current flow through the human body is given in Table 23.1.

TABLE 23.1. EFFECTS OF MACROSHOCK

Total current flow	Effect
0·5 mA	Perceptible shock; may cause a related accident due to dropping or spilling hot liquids, etc.
1 mA	Distinct shock, accompanied by mild muscular contraction. *Recommended as upper limit of safe range.*
5 mA	Severe shock and muscular contraction.
10–1C0 mA	Severe shock and muscular contraction, with high probability of ventricular fibrillation and respiratory paralysis. Some burning of tissue.
Above 100 mA	Violent muscular contraction, complete cardiac arrest, respiratory paralysis and severe burning of tissue.

It should be noted that effects produced by currents of 10 mA or above will depend on the regions of contact and on the duration of the shock. In general, fibrillation is more likely to occur with currents in the 10–100 mA range than with any other; this range is the commonest in cases of contact with domestic supply voltages.

Nearly all the resistance to current flow offered by a human body lies in the immediate region of contact, and in the skin itself; if this resistance is reduced for any reason a correspondingly smaller voltage can produce a hazardous current flow. This situation occurs whenever measuring or current-passing electrodes are deliberately attached to a patient. For example, a typical electrode-to-electrode resistance for ECG electrodes is 5 K; in this case 50 V would cause 10 mA to flow.

It should also be noted that if a patient's heart is already affected by a number of predisposing conditions, and in particular if its oxygen supply is inadequate, the threshold for fibrillation may be considerably lowered.

Microshock occurs when a current is caused to flow within the body, by reason of the fact that one or both of the points of contact between the body and a source of e.m.f. is a conductor deliberately placed within the tissues. The hazard is greatest for a conductor introduced into the chambers of the heart itself, but it also exists whenever electrodes are located in the vicinity of the heart. Since this invasion is most likely to occur in intensive and coronary care situations, particular care must be taken in the design and use of the equipment to be employed.

A number of research workers have undertaken threshold measurements for fibrillation, by passing current from a catheter tip within the human heart to an electrode placed outside the body. (These tests can ethically be done when it becomes necessary to fibrillate a heart deliberately in the course of cardiac surgery.) The consensus of these tests gives a mean value of about 100 μA, and a very approximate standard deviation of

$\pm$ 30 μA. In view of these measurements, most authorities agree on *an upper safe limit of* 10 μA. Since it is possible to meet this requirement without great technical difficulty or expense, it is considered that all equipment used for invasive procedures should comply with it.

23.3 Electrical wiring in a building

Reticulation of electricity in any particular country tends to follow either American practice (based on a 110 V 60 Hz 2 phase distribution to the consumer) or British and European practice (based on a 240 V 50 Hz 3 phase distribution). In both systems the consumer is supplied from a large step-down transformer fed from the area high-voltage reticulation, but the winding connections differ.

The American system is indicated in Fig. 23.1. In this system the transformer secondary gives two 110 V r.m.s. supplies, 180° out of phase with each other. 110 V outlets in modern practice have three pins, connecting to *active*, *neutral,* and a separate *ground* (often a convenient water pipe). 110 V appliances operate between active and neutral, and the separate ground connects to the metal frame of the appliance. If due to a fault within the appliance active should short-circuit to the frame, the fuse will at once blow. Older installations have no separate ground pin; such 2-pin outlets are highly dangerous in medical institutions. A 220 V supply is available as shown, and is often used for permanently wired heavy current devices.

In British practice, considerably larger step-down transformers are usual; in an installation such as a hospital one or more of these will be located in a *sub-station*, and will supply the whole building. *Three-phase* reticulation is standard; the high-voltage supply to the sub-station is by means of one neutral and three active cables. The three active cables each carry an identical voltage with respect to the neutral, but the three voltages differ in phase by 120°. By means of a suitable three-phase transformer a

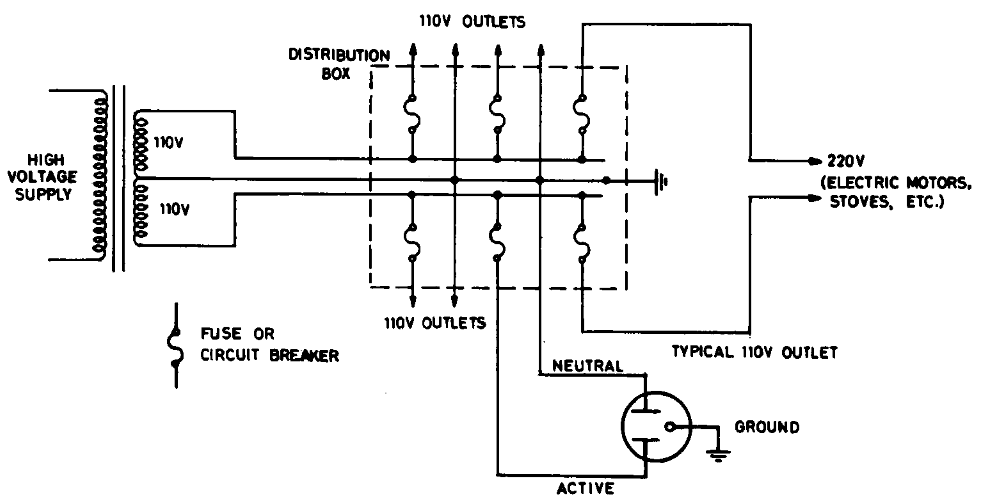


Fig. 23.1. American 110 V reticulation.

secondary three-phase supply is produced, in which each active is at 240 V r.m.s. with respect to the neutral. The neutral is earthed at the sub-station, and at each distribution box. The arrangement of a distribution box is indicated in Fig. 23.2. Only single-phase reticulation to 240 V outlets is shown in Fig. 23.2; it is also possible to utilise all three phases to operate heavy current devices, and this is often done. (A three-phase outlet has five pins, for the three phases, neutral, and earth.) As in the American system, single phase outlet sockets have three pins, providing a separate earth connection to the metal frame of the appliance in use.

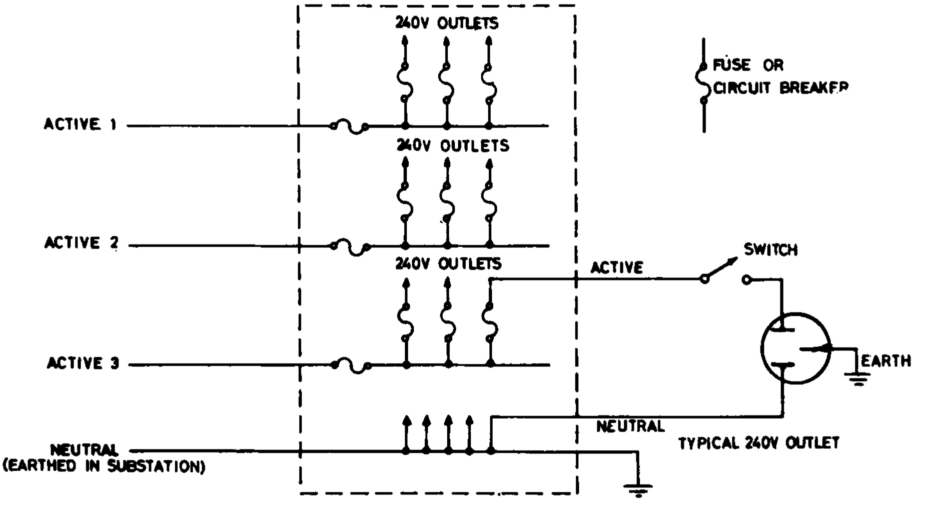


FIG. 23.2. British 240 V reticulation.

23.4 Possibility of electrocution

It will be seen from Figs. 23.1 and 23.2 that there are two ways in which the user can receive an electric shock; he may come into contact simultaneously with active and neutral, or with active and earth. Contact between active and neutral is relatively uncommon, and is likely only in the case where the user attempts modifications to an appliance while it is connected to the supply. Contact between active and earth will result in a shock, since neutral is connected to earth at the distribution box and substation. Since the frame of the appliance is deliberately connected to earth, it is relatively easy to come into contact with earth and active simultaneously if repairs to a "live" appliance are attempted; moreover, there is usually a variety of earthed metal objects in a room which the user may be contacting when the active is touched.

23.5 Macroshock hazards

Macroshock hazards in the hospital and laboratory can be minimised by rigid adherence to wiring regulations, by regular inspections and tests by a competent electrician, by training all personnel to recognise and report hazards, and by never purchasing equipment unless it is approved as safe by a recognised authority. Even with these precautions, hazards can arise. One of the commonest is the breakage of the earth wire in the flexible cord between the outlet and the appliance, usually at the plug itself. If now a second fault occurs, so that there is a connection between active and appliance frame, and the user or patient touches frame and another earth simultaneously, a severe shock will be received. A warning of the failure of an earth wire is the fact that a slight shock may be received on touching both frame and another earth simultaneously. This is due to capacitive coupling between active and frame, giving a small current which normally flows down the earth wire; shocks of this sort should at once be reported, and the equipment sent for repair. Difficulties in obtaining a hum-free ECG recording are frequently caused by a broken earth wire on some appliance in the vicinity.

23.6 Macroshock protection

No device can protect a person who places himself between active and neutral, since fuses and circuit breakers will invariably pass a lethal current without operating. Two systems of protection against active-to-earth electrocution are in vogue. Neither is a substitute for regular maintenance.

It is possible to give protection by the use of *isolating transformers*, as shown in Fig. 23.3.

Since neither side of the transformer secondary has any connection to earth, a user can touch both earth and either side without receiving a shock. The specifications of a typical transformer would call for a Faraday screen between primary and secondary, double thickness insulation, and a rating of 15 amperes.

In view of the bulk and expense of these transformers, their use is necessarily restricted to areas of relatively high macroshock hazard, such as operating theatres and renal dialysis units. (As will be discussed in § 23.8, they are also required in conjunction with equipotential earth systems in microshock-protected areas.) Installation in an old building is more difficult than in a new one.

In the simple system of Fig. 23.3, there is no warning if some appliance plugged into an outlet should develop a short circuit from one side of the wiring to frame, and thereby remove all the protection offered by the transformer. It is usual to supplement the system of Fig. 23.3 by the addition of an *earth leakage detector*, as shown in Fig. 23.4.

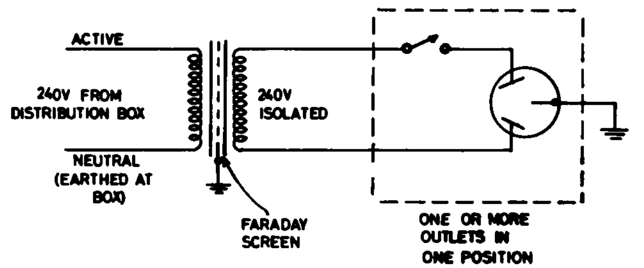


Fig. 23.3. Use of isolating transformer.

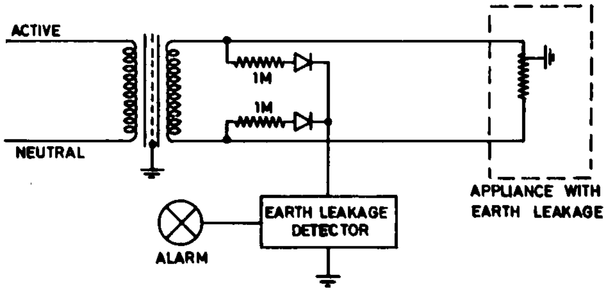


FIG. 23.4. Isolating transformer with earth leakage detector.

Should an earth leakage appear in an appliance plugged into the secondary, a current flows through the detector; this activates a warning signal. The detector can pass very little current (typically 10 to 100 μA), so does not itself constitute a breach of the isolation.

If installed, these detectors should be tested regularly during maintenance inspections. Any appliance that sets off an alarm should at once be reported.

The second protective system uses the *core balance relay*, as shown in Fig. 23.5. This device detects any *difference* in the currents between active and neutral; if a leakage appears between active and earth, the outward current does not return through the neutral. The device is then actuated, and interrupts the supply. An imbalance of 20 mA is sufficient to operate a good relay. This will of course give the user a severe shock if he is earthed and touches the active line, but the relay opens the circuit quite rapidly (in 100 msec or less) so the shock is not prolonged. These devices are reasonably reliable, and are less expensive to install than transformers. They have the major disadvantage that once interrupted the supply remains interrupted until reset. In an area such as an operating theatre this situation is quite unacceptable.

So far we have discussed patient and operator macroshock protection in terms of special installations, and we have seen that these are of limited applicability. A far better approach would be to use *macroshock-protected instruments* exclusively in the medical field. In the present state of the art, some instruments such as *electrocardiographs* and cardiac monitors can and should come into this category; other instruments such as *renal dialysis machines* employ large quantities of saline in the vicinity of electric wiring, and are more difficult to construct to an adequate degree of safety. At the present time, most *surgical diathermy units* require a large earthed electrode to be applied to the patient. It is clear that with an earth applied in this fashion, great care must be taken to avoid contact between the patient and any defective equipment. *An earthed person is*

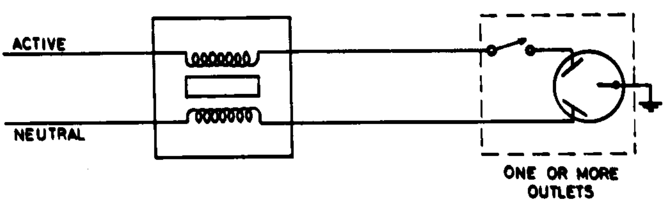


FIG. 23.5. Core balance relay.

always more at hazard than an unearthed person. When new surgical diathermy equipment is to be purchased, units with an *isolated output* are available and are recommended.

In the case of the *electrocardiograph, it is readily possible to avoid a direct earth connection to the patient.* This is best done by the techniques used to avoid microshock (see § 23.8 below), but a compromise design will give adequate macroshock protection (while still leaving a microshock hazard). The simple preamplifier of Fig. 22.10 is modified (a) by the inclusion of resistors in series with the patient leads, and (b) by the use of a protected *virtual earth* on the patient. The principle of the virtual earth is shown in Fig. 23.6. The patient is driven at the supply line frequency from a voltage divider formed by his leakage capacitances to earth and to the supply lines in the vicinity. At the beginning of a cycle, the virtual earth, and hence the inverting input of the operational amplifier, commence to rise in voltage. The output of the operational amplifier at once falls rapidly (remember it has a large gain) and current flows to the output through the 470 K resistor, thus minimising the voltage rise of the virtual earth. Should the virtual earth voltage commence to fall, the reverse will occur. This device is most effective in maintaining the patient at a potential close to earth without actually connecting an earth to him.

A complete macroshock-protected ECG preamplifier is shown in Fig. 23.7. The additional 470 K resistors in the patient leads *A* and *B*, and the 470 K resistor in the virtual earth input ensure that even if the patient is brought directly into contact with the active of the supply line, or if the frame of the ECG accidentally comes into contact with the active of the supply line while the patient is earthed by some external contact, the current flowing cannot reach a dangerous level.

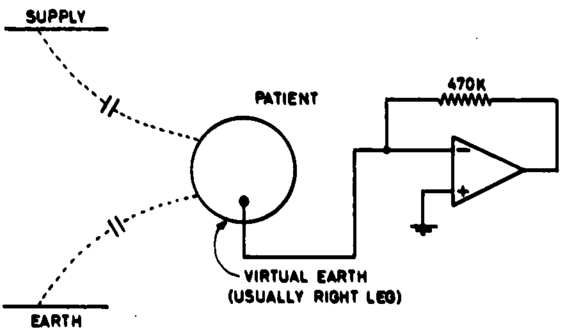


FIG. 23.6. Principle of virtual earth.

23.7 MICROSHOCK HAZARDS

By definition microshock hazards exist only where invasive procedures are practised; this includes cardiac catheterisation laboratories, coronary and intensive care wards, and operating theatres used for cardiac pacemaker implantation. (Unfortunately, in an emergency, invasive procedures such as the insertion of leads for a temporary external pacemaker may be undertaken anywhere.)

Microshock may occur due to either of two situations. (a) The patient is earthed externally, and the invading conductor acts as a source of e.m.f. (b) The patient contacts

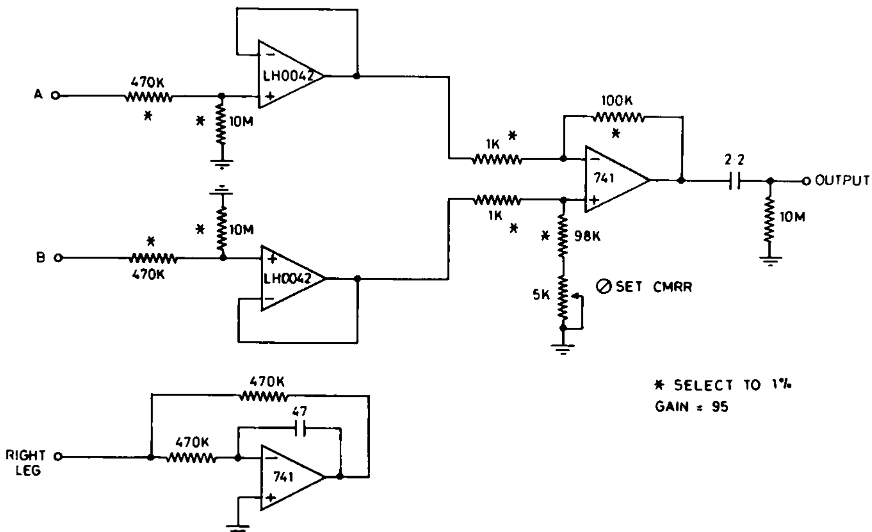


FIG. 23.7. Macroshock-protected ECG preamplifier.

a source of e.m.f., and the invading conductor is earthed. It must be realised that a very small source of e.m.f. may be lethal; 10 *mV will pass* 10 *μA through* 1 *K*. A typical example of (a) is the situation where the patient has a pair of pacemaker leads attached to his heart, with the exteriorised ends terminated in exposed plugs. The patient's arm is touching the metal frame of his bed, which is earthed. A nurse touches one of the exposed plugs while holding a bedside lamp whose metal frame is not earthed because its earth wire has broken at the plug end of its cord. In this case the leakage between active and earth in the cord and lamp may be quite sufficient to pass a lethal current through the patient's heart. A typical example of (b) is the situation where the patient has in his heart the tip of a saline-filled cardiac catheter, connected to a blood-pressure manometer whose frame is earthed. An electrocardiograph without microshock protection is also connected to the patient, and plugged into a wall outlet. A defective air-conditioning unit in an adjoining office is turned on, and passes 10 A through its earth line, which is common to the ECG outlet earth line (Fig. 23.8). A voltage drop of 10 V will appear across the patient, and assuming a total patient resistance of 10 K, a current of 1 mA will pass through his heart. Even if the instrument is macroshock-protected, the current passing *will still be* 100 *μA* if five leads each of 500 K are connected to the patient.

23.8 MICROSHOCK PROTECTION DEVICES

Microshock protection can be provided if a number of basic rules are followed.

(a) Areas in which invasive procedures are to be carried out must have an *equipotential earth system*. This requires special wiring, but is the first essential for safety. It implies that the earth pins of all sockets in the vicinity of the patient, and all metal objects in the vicinity of the patient, be connected by suitable earth wires to *one and only one common*

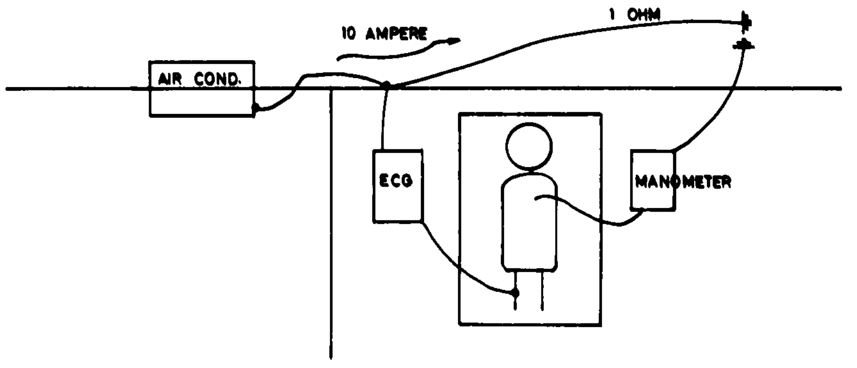


FIG. 23.8. Microshock by inadequate earth wiring.

physical earth point (Fig. 23.9). To avoid the necessity of using a very heavy copper earth bus, it is necessary to limit the fault currents due to possible defective appliances by the use of isolating transformers on all outlets in the room. Under these circumstances the earth bus and its connections may consist of insulated stranded copper wire (not less than seven strands of 1 mm diameter, or seven strands of 0·036 inch diameter). The bus should be a continuous loop, with branches to various outlets soldered on to it and the joints taped; the bus should not be "jumped" from each outlet to the next.

Beds may be connected to the bus by lengths of moderately heavy flexible insulated lead. These should be about 1 m in length, and provided with a spring clip at the bed end, and an attachment to go on a terminal on a wall plate at the other. The terminal attaches to the earth bus. Spare leads should be provided for each ward or theatre.

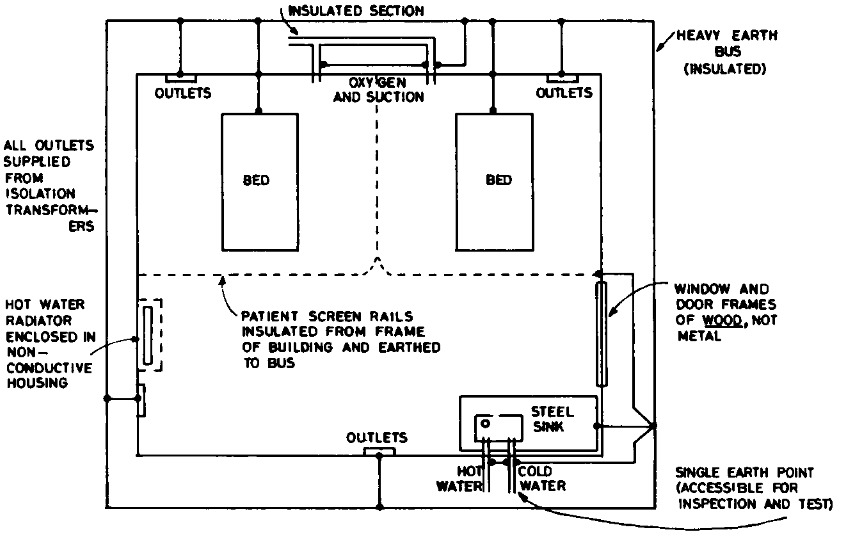


FIG. 23.9. Equipotential area.

Although it is preferable to treat a whole ward as a single equipotential area, this may be very difficult in the case of existing buildings. In this case each patient area may be treated separately, the common earth point in each case being the metal object most difficult to isolate from earth. (Under these conditions it is of course necessary to prohibit the operation of equipment in one area from outlets in adjacent areas.) The provision of insulating sections in oxygen, suction and waste lines may be necessary to obtain satisfactory isolation. It is best (i) to minimise or eliminate the provision of plumbing in equipotential areas (ii) to avoid the use of metal in doors, door frames, window frames, partitions, and so on. Conductive flooring should *not* be used, nor should terrazzo flooring with brass dividing strips.

(b) Exteriorised invasive connectors, such as pacemaker leads and catheters connected to manometers, should have their external connections insulated to avoid accidental contact with earth, and should be handled only with dry rubber gloves.

(c) Any electrical equipment brought into the area should be certified by a competent person as being free from microshock hazards, and should be subject to frequent inspection and testing for continuity of the earth wire in its flexible supply cord. Supply cords should be of adequate length (typically 5 m), and extension cords, multiple outlet expanders, and adapters of any type should be absolutely prohibited in these areas. Under no circumstances should appliances used in the area be plugged into outlets outside it. (This applies to floor polishers and cleaning appliances as well as to medical equipment.)

In the case of electrocardiographs, including patient monitor units, it must be possible (a) to connect the active of the supply directly to any combination of the patient leads without a current excess of 10 μA flowing (Fig. 23.10(a)) (b) to break the earth wire at the wall outlet without a current in excess of 10 μA flowing between any combination of the patient leads and earth (Fig. 23.10(b)) (c) to break the earth wire at the wall outlet and connect the active of the supply directly to the frame of the equipment (CAUTION!) without a current in excess of 10 μA flowing between any combination of the patient leads and earth (Fig. 23.10(c)).

These tests can be met only by an ECG built with a completely isolated preamplifier. This is usually done as indicated in Fig. 23.11. The preamplifier assembly is built to give a minimum of leakage capacitance to earth, and is supplied with power through a high

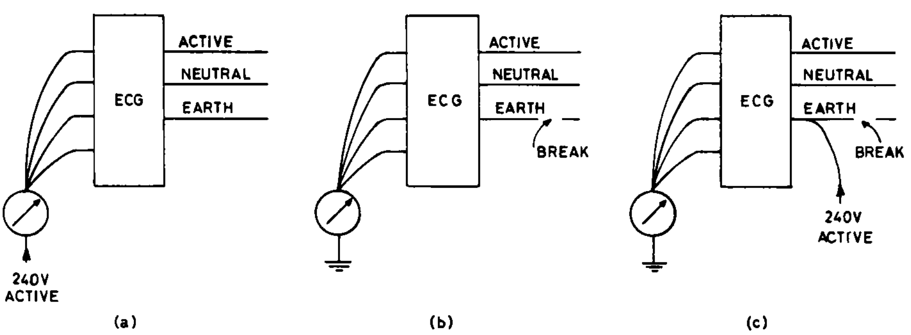


FIG. 23.10. Tests for microshock protection.

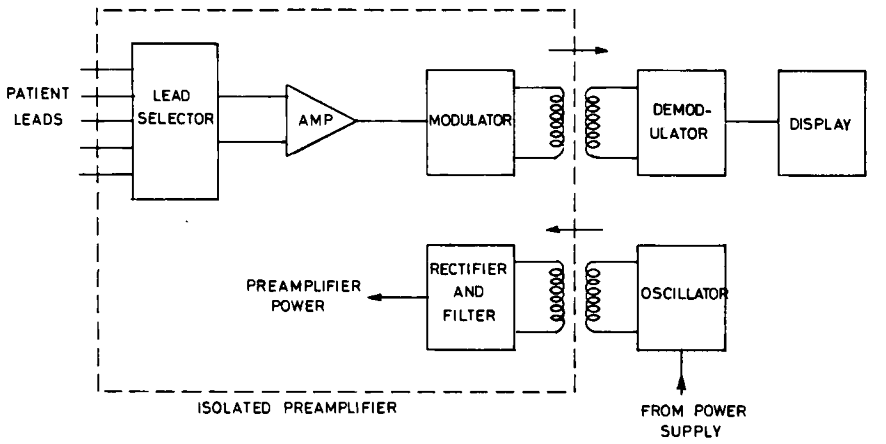


FIG. 23.11. Isolated preamplifier ECG.

frequency transformer whose primary and secondary windings are well insulated from each other, and whose leakage capacitance is very small. The signal output of the preamplifier is similarly coupled out through an isolating transformer. To give an adequate frequency response it is necessary to convert the ECG signal to some much higher frequency by means of a suitable modulator, couple it out through the transformer, and then reconvert it. (Alternative methods such as the use of optoelectronic or acoustic coupling are also used.)

Isolated preamplifier electrocardiographs are now readily available at a reasonable price, and there is no justification for not using them where invasive procedures are being undertaken.

23.9 SUMMARY OF ELECTROCARDIOGRAPH TYPES

TABLE 23.2

Type of ECG	Areas of use	Maximum leakage
Unprotected: direct earth on right leg lead, direct connection of signal leads to preamplifier	Unsuitable for human use	Excessive
Macroshock-protected: Resistance-protected virtual earth on right leg lead, resistance-protected signal leads to preamplifier	Suitable for human use except during invasive procedures (may be used in an emergency for invasive procedures in an equipotential area if a separate earth wire is connected externally from instrument frame to earth bus)	1 mA with earth lead in power cable broken, 10 μA with earth lead intact
Microshock-protected: Right leg and signal leads returned to isolated preamplifier	Suitable for human use under all circumstances	10 μA with earth lead in power cable either broken or intact

It should be noted that most microshock-protected electrocardiographs are more prone to interference from any prevailing alternating current leakages than are macro-shock-protected instruments. An almost certain cure is (a) to remove any flexible power leads from the immediate vicinity of the patient, unplugging those not actually in use (b) to ensure that the patient's bed is earthed to the equipotential system, and that the patient is not touching the metal of the bed.

23.10 Defibrillators

Reversion from ventricular fibrillation to a normal rhythm is carried out by the use of a massive surge of current through the heart. This may be administered directly to the surface of the ventricles if the heart is already exposed during surgery; more usually it is applied through the chest wall. For external defibrillation a pair of electrodes about 100 mm in diameter are used; these are fitted with well-insulated handles. The current pulse used has a duration of 3 to 4 msec, and a peak value of between 40 and 80 *amperes*. Since the resistance between these large electrodes is about 50 ohms, a peak voltage of 2000 to 4000 is used. Values in excess of 5000 V cause excessive cardiac tissue damage, and should be avoided. The pulse is usually formed by discharging a large capacitor (typically 20 μF) which has previously been charged to the desired value by a suitable power supply; a small inductor is often placed in series with the discharge circuit to shape the pulse to the desired form, as indicated in Fig. 23.12.

Since the pulse is produced by a capacitor discharge, it is customary to rate the pulse used in terms of joule of energy delivered to the patient. Common values are 75, 150 and 300 J, the latter being used most frequently on adults.

For internal defibrillation, electrodes of about 50 mm in diameter are used. The resistance between them when applied to the heart is about 50 ohms, and energies of 25, 50, and 100 J are most common.

Defibrillators must be simple in external panel layout, with a clear and unmistakable sequence of operation marked on them. They should be tested at regular intervals, and the tests should include power output and waveform measurements into a non-inductive 50 ohm dummy load of adequate dissipation.

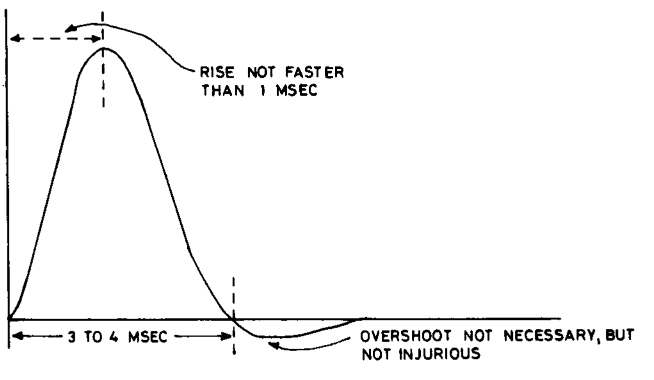


Fig. 23.12. Pulse for ventricular defibrillation.

Cardiac defibrillators are normally used in conjunction with an electrocardiograph, and it is important that the electrocardiograph used should resume normal registration of the ECG as soon as possible after defibrillation. Modern instruments are designed with this requirement in mind.

FURTHER READING

WALTER, C. W., *Electric Hazards in Hospitals*, National Academy of Sciences, Washington, 1970.
Hewlett-Packard, "Patient Safety", Application note AN718.
POCOCK, S. N., Earth-free patient monitoring, *Bio-medical Engineering* (1972) **7**, 47 and 67.

PRACTICAL

23.1 Examine and report on any type of electrocardiograph available in terms of (i) frequency response (ii) safety (iii) general comments.

23.2 Examine and report on any type of ventricular defibrillator available, in terms of (a) adequacy of panel layout and instructions (b) safety to patient and operator (c) delivered waveform and power. (Use a 50 ohm dummy load in series with a 1 ohm metering resistor. Observe the waveform generated by means of an oscilloscope across the 1 ohm resistor; sketch or photograph the waveform, calculate and graph the instantaneous power (e^2/R), and deduce the energy delivered by measuring the area under the instantan eous curve.)

(N.B. PROCEED WITH GREAT CAUTION!)

CHAPTER 24

REGULATED SYSTEMS

24.1 Introduction

A regulated system is one in which departure from a desired state results in an automatic response tending to restore that state. Examples have already been discussed, in the regulated power supply and the potentiometric recorder. Systems of this type are extremely important in the field of engineering, and equally so in biology; their basic property is that they actively resist the effect of changes in their environment. A knowledge of the general principles involved is applicable to both engineering and biology. Although a detailed mathematical study of these systems can be very involved, the main features may be brought out quite simply. The temperature regulation of a water bath will be taken as an example; it is also a useful study in its own right.

24.2 Behaviour of an unregulated water bath

We will first consider how an unregulated water bath behaves, and then apply a regulator to the system.

Fig. 24.1 shows a well-stirred water bath containing V ml of water, which can receive energy from an electric heater at a rate Q joule/sec, or Q watt. Let us suppose that initially the bath is at room temperature, and that the heater is turned on at a time $t = 0$. As the bath temperature commences to rise, heat is lost from the bath by radiation, conduction, convection and evaporation. Let this rate of loss for any temperature T degrees Celsius *above that of the room* be R watt. Initially the situation was that $T = 0$, $Q = 0$, $R = 0$, and the bath was in *thermal equilibrium* with its surroundings.

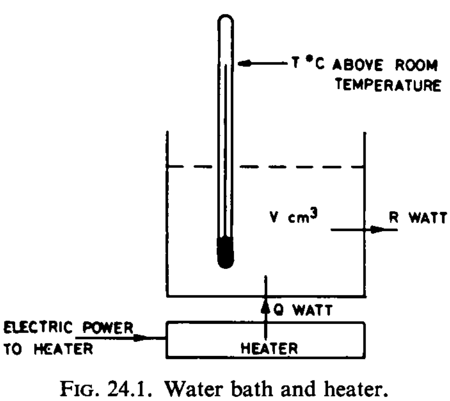


Fig. 24.1. Water bath and heater.

Experiment has shown that 4·18 J will raise the temperature of 1 ml of water by 1 deg C. Since the bath has a volume V ml, the application of Q J/sec will cause an initial temperature rise of $Q/4{\cdot}18V$ deg/sec. It so happens that the rate of heat loss R is directly proportional to the temperature T in excess of room temperature:

$$R = aT, \tag{24.1}$$

where a watt per degree Celsius is a constant for the particular water bath used. If at some time t after turning on the heater the bath is at T°C above ambient, the net rate of gain of heat will be $(Q - R)$ watt, and so the bath temperature must be rising at a rate $(Q - R)/4{\cdot}18\ V$ deg/sec. In calculus notation:

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{Q - R}{4{\cdot}18V} \tag{24.2}$$

but since $R = aT$,

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{Q - aT}{4{\cdot}18V}$$

$$= \frac{Q}{4{\cdot}18V} - \frac{a}{4{\cdot}18V}T. \tag{24.3}$$

As the temperature rises, so will R rise, and the net rate of gain of energy $(Q - R)$ will become less and less, and so will the rate of rise of temperature. Eventually the system will reach a *steady state*, at which $R = Q$, and $\mathrm{d}T/\mathrm{d}t = 0$, at some constant temperature T_C. (It can be seen that the initial equilibrium condition is a special case of a steady state, where $R = Q = 0$, and $T = 0$).

The differential equation (24.3) may be solved for T by a variety of standard methods; it is found that

$$T = \frac{Q}{a}(1 - \epsilon^{(-at/4.18V)}), \tag{24.4}$$

This is the equation for an exponentially rising curve of a form identical to that of a capacitor charging through a resistor, as will be seen from Fig. 24.2.

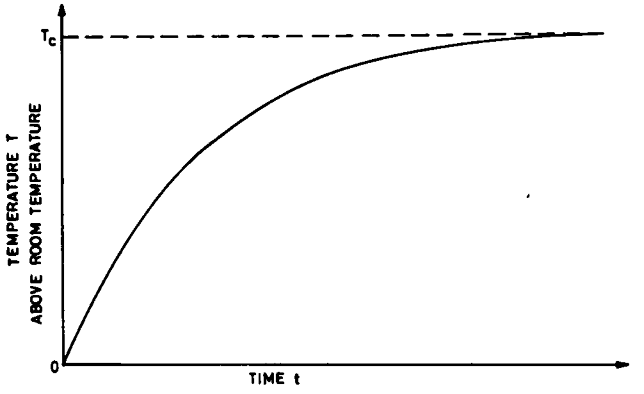


FIG. 24.2. Heating curve for a water bath.

From eqn. 24.4 it will be seen that after a very long time t,

$$\epsilon - \frac{at}{4{\cdot}18V} = 0,$$

and

$$T_c = \frac{Q}{a}. \tag{24.5}$$

Further, the curve has a time constant τ, given by

$$\tau = \frac{4{\cdot}18V}{a}. \tag{24.6}$$

The laws of heating and cooling of water baths were first set out by Newton.

A water bath is characterised by its bath constant a; the time constant is then fixed.

Knowing the bath constant in watt/degree, the heating power for any required temperature above ambient is easily determined. This constant is best determined by filling the bath with warm water, and observing the rate of cooling in degree per second for a given temperature T above ambient. Then, since $Q = 0$,

$$\frac{dT}{dt} = -\frac{aT}{4{\cdot}18V} \tag{24.7}$$

in which a is the only unknown quantity. By doing this operation with a new water bath before the heating element is selected, much subsequent waste of time may be avoided.

For a bath in which $Q = 100$ W, $V = 500$ ml, and $a = 5$ W/deg

$$T_c = \frac{Q}{a} = 20^\circ \text{ C above ambient}$$

$$\tau = \frac{4{\cdot}18V}{a} = 418 \text{ sec}$$

and the initial rate of rise $= Q/4{\cdot}18V = 0{\cdot}048$ deg/sec. The bath will take 4 time constants, about 25 minutes, to come to within 1 % of its final temperature.

24.3 Regulated water bath: on–off control

If a pair of contacts is sealed into the stem of the thermometer of Fig. 24.1, and these are connected into a circuit controlling the electric power to the heater, a regulated system results, as shown in Fig. 24.3.

If the bath is initially at room temperature, and the power is turned on at a time $t = 0$, the thermometer contacts are open, the relay contacts are closed, and the bath receives Q' watt of heating (Q' is made considerably larger than Q in the previous case).

Following the argument of the previous section, it will be seen that the bath constant a and time constant $4{\cdot}18V/a$ are unchanged, but that T_c $(=Q'/a)$ will be at temperature far above that required, and that the initial rate of rise $(=Q'/4{\cdot}18V)$ is much greater than

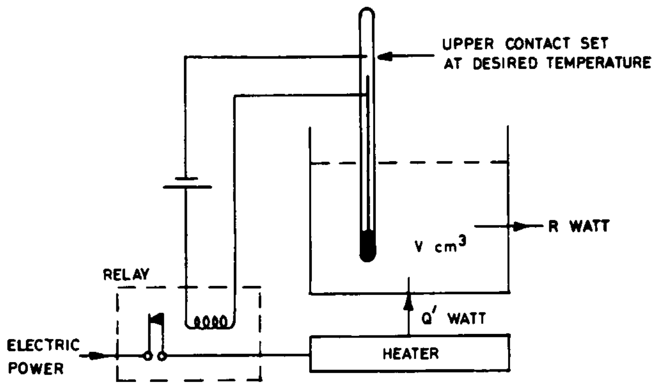


FIG. 24.3. Regulated water bath.

previously. Using the previous example of $V = 500$ ml, $a = 5$ W/deg, but taking Q' as 300 W,

$$T_c = 60°\text{C above ambient}$$

$$\frac{Q'}{4{\cdot}18V} = 0{\cdot}144 \text{ deg/sec}$$

The temperature–time curve will follow this steeper exponential rise until the thermometer contacts close. The power to the heater will then be shut off, and the curve will cease to rise. Owing to the thermal capacitance of the heater unit, the bath will rise a little above the temperature set by the upper thermometer contact. It will then start out on an exponential *cooling* curve towards room temperature, as shown in Fig. 24.4. The cooling curve, as before, will be governed by eqn. (24.7), and will be much slower than the heating curve.

As soon as the thermometer bulb temperature falls to the point where the thermometer contacts are broken, the heating power is restored. The heater, however, will take a short time to warm up before its output rises to Q', and consequently the bath temper-

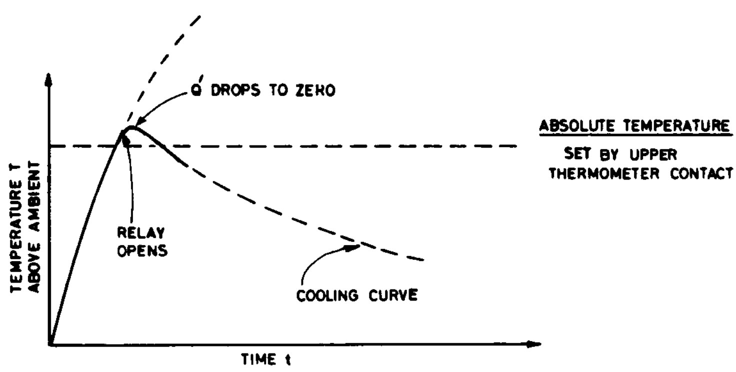


FIG. 24.4. Overshoot of bath temperature.

ature will fall further before again starting to climb. This cycle of rising and falling temperature will repeat indefinitely. The bath temperature will fluctuate up and down over a small range (the *differential* of the system), but this can be made as small as desired by suitable design (Fig. 24.5). Further, the mean temperature is now *absolute*, and independent of room temperature or fluctuations in the heater supply. To minimise the differential, the thermal capacitance of the heater must be minimised; the heating element is in practice usually immersed in the water bath. Further, the value of Q' selected must not greatly exceed that required for the desired bath temperature without regulation. This, however, makes initial heating from room temperature slow, and it is usual to provide a booster heater for the initial heating. This is cut out manually when the desired operating temperature is reached.

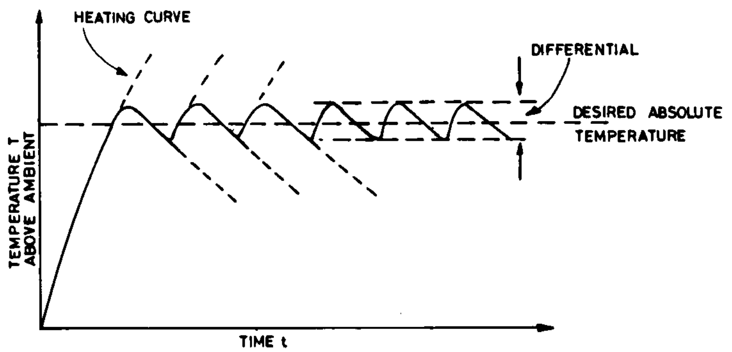


FIG. 24.5. Behaviour of regulated water bath.

24.4 PROPORTIONAL CONTROL

If the thermometer contacts and relay of the previous section were replaced by a heater regulator arranged to vary *Q in proportion to the error* between the actual and desired bath temperatures at any instant, a much smoother control of temperature could be achieved. In the case of a water bath this precision is not likely to be needed, but there are many cases where proportional control is desirable. The electronically regulated power supply is an example of proportional control; the potentiometric recorder is another (except that this system overloads by large errors, and is truly proportional only when close to its desired position). Many physiological examples of proportional control are available.

Systems which embody proportional control require careful design if their operation is to be stable. In mathematical terms, the equation for the system is a differential one of second or higher order, and the solution of such an equation may take one of three forms, depending on the relationship between the physical constants in the system. The system may be *overdamped*, *critically damped*, or *underdamped*. For a temperature regulator the behaviour after switching on will follow one of the patterns of Fig. 24.6, depending on the various thermal time constants in the system. The critically damped curve is the one to be desired, since it gives the fastest possible approach to the desired temperature without overshoot. Provision is usually made to adjust the damping of

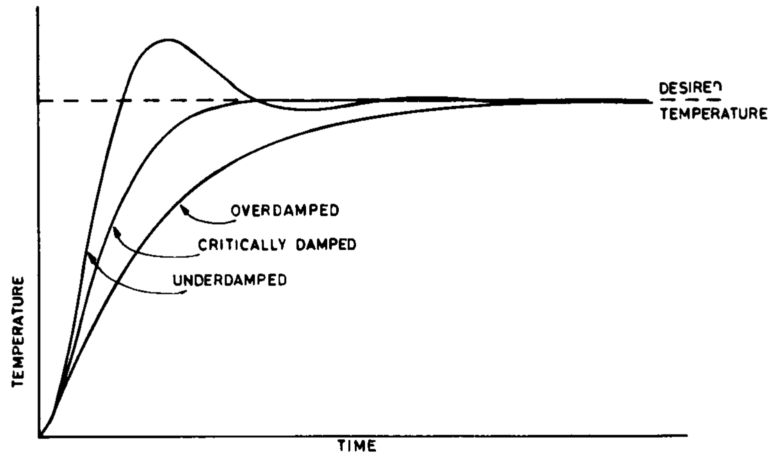


FIG. 24.6. Response of a water bath using proportional control.

proportional control systems; a typical method will be discussed in Chapter 27 for a potentiometric recorder.

24.5 NON-LINEAR SYSTEMS WITH PROPORTIONAL CONTROL

Many practical systems incorporating proportional control, both in the fields of engineering and of physiology, have elements in their control pathways which themselves are not constant, but which depend on the level of the process being controlled. If such a system is underdamped, it is common to find not the behaviour of the underdamped curve of Fig. 24.6, in which the oscillation dies out, but a continuous oscillation about the desired temperature, giving a differential similar to that obtained in on–off control. This may usually be eliminated if desired, but providing that the differential is not excessive, the system may be quite adequate for many purposes. This behaviour is typical of many physiological systems.

24.6 REGULATED SYSTEMS IN GENERAL

The types of regulated system described above may be represented by a block diagram, as in Fig. 24.7. A second type of regulated system, incorporating "feed-forward", is shown in Fig. 24.8. It will be seen that this arrangement is less economical of energy than the first, since Q is permanently at its maximum value, and T is controlled by wasting a

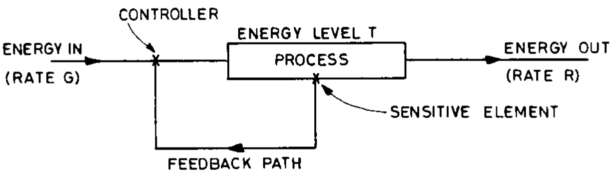


FIG. 24.7. Regulation by control of input.

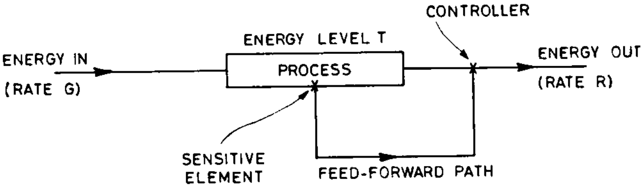


FIG. 24.8. Regulation by control of output.

greater or lesser amount of energy by way of R. It may have certain practical advantages, especially as it produces a slow rise and a rapid *fall* in T. A combination of both types is not uncommon, especially where T may be required to lie either above or below the ambient level in the room.

FURTHER READING

HAMMOND, *Feedback Theory*, English Universities Press, London, 1958.
ASHBY, *An Introduction to Cybernetics*, Methuen, London, 1956.
CLYNES and MILSUM, *Biomedical Engineering Systems*, McGraw-Hill, New York, 1970.

PRACTICAL

24.1 Determine the bath constant of a large beaker of water, using the method described in this chapter, and hence derive its time constant. Keep it well stirred. What electrical heating power would be required to maintain it at 37°C in a room at 25°C, and at 10°C? Light a low burner under it, and plot the initial rate of rise of temperature. What value of Q have you set in? What final steady-state temperature would the bath reach, and about how long would it take?

24.2 Sketch a block diagram of the system involved when you move your finger to touch a spot of ink on the surface of a table. How could this system become oscillatory?

24.3 Regulated systems often embody a human operator as part of the feedback loop, and some mechanical device as the remainder. Sketch the system involved in manipulating a hot shower for satisfactory results. How can this system become oscillatory, and how is it normally stabilised?

24.4 If available set up a water bath with electrical heater, thermostat and stirrer, and record its temperature by means of a resistance thermometer, pseudo-Wheatstone bridge and potentiometric recorder. (The thermometer and Wheatstone bridge can easily be improvised.) How good is the thermostat? What is the effect of excessive heating power?

24.5 Design features of any other available regulated system should be examined, and the accuracy of control observed.

CHAPTER 25

TRANSDUCERS

25.1 Transducers

Transducers are devices for the conversion of one variable into another; as generally used, the term is applied specifically to devices for converting mechanical variables such as pressure, displacement, force and so on into corresponding electrical changes. In this chapter a section is devoted to each type of mechanical variable, and the most suitable means of transducing it in modern practice.

Any transducer is used in association with suitable electronic circuitry to give a final electrical output of convenient amplitude; for data storage and processing, as discussed in Chapter 27, it is usual to produce an output in the range +1 V to −1 V.

25.2 Displacement transducers

25.2.1 *Potentiometer*

Potentiometers for use as transducers are available with almost frictionless bearings and wipers, in both rotary and linear forms. The *resolution* of such a transducer is limited by the distance between turns of the wire used; it is commonly about 0·1 % of the total mechanical travel. High quality transducer potentiometers are fairly expensive, but in many cases they can give the desired electrical output with a minimum of associated circuitry. A typical arrangement using a buffer operational amplifier to avoid loading errors is shown in Fig. 25.1.

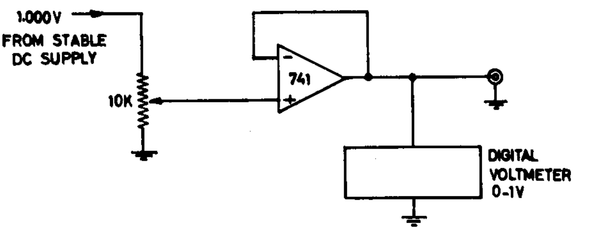


Fig. 25.1. Potentiometer transducer and buffer.

25.2.2 *Linear variable differential transformer*

The linear variable differential transformer (LVDT) consists of three windings, connected as shown in Fig. 25.2. Windings *A* and *C* are connected in opposition, so that no net output is induced in them by *B*. If now a small slug of iron is introduced into the

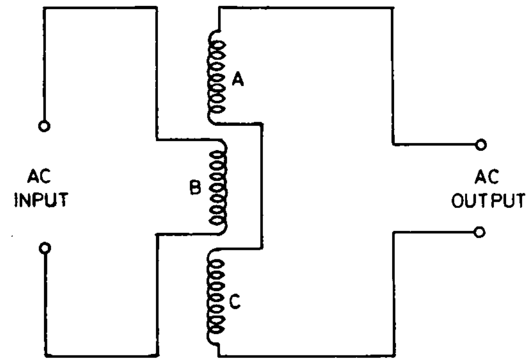


FIG. 25.2. Linear variable differential transformer.

centre of *B*, the balance will be preserved; but if this slug is displaced up or down, it unbalances the system, and an output appears. The phase of the output will be reversed as the core goes through the null position, so it is possible to detect on which side of balance the core lies by comparing the output and input phases. A very common arrangement is as shown in Fig. 25.3. The LVDT is supplied with a sine wave of constant amplitude from a suitable oscillator (at 400 Hz in the example). Constancy of amplitude is ensured by the thermistor in the feedback path, and the frequency is determined by the values of *R* and *C* used:

$$f = \frac{1}{2\pi RC} \tag{25.1}$$

As the position of the LVDT core varies, it is said to *modulate* the 400 Hz *carrier*, causing the output amplitude to vary. This output variation is then converted into a DC variation by a *demodulator* package, as shown; the demodulator also requires a *reference input* directly from the oscillator.

LVDTs are commercially available in many different forms, and if necessary may be constructed in an experimental workshop; the associated electronic circuitry is also available commercially.

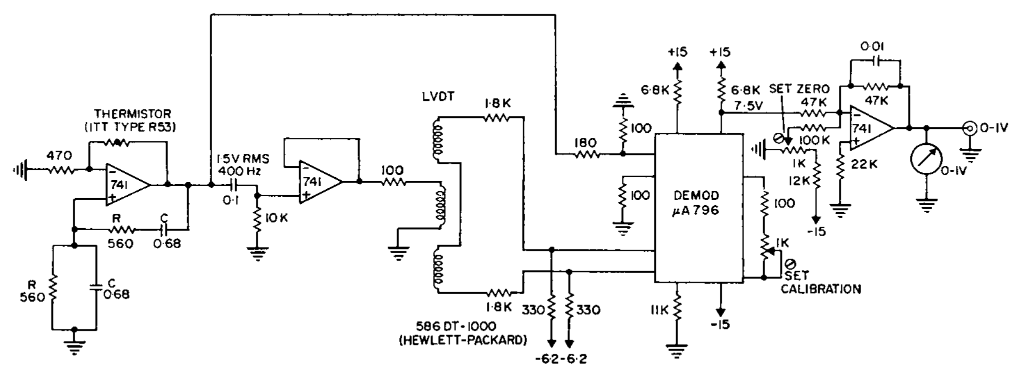


FIG. 25.3. LVDT and associated circuitry.

A particularly convenient unit is the *DC differential transformer* (DCDT). This consists of an LVDT with its associated circuitry in a single package. It is only necessary to supply a constant DC source of power, and the unit will produce a DC output which is a function of its core position. These units are manufactured by Hewlett-Packard, in a variety of sizes.

25.2.3 *Photoelectric transducers*

For the range of displacements from about 0·1 mm to 10 mm, a simple photoelectric transducer is often suitable. This consists of a constant light source, a variable width slit, and a photocell, as shown in Fig. 25.4. The lamp must be maintained at a constant voltage from a regulated supply. The collimating lens produces a parallel beam of light, which is intercepted by a slit whose width is varied by the displacement. The light is then focused on a suitable photocell, whose output is a measure of the displacement. For large displacements, a single slit is used; for small displacements, greater sensitivity can be obtained if a number of parallel slits are employed.

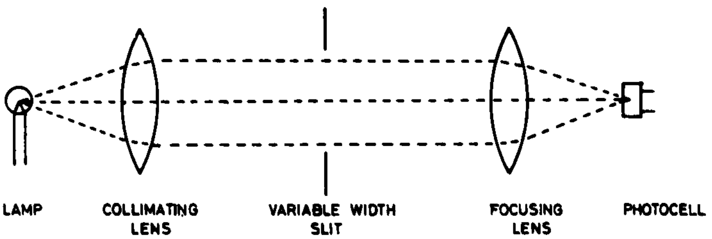


FIG. 25.4. Optical transducer.

A number of *digital output* optical methods are available, for both angular and linear displacements. If it is only necessary to measure the amount of a displacement, irrespective of its direction, a light source, fixed slit, and photocell may be used in conjunction with a graticule attached to the object whose displacement is to be measured (Fig. 25.5). The graticule may be made photographically from a much larger India ink drawing, and attached to a Perspex or glass strip.

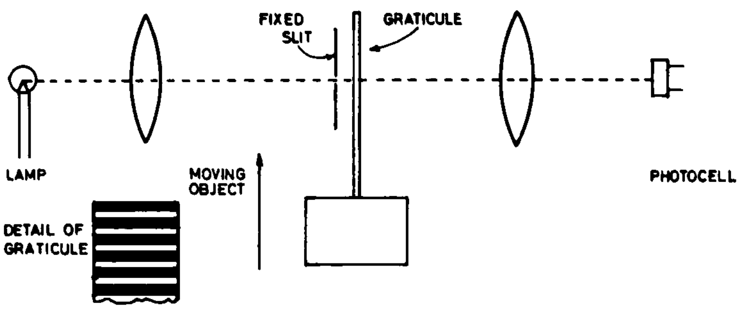


FIG. 25.5. Digital displacement transducer.

The output of the photocell is a series of rectangular pulses, whose number represents the amount of displacement. These pulses may then be passed to a digital counter, which is set to zero when the object is in its initial position.

To obtain a digital readout which also takes into account direction of travel, a second fixed slit is placed alongside the first, but displaced from it by a *quarter of the graticule spacing*. Its light output passes on to a second photocell (Fig. 25.6). As the graticule

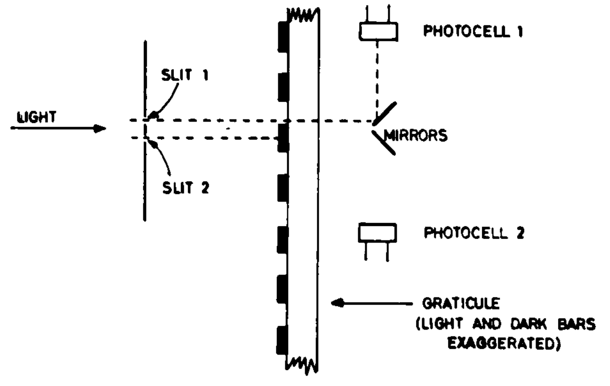


FIG. 25.6. Arrangement to give direction of movement.

moves *up* photocell 2 will always receive light *before* photocell 1; as the graticule moves *down*, photocell 2 will receive light *after* photocell 1. Fig. 25.7 shows the outline of a logic circuit for separating these two situations. The outputs of both photocells are first sharpened by Schmitt triggers; photocell 1 output is then passed through a monostable, which gives a short positive pulse out whenever it encounters a *leading edge* of a pulse. Photocell 2 output is fed into two AND gates, one by way of an inverter. It will be seen that there will be a series of output pulses from the upper gate if the graticule is moving up, and a series from the lower gate if the graticule is moving down. These are fed into a counter with two inputs; one input adds counts to its reading, and the other subtracts them.

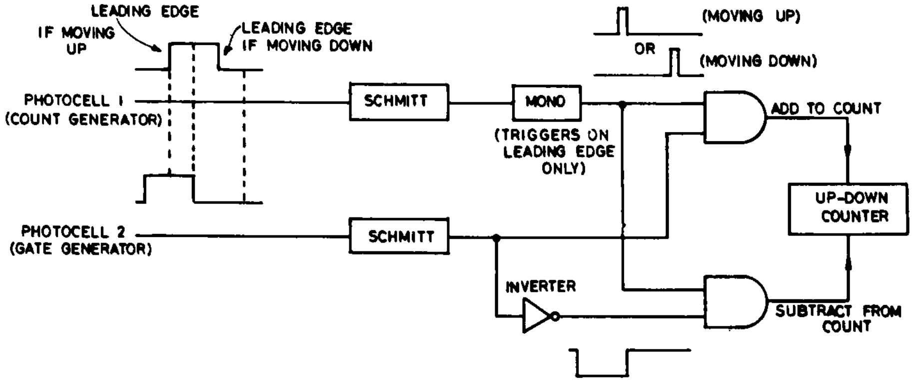


FIG. 25.7. Directional displacement transducer.

For smaller displacements or greater accuracy, the same arrangements as discussed above may be used, but with the slits and graticule replaced by a fixed and a moving graticule. Each has alternate light and dark bands, but one has, say, 10 bands per cm and the other 11. These will generate *moiré fringes*; as the moving graticule is displaced, a series of dark bands appears to travel from one end of it to the other. For each movement of 1 mm in the example above, one band will travel its length. These may be counted, and their direction of movement detected, exactly as before. In practice the rulings on the graticules may be very fine, and extremely small displacements recorded.

By the use of a graticule with multiple tracks, and a photocell for each track, it is possible to produce an output which represents the graticule position directly in binary or binary coded decimal form, and this is often done. *Shaft encoders* for measurement of angular position are available commercially, and often provide a simple and economical solution to the problem of encoding an angular position. The first four tracks of a pure binary encoding graticule are shown in Fig. 25.8. This method has the advantage of transmitting a shaft position unambiguously, without needing any initial zero setting. It will be seen that even with only four bits the position can be defined to the nearest 22·5°.

Pure binary encoding has the disadvantage that several bits change value together at a number of points around the periphery; if one changes slightly before another it is possible for a grossly incorrect reading to be transmitted in these positions. For this reason a special code, known as *Gray code*, has been developed for encoders; in this code only one bit at a time changes at any transition region. Gray code is readily translated into binary by means of a standard diode matrix, and packaged converters are available for this purpose.

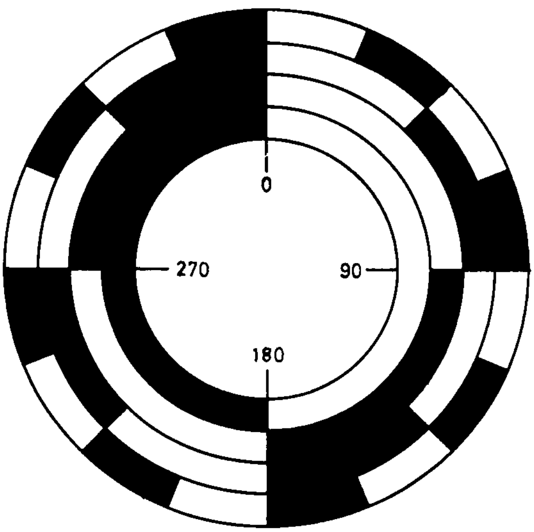


FIG. 25.8. Four-bit binary shaft encoder.

25.3 Velocity transducers

25.3.1 *Velocity by computation*

Velocity may be obtained from a displacement transducer, by differentiation, or from an acceleration transducer, by integration. Under some circumstances differentiation is possible, but it must be realised that any noise present in the original signal will be greatly magnified by this process; for example, differentiation of the output of a wire-wound potentiometer is quite unacceptable. Integration of an accelerometer output is possible, provided that an adequately drift-free accelerometer is used in the first place.

25.3.2 *Electromagnetic transducer*

If a conductor of length d metre is caused to move at right angles to a magnetic field of density B tesla at a velocity S metre/sec, an e.m.f. of E volts will appear across its ends; E is given by

$$E = dBS$$

Since a density of 0·2 T is about the maximum readily available, it will be seen that a conductor 1 cm long can produce an e.m.f. of 20μV per centimetre per second of velocity. Most types of velocity transducer are based on this principle.

Packaged velocity transducers containing a permanent magnet and a coil are available; these produce outputs of a magnitude which can be used without amplification. A buffer unity gain operational amplifier is generally used in conjunction with them, to drive a meter or output circuit.

25.3.3 *Electromagnetic fluid flowmeter*

The principle of the electromagnetic transducer may also be used for measurement of fluid flow velocity. Provided that the fluid is a conductor, it will generate an e.m.f. proportional to its velocity if placed in a magnetic field at right angles to its direction of flow. However, the steady e.m.f. of the small magnitude occurring in practice is very difficult to distinguish from small error potentials inevitable with the use of reversible electrodes in contact with the fluid. The use of an alternating field will produce an alternating potential at the same frequency, and proportional to the velocity, but also it generates another alternating potential in the associated connecting leads. Fortunately, this latter potential is in a phase at 90° to the velocity component, so they can be separated by suitable circuitry.

An alternative approach is to use a square-wave magnetic field, which is the equivalent to a constant field, but reversed often enough to eliminate error potentials. This also produces spurious potentials each time it reverses, but by suitable gating, these potentials can be eliminated at the output of the associated amplifier.

Instruments using both these principles are commercially available; they can be used with fixed probes mounted in a plastic cannula, or the magnet and probes may be applied directly to the walls of an intact blood vessel. In the latter case the accuracy is lower, but much less damage is done to the circulating blood. It is quite practicable to leave the whole transducer assembly chronically implanted about a blood vessel in an experimental animal.

25.3.4 *Ultrasonic (Doppler) fluid flowmeter*

Doppler flowmeters depend on the fact that when a sound wave is reflected from a substance moving in a direction along the line of the sound wave, its apparent frequency is altered. The arrangement of a typical flowmeter is shown in Fig. 25.9. A high frequency oscillator drives a *transmitting crystal* (see § 25.4, piezoelectricity) in the flowmeter probe; this is coupled acoustically to the desired blood vessel through the intervening tissues and a blob of gelatinous coupling compound. The reflected frequency, which is slightly different, is picked up by a receiving crystal, amplified, and multiplied by a sample of the original frequency. The output contains the two original inputs, together with a signal at a frequency representing their difference. This difference frequency, which is proportional to the flow velocity, is separated by a low-pass filter, and its value displayed on a ratemeter calibrated in flow velocity.

Doppler flowmeters have outstanding possibilities, particularly since they are non-invasive. In practice they are excellent for relative measurements, and as indicators that flow in a vessel has been occluded, but are rather difficult to use quantitatively.

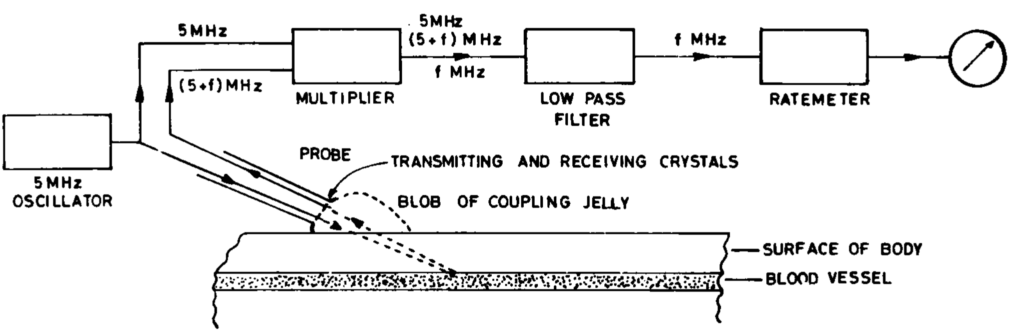


FIG. 25.9. Principle of Doppler flowmeter.

25.3.5 *Resistance flowmeter (gas)*

The measurement of gas flow velocity is of considerable importance in studies on respiratory function. In modern practice, this is done by introducing a resistance into the tube through which the gas is flowing, and measuring the pressure drop across it by means of a thin diaphragm, whose deflection is sensed by an LVDT (see § 25.6). A typical arrangement is shown in Fig. 25.10. The resistance offered to flow must be kept to a minimum, to avoid interference with the patient's normal respiratory pattern. The entry and exit tubes to the resistance are typically 25 mm in diameter, and the resistance itself consists of a sheet of gauze 75 mm in diameter, and having twelve to sixteen wires per mm. Pressure drop across this system is accurately proportional to flow velocity; the value of the resistance needs to be determined at one point only (in terms of pressure drop per litre/sec of flow).

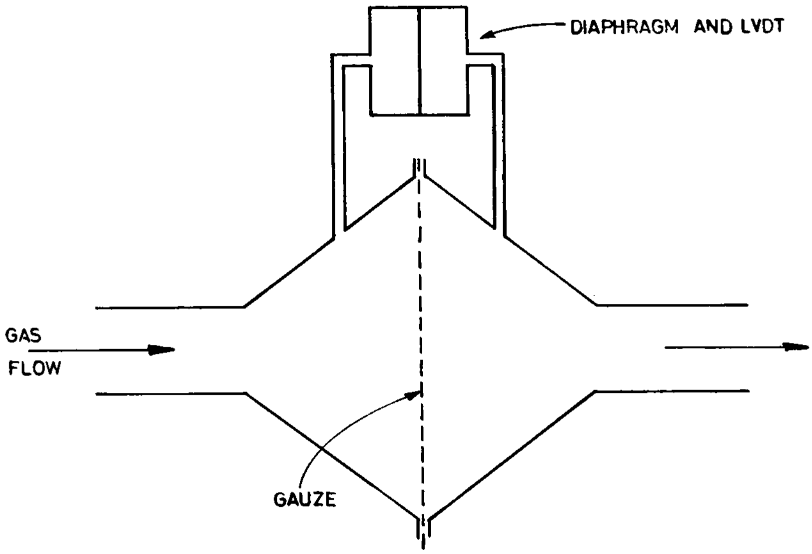


FIG. 25.10. Resistance to flowmeter.

25.4 FORCE TRANSDUCERS

25.4.1 *Resistance strain gauge*

If a wire is subjected to tension, it both increases in length and decreases in diameter; provided that the tension is within the elastic limits of the wire, it will return to its original shape when the tension is removed. Both the increase in length and the decrease in diameter contribute to an increase in electrical resistance, which is quite accurately proportional to the tension applied. The *strain* in the wire is defined as extension per unit length, $\delta l/l$. The increase in the resistance per unit resistance, $\delta R/R$, is related to the strain by a *gauge factor*, which depends on the wire material, and is usually about 2. The maximum change in resistance that can be produced without exceeding the elastic limits of a wire is usually about 1 %. By connecting the wire as one arm of a Wheatstone bridge, the tension applied can be measured. However, the temperature coefficient of resistance of the wire will be about 1 % for every 2·5°C, so this system is unusable in practice. To compensate for ambient temperature changes, it is usual to make all four arms of the bridge tension-detecting elements; this also gives four times the output for a given tension.

Two types of gauge are in common use. The *unbonded* gauge is shown diagrammatically in Fig. 25.11. The four tension-detecting elements are set up initially to be under tension; an added tension then increases the strain on B and C, and reduces it on A and

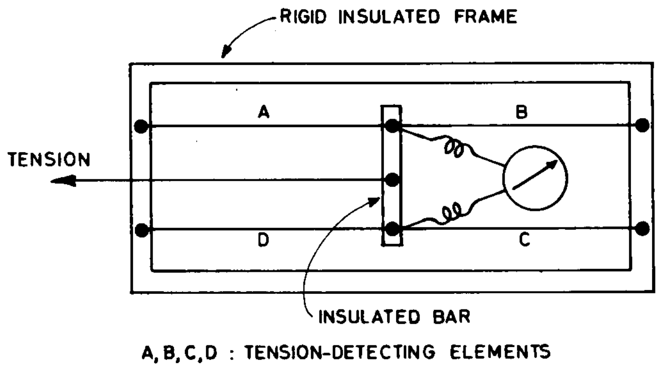


FIG. 25.11. Unbonded strain gauge.

D. The four elements are *cross-connected* as a Wheatstone bridge, as shown in Fig. 25.12. Using these connections, the changes in all four arms contribute to the output, but a temperature change merely alters all four simultaneously, and has no effect.

Many commercial forms of unbonded gauge are available; they are often used in parallel with a much heavier spring, to increase the force required to give maximum output.

The *bonded* gauge consists of a very fine etched metal film attached to a thin elastic backing; in use four gauges are rigidly cemented by an epoxy resin to the surface whose extension is to be measured, and connected as a Wheatstone bridge, as before, so that two are lengthened and two either unchanged or shortened. One of the most satisfactory ways of using bonded gauges is in conjunction with a *proving ring*. This is a metal ring of suitable dimensions; force is applied across it, and the gauges are attached to its surfaces, as shown in Fig. 25.13.

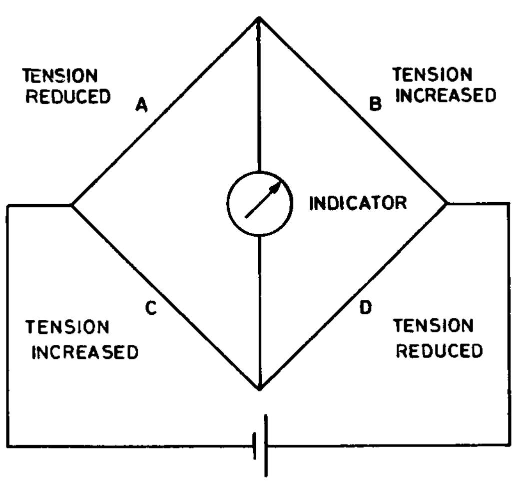


FIG. 25.12. Strain gauge bridge connections.

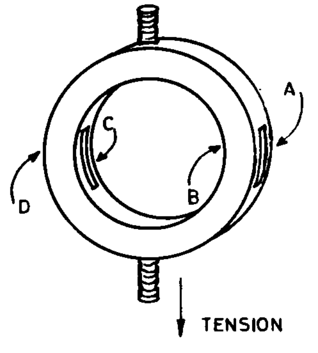


FIG. 25.13. Proving ring.

The maximum extension of a proving ring should not exceed about 0·5% of its mean diameter; this is often expressed, in terms of *microstrain* ($\delta l/l \times 10^6$), as 5000. If this is exceeded, permanent deformation will result. On the other hand, if the full extension of the ring is not used, the electrical output will be low, and slow drifts in zero due to temperature changes will become significant. The basic design equation for a proving ring is:

$$P_{max} = \frac{5{\cdot}51 \times 10^{-3} bh^3 E}{r^2} \tag{25.2}$$

where P_{max} is the maximum permissible force (newton)
b is the width of the ring(metre)
h is the thickness of the ring(metre)
r is the mean radius of the ring(metre)
E is Young's modulus of elasticity for the material used (newton/metre^2)

For mild steel, $E = 2{\cdot}1 \times 10^{11}$ N/m^2, and the equation then becomes:

$$P_{max} = \frac{1{\cdot}57 \times 10^{9} bh^3}{r^2} \tag{25.3}$$

If G is the gauge factor of the gauges used, the sensitivity K of this arrangement, in microvolt/newton for each volt applied to the bridge, is given by

$$K = \frac{1{\cdot}04 \times 10^{-5}\, r}{bh^2} \tag{25.4}$$

Table 25.1 shows the characteristics of a typical proving ring, using Philips bonded gauges type PR9211. These gauges have a nominal resistance of 120 ohms each. Bonded gauges of semiconductor materials are available. These have a very high gauge factor, but they also have a very high temperature sensitivity, are expensive, and have limited life. They are not recommended for most purposes.

Bonded gauges are sold in matched sets. Only gauges from the same set should be combined into a bridge, or the temperature effects will be excessive.

TABLE 25.1. CHARACTERISTICS OF TYPICAL PROVING RING STRAIN GAUGE

Material	mild steel
Mean radius	$2{\cdot}25 \times 10^{-2}$ metre
Thickness	$2{\cdot}4 \times 10^{-3}$ metre
Width	$1{\cdot}28 \times 10^{-2}$ metre
P_{max} (calculated from eqn. (25.2))	383 N
Bridge supply voltage	5·0 V
Bridge output (meas)	16·5 μV/N; 6·32 mV at P_{max}
(calc)	16·0 μV/N

25.4.2 *Piezoelectric strain gauge*

These devices depend on the fact that many crystals show the property of *piezoelectricity*. If stressed along a suitable axis, an electric charge appears between opposite faces of the crystal. Conversely, if a voltage is applied between these faces, the crystal deforms along the original axis. Materials normally used in electronic equipment are either quartz, or synthetics such as barium titanate.

If the stress along the axis is maintained, the charge produced slowly leaks away, both around the crystal and through the associated measuring circuit. This limits the use of piezoelectric strain gauges to the measurement of sudden changes in applied force, or of alternating forces, rather than to the measurement of slowly varying or steady forces. They are always used with a FET input operational amplifier, usually connected in the unity-gain non-inverting configuration.

25.4.3 *Displacement transducers as force transducers*

By using a sensitive displacement transducer such as an LVDT or DCDT in parallel with a spring or proving ring, a perfectly satisfactory strain gauge may be set up. The main disadvantage of this arrangement is its physical size; it is often used.

25.4.4 *Resonance in force transducers*

Whenever a force is applied to a spring, a *mechanically resonant system* is set up. As in the case of the water bath temperature of Chapter 24 (Fig. 24.6), a change imposed on the extension of the spring will produce one of three possible types of behaviour, depending on the amount of damping imposed on the system by air resistance, internal energy loss in the spring, and any deliberately introduced energy losses in parallel with the spring. Unless deliberate measures are taken, almost every force transducer is underdamped, and will vibrate for some time after the application of a force, at a frequency dependent on the stiffness of the transducer and the mass associated with the system. The output from the transducer then gives a most misleading picture of the time course of the applied force.

In many cases it is not practical to add sufficient mechanical damping to the system to bring it into the region of critical damping, though this is sometimes done in commercial instruments. (Slight underdamping is common in these cases, to give a faster rise time and linear phase shift at the expense of a slight overshoot; a typical value is a *damping factor* of 0·7, which gives 4% overshoot when a sudden step of force is applied.)

The usual procedure is to tolerate the vibration which occurs, but to use a low-pass filter in the associated amplifier to eliminate it at the output. The rise time of the system will of course be grossly reduced, but this may still give a perfectly acceptable instrument.

If the stiffness of the transducer is measured (in terms of force to produce unit extension) and the mass in the system is known, then the natural resonant frequency is given by

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}}, \tag{25.5}$$

where f is the frequency (Hz)
k is the stiffness (N/m)
m is the system mass (kg)

For the proving ring of Fig. 25.13, k may be calculated from P_{max}:

$$k = \frac{P_{max}}{r} \times 10^2. \tag{25.6}$$

For example, if the proving ring of Table 25.1 has a mass of 1 kg attached to it, the natural resonant frequency may be calculated as follows:

$$k = \frac{383 \times 10^4}{2{\cdot}25} = 1{\cdot}70 \times 10^6$$

$$f = \frac{1}{2\pi}\sqrt{\frac{1{\cdot}70 \times 10^6}{1}}$$

$$= 207 \text{ Hz}.$$

To obtain adequate filtering in the associated amplifier, an active low-pass filter of the type described in Chapter 12, with f_0 at 40 Hz, would be adequate. The overall rise time of the system would then be about 25 msec. If this is too slow, either the mass must be reduced, or a stiffer proving ring used; in the latter case the electrical output will of course be reduced.

When a bonded strain gauge bridge is used for small deflections only, it will give outputs of 100 μV or less. Particular care must then be taken in the choice of the operational amplifier, to avoid thermal drift in it. A high stability type such as the Analog Devices 260J should then be used rather than a μA741.

25.5 Accelerometers

25.5.1 *Principle of the accelerometer*

If a mass attached to a force transducer is subjected to acceleration, a force proportional to the acceleration will be developed, and an electrical output proportional to the acceleration will be obtained. The problem of damping must be carefully considered in these instruments; oscillations may be set up by any sudden change in the acceleration being measured. Many accelerometers are filled with oil; others use air damping, or some form of electrical damping. A damping factor of 0·7 is common.

25.5.2 *Piezoelectric accelerometer*

Many accelerometers are of this type; although they are simple and inexpensive, they have all the limitations of piezoelectric force transducers, and also an extremely high temperature coefficient. They are quite suitable for vibration measurements at frequencies of a few hertz and up, but not for relatively slow sustained accelerations.

25.5.3 *Displacement accelerometer*

A wide variety of accelerometers suitable for measurement of slow sustained accelerations is available. Most of these contain a mass, a spring, some damping system, and an LVDT or DCDT to measure the displacement produced. A typical unit has a frequency response from 0 to 400 Hz, an output of 50 mV per m/sec^2, and a weight of 80 g.

The best units contain an electrical feedback system; as the mass commences to displace under acceleration, its movement is sensed, and an opposing force is generated by a current passed through a core in a magnetic field. The magnitude of this current is a measure of the force produced, and hence of the acceleration experienced.

25.6 Pressure measurement

Gas and fluid pressures are measured by permitting them to deflect a diaphragm, and using a displacement transducer to measure the deflection. Provided that the deflection is less than about one-third the thickness of the diaphragm, it will be directly proportional to the applied pressure. The two most commonly used displacement transducers in this application are the LVDT (or DCDT) and the unbonded strain gauge; both are equally suitable.

The types of pressure transducer most often required in the biomedical field are for low pressure respiratory measurement, for venous blood pressure measurements, and for arterial blood pressure measurements. The first two of these are not very critical as regards design, since very rapid response to pressure changes is not required. The arterial transducer, however, must be very carefully designed. In particular, its diaphragm must be small in diameter, and must deflect by a minimum amount under its working pressure, otherwise it will be unable to respond to rapid pressure changes when connected to an artery through a needle or catheter. The maximum acceptable volume change in the measuring chamber is about 0·04 mm^3 for a pressure change of 100 mm of mercury. This figure represents a hydraulic capacitance, which in conjunction with the hydraulic resistance of a needle or catheter forms a low-pass filter when measuring a fluctuating blood pressure.

FURTHER READING

DONALDSON, *Electronic Apparatus for Biological Research*, Butterworths, London, 1958.
HANSEN, *Pressure Movement in the Human Organism*, Teknisk Forlag, Copenhagen, 1949.
TIMOSHENKO, *Strength of Materials* (2 vol.), Van Nostrand, New Jersey, 1955.
OSTER and NISHIJIMA, Moiré patterns, *Scientific American*, May 1963.
KUHL, *Strain Gauges*, Philips Elektronik Industrie GmbH, Hamburg.

PRACTICAL

25.1 Construct a linear variable differential transformer, as described in this chapter. Supply the input from a low voltage secondary power transformer, using a series resistor if necessary, and observe the output with an oscilloscope as a piece of soft iron is moved through the null position. Set up a balanced demodulator as in Fig. 25.3, using a μA796 or the equivalent; drive it from the LVDT, and observe its output on the oscilloscope.
25.2 Construct a pair of moiré gratings by ruling thick India ink lines at regular close intervals on two strips of tracing paper, as described in this chapter. Place one behind the other, hold them up to the light, and move one with respect to the other. Which way do the fringes move? Why?
25.3 Observe the operation of any other types of transducer which may be available, and carry out calibration procedures on them.
25.4 If two arterial pressure transducers and recorders are available, connect them both to a fluid-filled system consisting of a rubber bulb and length of thin-walled rubber tubing. One connection should be made through a tube of 2 mm diameter or more, the other through a thin (21 gauge or less) 75 mm hypodermic needle. Press the bulb rhythmically, and observe the phase and amplitude of the two pressure tracings. What is the hydraulic time constant of the needle-transducer system?

CHAPTER 26

SOME BIOLOGICAL ANALYTICAL METHODS

26.1 Scope

A modern biochemical or respiratory function laboratory contains a considerable number of highly sophisticated and often automatic analytical instruments, whose description is beyond the scope of this book. The basic principles used in them, however, are quite simple, and this chapter will outline some of the more common methods, and the instrumentation required.

26.2 Ionic composition of tissue fluids

26.2.1 *Sodium and potassium*

These ions are routinely estimated by the flame spectrophotometer. In this instrument a burner using an air–gas mixture produces a normally non-luminous flame. Into this flame is introduced a fine spray of a dilute solution of the ions to be estimated, at a constant rate. Solute atoms are excited at the high temperature; outer shell electrons are moved up to higher energy levels, and in falling back emit quanta of light at wavelengths characteristic of the atomic species present. This emitted light is filtered into its separate wavelengths by optical means, and the intensity of one wavelength corresponding to each of the species present is measured by means of a photocell, and recorded on a meter or chart recorder.

A typical concentration of sodium or potassium ion sprayed into the flame is about 0·5 mmol/l, which is quite dilute; considerable care must be taken to use very pure distilled water and scrupulously clean glassware.

The optical filtering is commonly carried out by means of suitably dyed transparent films, although the better instruments use *half-wave filters*, in which two glass plates are mounted with great accuracy at a spacing of a half wavelength at the desired optical wavelength. For laboratory use a *prism* or *grating spectrometer* may be used for analysis; in this instrument the wavelength to be measured may be continuously varied by rotating a prism or optical grating by means of a finely calibrated dial.

In simple instruments the photocell is used directly as a source of current, which is measured by being passed through a sensitive microammeter. With care an accuracy of about 1% can be obtained from an instrument of this type, by repeated comparison between a standard solution and the unknown one. The more expensive instruments use a photomultiplier tube, whose output current is passed through a resistor, and the resulting voltage drop measured by a digital voltmeter or by a potentiometer. The reference by Kay at the end of this chapter should be consulted for a comprehensive review of these techniques.

26.2.2 *Calcium, magnesium, and trace ions*

Many ions, including calcium and magnesium, are insufficiently ionised in an air–gas flame to emit much light, even if acetylene is used as the gas. These, however, may readily be estimated by use of an *atomic absorption spectrometer*. This instrument consists of an air–gas burner (usually air–acetylene) into which the solution to be estimated is sprayed as before. Light is passed *through the flame* from a special discharge lamp containing a trace of the element being investigated; the emitted light from this lamp then contains the characteristic wavelengths for that element. This light is received by a grating spectrometer and photomultiplier, which operates an output digital voltmeter or potentiometer as before. The spectrometer is set to one of the wavelengths from the lamp, and the meter is adjusted to give a *full scale reading* when distilled water only is sprayed. When the solution to be estimated is sprayed, the flame *absorbs* energy from the lamp if the element being investigated is present, and the meter reading *drops* by an amount which is a function of the amount of element present.

Very small concentrations of ions may be estimated in this fashion; the method is particularly useful for detecting potentially toxic amounts of heavy metals, in foodstuffs or accumulated in the body.

26.2.3 *Chloride*

Chloride estimations are commonly required in laboratory investigations, although they tend to be avoided in clinical biochemistry. They are usually performed by titration with a solution containing silver ions; the end-point is reached when the precipitation of solid silver chloride just ceases. In practice this is impossible to determine by eye, and an *electrometric titration* is performed. This is based on the concept of reversibility of a silver electrode, as discussed in Chapter 22. If two silver electrodes have a potential of about 0·25 V applied between them, as in Fig. 26.1, a significant current will flow in the

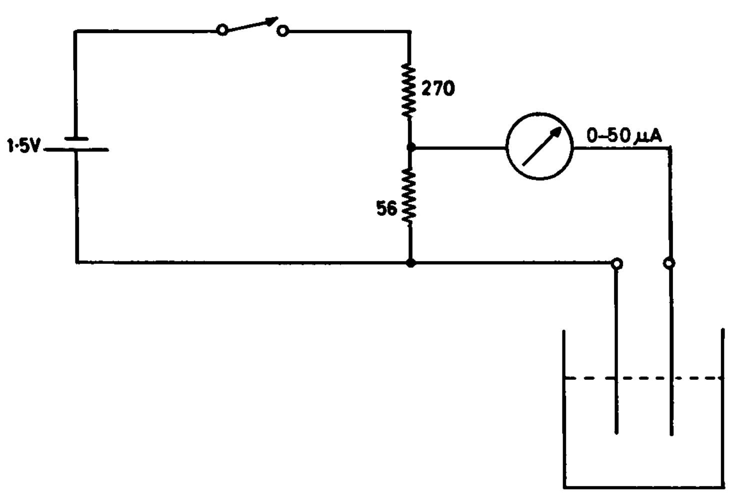


FIG. 26.1. Electrometric endpoint indication for chloride titration.

circuit only if free silver ions are present in the solution; this can occur only after all chloride ions have been precipitated. The electrodes are wires of spectroscopic grade silver, of about 1 mm diameter. To give a sharp end-point, the conductivity of the solution is increased by the use of a *supporting electrolyte.* It is usual to dilute the original chloride-containing biological solution considerably by a known fraction and this is done using a solution of 0·1 mol/l. chloride-free nitric acid with 10% acetic acid added, rather than with distilled water. A few drops of chloride-free gelatin solution are added to give a uniform precipitate, and the solution is well stirred during the titration. The end-point is reached when the meter suddenly commences to swing upward; an arbitrary point of 20 μA is suitable as a threshold. The small zero error this produces may be allowed for by titrating a chloride-free solution.

Instead of adding silver ion by means of a burette in the conventional fashion, Cotlove's method may be used. Silver ion is added electrolytically, by inserting two more silver electrodes in the solution, and passing a constant current between them (9·46 mA will precipitate 0·1 μmol/sec). The time for which the current is passed before the end-point is reached is then a measure of the amount of chloride present.

26.3 Estimation of oxygen and carbon dioxide in biological solutions

26.3.1 *Oxygen in solution*

Oxygen concentration in solution is most simply estimated electrolytically. The basic principle is that if a bright platinum electrode is placed in a solution, and +0·7 V is applied to it with respect to a reversible electrode (usually silver–silver chloride in biological systems), oxygen will be reduced at it, and the current flowing is a measure of the rate at which oxygen is being reduced. By enclosing the platinum electrode in a film of collodion or other material permeable to the solution, and keeping the solution in contact with its outer surface well stirred, a constant flow of oxygen is set up through the film, proportional to the external oxygen concentration, and so a current flow proportional to oxygen concentration is obtained. A number of commercially produced oxygen electrodes are available. Typical techniques are described in the review by Kay.

26.3.2 *Carbon dioxide in solution*

The ability of whole blood to transport carbon dioxide from the body tissues to the lungs is of vital importance, and an estimate of this function is frequently required. In modern practice the estimation is carried out on a single drop of blood (refer to the paper by Sigguard *et al.* cited at the end of this chapter). The blood sample is held at a constant temperature, and its hydrogen ion concentration measured electrometrically by a pH electrode, as described in § 26.5. It is then equilibrated in turn with two known concentrations of carbon dioxide gas, and its pH redetermined for each; a calculation using the Henderson–Hasselbalch equation then gives the original blood bicarbonate level, and hence its ability to transport carbon dioxide.

26.4 Gas analysis methods

26.4.1 *Oxygen*

The measurement of gaseous oxygen is often required in respiratory studies. Unfortunately, apart from its oxidising properties, oxygen has few characteristic features. It is very weakly *paramagnetic*; the relative permeability of pure oxygen at atmospheric pressure is 1·0000018, as compared with 1·0000000 for a vacuum. This difference is sufficient to allow its estimation; the most satisfactory and simple method is due to Pauling. A number of commercial oxygen meters using this principle are available. A small glass "dumbbell" is suspended by a fine quartz fibre on a non-uniform magnetic field, and the gas to be analysed is admitted so that it surrounds the dumbbell. In the absence of oxygen, the dumbbell takes up a position due to the torque of the fibre alone. If oxygen is present, it is concentrated very slightly in the magnetic field, and the dumbbell is forced to rotate. The degree of rotation depends on the concentration of oxygen present; this can either be measured simply by means of a mirror attached to the fibre, which deflects a light beam across a scale, or by a null method in which the torque due to the oxygen is exactly balanced by passing a current through a coil attached to the dumbbell, and measuring the current required.

Since the suspension is quite fragile, the gas to be measured must be admitted slowly through a small leak. This gives a slow response, several seconds being required to take a reading.

It is often desired to follow the breath-to-breath changes in oxygen concentration of expired gas; this can only be done by fairly elaborate methods. The very small change in inductance of an inductor when immersed in the gas sample may be measured, or a mass spectrometer used to determine the amount of oxygen present by using its atomic mass to characterise it. The first method is not commercially available; instruments for the second are available, and are relatively simple to use, but are quite bulky and very expensive.

26.4.2 *Carbon dioxide*

Carbon dioxide is characterised by its very poor thermal conductivity, and by its infra-red absorption spectrum.

Thermal conductivity carbon dioxide meters consist of a Wheatstone bridge made up of four fine platinum wires, each suspended down the axis of a cylindrical cavity drilled in a copper block. Two of the cavities are air-filled; the gas sample to be analysed is slowly admitted to the other two. The two wires surrounded by the carbon dioxide lose heat more slowly than the two surrounded by air, and consequently their temperature and their resistance rise. The bridge is then unbalanced by an amount depending on the carbon dioxide concentration.

This method is quite effective and reasonably cheap, but since the system must come to a steady state before a reading can be taken, its response is quite slow.

For following breath-to-breath changes, the infra-red absorption of carbon dioxide is used in a number of commercial instruments. The principle is shown in Fig. 26.2. The output from a source of infra-red radiation is passed through two parallel tubes, each

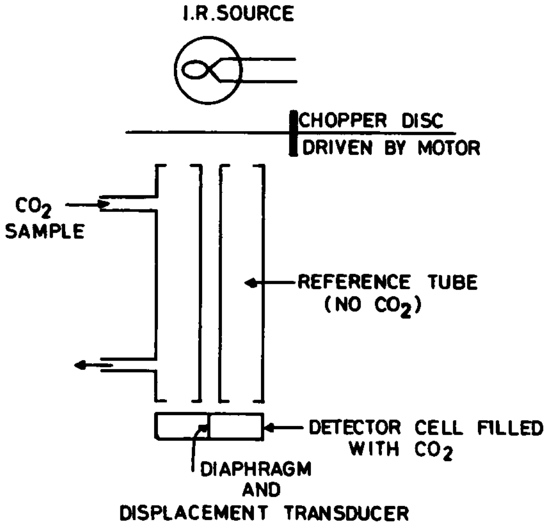


FIG. 26.2. Infra-red carbon dioxide meter.

closed at either end by a window transparent to infra-red radiation. The beam entering the tubes is periodically interrupted by a chopper disc, at a rate of about 10 Hz. At the far end, the two beams fall on the two compartments of a detector cell, which are divided by a diaphragm; both compartments are filled with carbon dioxide. The diaphragm is attached to a sensitive displacement transducer.

One of the two long tubes is filled with carbon dioxide-free dry air; the other contains the dried carbon-dioxide-containing sample of gas to be analysed. This sample absorbs infra-red radiation, while the air does not; thus in each pulse of radiation more falls on the right-hand side of the detector cell (Fig. 26.2) than on the left. The left side is thus heated less than the right, the pressure in it rises less, and the diaphragm is driven to the left. In between pulses, the detection cell cools, and the diaphragm moves back. In this way the diaphragm deflects at a 10 Hz rate, by an amount depending on the carbon dioxide concentration in the sample.

Instruments of this type are relatively bulky; they can however respond very rapidly to changing gas concentrations, and may be made sufficiently sensitive to handle not only the gas concentrations in human respiration (about 3000 parts per million) but the much lower atmospheric concentrations (about 30 parts per million) vital for the respiration of green plants.

26.4.3 *Helium*

Even at the end of a forceful expiration, human lungs still normally contain about 2 l. of air. The determination of this residual volume is often of clinical significance; it is frequently determined by a dilution method, which consists of adding a known volume of pure helium to a closed system into which the patient is breathing. After allowing for complete mixing, a sample is withdrawn and analysed for helium. Since the *external* volume of the closed system is known, and the concentration of the diluted helium is known, the internal volume of the patient is easily determined.

The helium analysis is usually done by a thermal conductivity method, as for carbon dioxide, and often in the same instrument. (Any carbon dioxide in the gas sample must of course be removed before the estimation is carried out.) The use of a mass spectrometer is also entirely possible, if one is available.

26.4.4 *Nitrogen*

Nitrogen is not at all easy to measure, and is usually estimated by subtraction. The only satisfactory method of direct measurement is by mass spectrometer.

26.5 Specific ion-sensitive electrodes

26.5.1 *Concentration and pH*

pH, which is the negative logarithm of the hydrogen ion concentration in a solution, is one of the most commonly required measurements in the biological field. It is normally carried out electrometrically by the use of a *glass electrode*, which is essentially a special thin glass membrane containing a reference electrode in hydrochloric acid. This arrangement behaves in every respect as an electrode reversible to hydrogen ion, and consequently develops a potential proportional to the logarithm of the hydrogen ion concentration. It is used in conjunction with a reference electrode dipped into the solution being measured. Such a reference electrode commonly consists of a silver–silver chloride electrode immersed in potassium chloride, with the latter brought into contact with the unknown solution through a porous glass plug or glass capillary. Although very thin, the glass electrode has a very high resistance, and the system requires the use of a unity-gain non-inverting FET input amplifier circuit to avoid drawing an error-producing current. The voltage reading is usually displayed on a digital voltmeter; an amplifier of adjustable gain is inserted between the input operational amplifier and the voltmeter to permit direct calibration in pH units. Each individual glass electrode has its own small zero error potential; this is allowed for by a preliminary calibration in a standard *buffer solution* of known pH. The reading is also affected by the temperature of the solution being measured, and a correction for this is applied to the reading.

Where an instrument of this type is found to give an erratic or incorrect reading in a known solution, it is almost always due to a defective glass electrode; the remainder of the equipment should not be blamed until defective electrodes are definitely excluded.

26.5.2 *Glass electrodes for other ions*

Glass electrodes intended for sodium, potassium, chloride and other ions are now freely available. In general they are both bulky and expensive, but may very well meet specific experimental requirements. They should be checked carefully before use.

FURTHER READING

DELAHAY, *New Instrumental Methods in Electrochemistry*, Interscience, New York, 1953.
KOLTHOFF and LINGANE, *Polarography* (2nd ed.), Interscience, New York, 1952.
TSAO and VADNAY, Continuous measurement of transient blood pO_2, *J. App. Physiol.* **15**, 712 (1960).
COTLOVE *et al.*, An instrument and method for chloride titration, *J. Lab. Clin. Med.* **51**, 461 (1958).
SIGGUARD, ANDERSON and ENGEL, Micro determination of pH and CO_2 tension, *Scand. J. Clin. Lab. Invest.* (1960).
KAY, *Experimental Biology*, Chapman & Hall, London, 1964.
MELOAN and KISER, *Instrumental Analysis*, Merrill, Columbus, Ohio, 1963.

PRACTICAL

Practical work will be largely dependent on the apparatus available and the interests of each individual taking the course. Typical experiments are the estimation of sodium and potassium by flame photometer, the estimation of chloride by Cotlove's method, gas analysis of respiratory samples, and pH measurements.

CHAPTER 27

STORAGE AND PROCESSING OF DATA

27.1 TRANSMISSION OF DATA

For any quantitative discussion on the rate of transmission of information from one point to another, or from one person to another, a method of estimating "amount of information" is required. The way in which this is usually done is to consider the change in the *probability* of an event, in the mind of the person receiving information about it. If a man is waiting for a bus, and he is informed by an official "There will be a bus here within the next hour", this statement evidently conveys little information; he already knows that this is very probable. But if he is told "There will be no bus here within the next hour", this conveys a lot of information. In fact, *as far as the prospective traveller is concerned*, the probability of arrival of a bus is altered considerably by the information.

If a prospective father is waiting to learn the sex of his child, and is reliably informed "It's a boy", the probability of the birth of a son is altered in his mind from 0·5 (boy or girl equally likely) to 1·0 (certainty of a boy).

It is convenient to express quantity of information transmitted in logarithmic terms. If an event occurs, and a message is transmitted telling about the occurrence, the amount of information Q transmitted is defined as

$$Q = \log_2 \frac{\begin{array}{l}\text{probability of the event}\\ \text{to the receiver afterwards}\end{array}}{\begin{array}{l}\text{probability of the event}\\ \text{to the receiver before}\end{array}} \tag{27.1}$$

Using this definition, Q will be expressed in *bits*; these are in fact simply binary numbers.

In the case of the prospective father, eqn. (27.1) works out as

$$Q = \log_2 \frac{1{\cdot}0}{0{\cdot}5}$$

$$= \log_2 (2) = 1$$

In other words, one bit of information converts complete uncertainty about a binary choice to complete certainty. In the same way, two bits of information will determine whether the child is a blue-eyed boy, a blue-eyed girl, a boy other than blue-eyed, or a girl other than blue-eyed. By increasing the number of bits transmitted, any degree of complexity or gradation of information can be transmitted.

How much information does the fuel gauge of a car need to transmit? In fact, such gauges are usually divided into four quarters, and this will transmit two bits of information, all that is really needed. A gauge divided into 16, and thus transmitting four bits, would be no more useful. Similarly an engine temperature indicator is usually only a warning light, transmitting one bit of information: too hot or not too hot.

The rate of transmission of information is expressed in bits per second. For any given channel of transmission, a maximum rate of transmission exists, governed by the frequency response of the channel. If information is supplied at a higher rate, part or all of it will be lost. The upper limit of rate of transmission to a human observer, either visually or aurally, is about 2·5 bits per second. This limitation is a vital factor in the design of machines to suit an operator's capabilities.

27.2 Data storage systems

27.2.1 *Types of data storage*

Data storage systems may be divided into two classes: those from which the data may be reconverted into an electrical signal, and those from which it may not. In selecting a storage system, a decision must first be made between these two types.

27.2.2 *Irretrievable systems*

(a) *Pen recorders*

Pen recorders are widely used for storage and display of biological information, and in their modern versions are very reliable, sensitive and accurate. They can readily be manufactured with a frequency response up to 100 Hz or better. Various types of *pen motor* are in use; most are based on the d'Arsonval moving coil meter movement. The pens used are of two types; one is heated by passage of a current, and writes on heat sensitive paper, while the other uses ink. In modern instruments the record is always written on an ordinary rectilinear graph, and this requires either a rather elaborate linkage to convert the radial pen motor movement into a rectilinear pen movement, or the use of a chart passing between an elongated pen and a straight knife edge. The first method is used with ink recording, the second with heat recording.

A good ink recorder will use special paper to give rapid drying of the ink, and some system of sterile ink transfer from reservoir to pen; many of the earlier instruments clogged readily as moulds grew in the ink channels, and spattered ink badly at high frequencies. The line drawn by the pen should be clear and continuous at all chart speeds, pen amplitudes, and frequencies.

A good heat recorder has a pen which is robust, and is tolerant of variations in paper quality. The heat should be automatically varied as the chart speed is varied, with a manual control to allow for various pen amplitudes. The line drawn should be uniform and thin.

Modern recorders of either type are driven by transistor amplifiers; the basic instrument usually has sufficient gain to take a standard 0–1 V signal input. Many instruments also include preamplifiers for ECGs or various types of transducer, and may be obtained with one, two, four, six or eight pen motors to allow comparison of various parameters. In this form they are often referred to as *polygraphs*.

Adequate damping of the pen is important; for reliable operation some form of electrical damping is essential. Pen-to-paper friction should represent a negligible fraction of the total damping, since it varies considerably with paper quality, temperature, and minor changes in pen contact pressure.

Apart from these mechanical considerations, an unfamiliar instrument should be tested electrically before purchase. The frequency and transient responses should be checked; in addition, the use of a triangular wave from a signal generator will quickly show up any lack of linearity at large amplitudes.

(b) *Potentiometric recorders*

Potentiometric recorders have already been discussed in principle in Chapter 4. They are much slower in response than pen recorders, but use larger charts, and are of much higher accuracy. A typical recorder would have a chart width of 250 mm, and the pen would take a second to traverse the full width. The *slewing rate* of a chart recorder is usually expressed in seconds for full chart traverse. Sensitivity is high, usually in the range 0·5 to 100 mV for full scale deflection. Modern recorders have an input impedance of 1 megohm or better; the older units can be used only with sources having a resistance less than about 500 ohms or so.

Recorders are available with two complete potentiometer systems writing on the same chart, to allow simultaneous recording of two variables. Sampling recorders, printing out distinctive dots for six or more variables in rotation, are also in common use, but are not all easy to read with accuracy. A further variant is the X–Y recorder, in which two potentiometer movements drive the pen along two axes at right angles to each other, and so allow the plotting of functions of a variable other than time.

The block diagram of Fig. 27.1 is typical for a simple recorder. The slide wire is fed from a precision regulated power supply, and the voltage picked off from it is connected

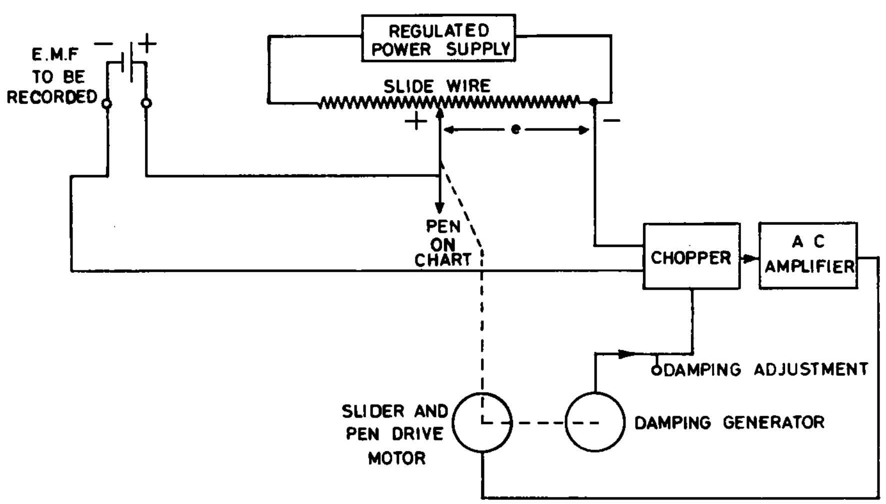


FIG. 27.1. Potentiometric recorder.

in opposition to the e.m.f. to be recorded. The differences between the two voltages is converted into a.c. by a *chopper unit*; in older recorders this is a mechanical vibrating reed carrying suitable contacts, and in more modern units is a solid-state device. The resulting a.c. is amplified and used to drive a small motor which moves the slider and pen. For small values of chopper output the motor speed is proportional to the degree of imbalance of the potentiometer; if however the imbalance is considerable, the amplifier saturates, and the pen *slews* at a constant maximum speed.

On the same shaft as the motor is a small d.c. generator, whose output is proportional to the motor speed. This output is added into the amplifier in such a way that it provides a braking force proportional to the motor speed, and hence gives control of the damping. (In colloquial terms, the faster it goes, the more it can't!) When the damping is correctly adjusted, the pen will move rapidly up to the correct reading when an e.m.f. is applied, and then stop. All precision potentiometers are provided with a standardising circuit; older instruments use a standard Weston cell, while more modern ones have a solid-state precision Zener reference unit. This circuit is not shown in Fig. 27.1.

(c) *Incremental recorders*

Incremental recorders are also *X–Y* recorders, but do not use the potentiometric principle. In them both the pen and the chart are advanced or withdrawn in very small steps (usually 0·25 or 0·125 mm) by means of two *stepping motors*. Each of these produces a single step whenever an electrical impulse of suitable size and duration is fed into it. It will be obvious that these must be used with external circuitry to keep an account of the actual location of the pen on the chart, since the recorder has no automatic return to zero. Incremental recorders can be made to respond very rapidly, at up to several hundred impulses per second. They are used extensively to plot computer-generated graphs, and are readily caused to draw very complex diagrams, and even to add any required figures or explanatory text. Data input may be provided directly from a computer, or from punched paper or magnetic tape (see below).

(d) *Photographic recording of data*

Photographic recording of data, particularly in graphical form, has been widely used in the past. It has the obvious disadvantage that processing is required before the record can be inspected. For frequencies above about 500 Hz, the combination of a camera and

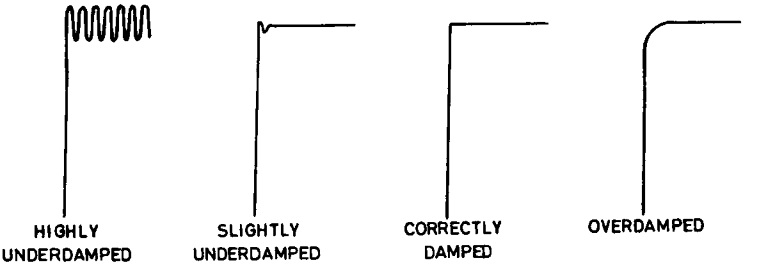


FIG. 27.2. Damping adjustment of potentiometric recorder.

oscilloscope is still the best method available. For a continuous record of non-recurrent phenomena, such as an electromyogram, no time base is used on the oscilloscope, and a strip of film (usually 35 mm) is moved through the camera at constant speed to generate a time base. For recurrent phenomena, or short samples of non-recurrent phenomena, a time base is used, either repetitively or with a single sweep, and the screen is simply photographed in the usual way. This may be done with ordinary film, or by the use of "Polaroid" film. The latter is a very fast film, and gives immediate results, but is quite expensive.

Single frame photography on ordinary 35 mm film may be processed rapidly and effectively by developing in the usual way, enlarging on to special printing paper to bring each print up to about 125 $\times$ 100 mm, and then passing the prints through the rollers of a special developing tank which takes only a few seconds to give a final picture. By the use of suitable multiplexing, up to eight variables may be recorded simultaneously from a 125 mm oscilloscope screen in this fashion.

For multi-channel recording of phenomena in the range 0 to about 5 kHz, a bank of mirror galvanometers can be used, in conjunction with wide film moving at constant speed. Special sensitised papers used with an ultra-violet light source have made it possible to obtain a record without processing the film, but the system has a number of serious disadvantages. Many cardiologists favour this type of recording.

(e) *Storage oscilloscopes*

A number of manufacturers of oscilloscopes produce one or more models incorporating special cathode ray tubes and circuitry to retain an image on the screen once a sweep has taken place. Erasure is effected by pressing a switch. These instruments are more expensive than the simple oscilloscope, but they have great value in biological and medical investigations, since a transient phenomenon such as an electromyogram can be examined in detail immediately after its occurrence, and photographed if it is desired to maintain a permanent record.

Storage oscilloscopes are not capable of storing very fast transients; their present limit is about 25 kHz.

(f) *Rolling display*

A number of manufacturers produce oscilloscopes which consist of an ordinary cathode ray tube, together with digital data storage circuits (see below) which hold a sample of data taken over a fixed time interval and which display it repetitively on the screen. This gives an illusion of a stationary display of the sample, with infinite persistence. If at the same time old data at one end of the time interval are dropped, and new data are added continuously at the other, the oscilloscope trace appears to drift across the screen, as though it were being written on a moving chart. This display is far more easily read than that on a P7 phosphor, since it does not fade out. Oscilloscopes of this type with apparent sweep speeds suitable for electrocardiograms or pulse pressure recordings are not much more expensive than ordinary display units.

27.2.3 *Retrievable storage systems*

(a) *Magnetic tape—analogue storage*

Magnetic tape recorders as used for the recording of speech or music are unsuitable as they stand for recording biological data, since they are incapable of reproducing the low frequencies contained in the signals. If, however, a constant carrier frequency in the audible range (for example, 3000 Hz) is generated, and this is then *frequency modulated* by the biological signal, the result is readily recorded or retrieved. Frequency modulation consists of *deviating* the carrier frequency by an *amount* proportional to the *amplitude* of the modulating signal, at a *rate* corresponding to the *frequency* of the modulating signal. Thus if a 3000 Hz carrier is frequency modulated by a 1 V peak 10 Hz signal, it might deviate from, say, 2500 Hz to 3500 Hz and back ten times a second. If the signal rises to 2 V peak, the deviation would then double, giving a variation from 2000 Hz to 4000 Hz and back at ten times a second.

Frequency modulation and demodulation are simply produced by the use of integrated circuits. It is possible to superimpose several frequency-modulated carriers on the one tape, provided that they are sufficiently far apart to allow subsequent separation by active filters.

Analogue tape recorders are particularly useful for temporary data storage. They permit editing of material before a permanent record is made, so that a considerable waste of recording paper or film is avoided. Moreover, the playback can be at a lower speed than the recording, so that the apparent frequency response of a pen recorder is greatly increased; it can also be at a higher speed than the recording, so that hours of record can be quickly scanned for a region of interest. One, two, and four-track recorders are readily available, and recorders having up to sixteen tracks may be obtained if required.

In using data tape recorders, it must be remembered that they are inherently noisy devices. In the best instruments, the noise is about 44 db below the maximum signal. Thus if the maximum signal is 1 V peak to peak, the noise is 6·3 mV peak to peak; if the maximum signal is 4 cm peak to peak on a recorder, the noise is 0·3 mm peak to peak, which is barely visible. If however the recorded signal is recorded at only $\frac{1}{10}$ of maximum amplitude, and the playback is then amplified ten times to compensate, the noise will be 3 mm peak to peak, which is certainly not acceptable. An active low-pass filter is often used following a tape recorder to reduce the noise to an acceptable level; if the bandwidth is reduced four times, the peak-to-peak noise will be halved.

(b) *Magnetic core memory—analogue-to-digital conversion*

Data may be stored indefinitely by *digitising*, and then storing as binary numbers in a magnetic core memory. Digitisation of an analogue signal consists of *sampling* it at regular intervals, and *converting* each sample into a binary number. Sampling is accomplished by means of a sample-and-hold circuit, which consists of an FET gate and capacitor. In the sample mode, the FET gate is conducting, and the capacitor voltage follows the incoming analogue signal. When digitisation is to take place, the FET gate is biased off, and the capacitor is left holding the analogue voltage at the instant of cut-off. This voltage is then converted into a series of binary numbers, representing its amplitude. Several standard methods of doing this are available; one of the commonest is by successive approximation, in which the first bit is determined by whether the voltage is in the upper or lower half of the maximum range, the second bit by whether it is in the upper or lower half of that half, and so on. Digitisation to an accuracy of 10 bits (0·1 %) is common, and integrated package circuits are available for both sample-and-hold and digitisation; the process typically takes 10 μsec.

The resulting 10-bit digitised value of the voltage is stored in a *buffer register* consisting of ten flip-flops, and may then be transferred into a magnetic core memory. In this device each bit is stored by magnetising a very small ring of ferrite, in either one direction or the other; if clockwise represents 0, anticlockwise represents 1. To store 1000 10-bit samples, 10,000 of these rings are required, and this is a relatively small store by modern standards. A magnetic core store is the heart of nearly all present-day computers.

Stored information can be read out of a core store repetitively, and reconverted to analogue form by digital-to-analogue conversion; such converters are also available in integrated form. The analogue output is then a replica of the original analogue input, and may be examined on an oscilloscope, or plotted on a pen or potentiometric recorder.

(c) *Magnetic tape—digital storage*

Digital data may be stored on magnetic tape, using one track for each bit required, and magnetising the track briefly one way or the other to give 0 or 1. A typical data tape is recorded at 550 bits per inch, giving a very economical method of storing large quantities of data in very accurate form. The noise problem of analogue tape recorders is almost completely overcome in this form of storage; even so, each bit is usually recorded twice, to eliminate possible errors due to "drop-outs" where the magnetic material is slightly defective. It is possible to effect direct transfer from incoming data to a digital tape recorder by use of an analogue-to-digital recorder, but it is more usual to pass the data into a core store first, and then to transfer whole blocks of it to the digital tape.

(d) *Punched paper tape—digital storage*

Much biological data is generated at quite slow rates. It is often economical to have a number of laboratories each equipped with an analogue-to-digital converter and a *paper tape digital punch*, which transfers the digital data to paper tape in the form of punched holes for 1s, and blanks for 0s. The resulting tape can then be read at high speed into a core store, using an optoelectronic reader, and the data subsequently plotted out as before.

Paper tape punches operating at speed up to about 1000 bits/sec are available. Older punches are very noisy, but the more modern types are quite unobtrusive.

(e) *Random access memory—digital storage*

Digital stores consisting of large arrays of flip-flops on a single chip are becoming available at quite moderate prices, and undoubtedly this method of data storage will largely replace core storage in biological data handling as prices diminish steadily.

27.3 Data processing

Digital data may be processed in a number of ways to facilitate analysis. This processing in modern practice is carried out in a small computer, which is used in conjunction with an experiment rather in the same manner as is an oscilloscope. Statistical treatment of successive trials in an experiment, such as the calculation of running averages and standard deviations, may be carried out as the trials proceed; histograms of various types may be progressively accumulated. *Digital filtering* of signals is readily done; any filter that may be constructed using active or passive elements may be simulated by calculations of running averages on a curve, and many ideal filters, such as *zero phase shift filters,* which cannot ever be realised with analogue techniques, may be simulated.

One of the most common processes carried out on a succession of trials is that of averaging successive sweeps, point by point. If an experiment is repeated a number of times, the response will usually contain a number of *regular* features, which are related to the experimenter's manipulations, and a number of random features, which are not. By taking the average of a series of sweeps, the regular features remain constant; random features ideally diminish in amplitude in proportion to $\sqrt{n}$, where n is the number of experiments. The net result is a gain in *signal-to-noise ratio* of $\sqrt{n}$. In practice the "random" features are never truly random, and the improvement in signal-to-noise ratio is not quite as good as $\sqrt{n}$; the technique is, however very valuable, and improvements as great as 10 times in signal-to-noise ratio are quite possible. The limitation to the process is the number of repetitions which may be carried out before the experimental material has varied appreciably by fatigue, drying out, or similar changes. It is essential to calculate error bars of standard deviation, *and to show them in published results,* whenever digital averaging is used.

FURTHER READING

Goldman, *Information Theory*, Prentice-Hall, New York, 1952.
Renwick, *Digital Storage Systems*, Spon, London, 1964.

PRACTICAL

27.1 Listen, my children, and you shall hear
Of the midnight ride of Paul Revere
On the eighteenth of April, in Seventy-five

He said to his friend "If the British march
By land or sea from the town tonight
Hang a lantern aloft in the belfry arch
Of the North Church tower as a signal light—
One if by land, and two if by sea. . . .

Henry Wadsworth Longfellow

How many bits of information were conveyed? Why not "None if by land, and one if by sea?"

27.2 Experiments will depend largely on available equipment. Typical good experiments are the ajustment of damping on a potentiometric recorder, the measurement of frequency response and accuracy of a pen recorder, and the fidelity and noise of a data tape recorder, using square-wave recording.

SUBJECT INDEX