REVISED

NUFFIELD ADVANCED SCIENCE

PHYSICS

TEACHERS' GUIDE 2 UNITS H to L

Ce P 132

PHYSICS
TEACHERS' GUIDE 2
UNITS H to L

Revised Nu

General editor,
Revised Nuffield
Advanced Physics
John Harris

Consultant editor
E. J. Wenham

Editors of Units in this *Guide*
Stephen Borthwick
Peter Bullett
David Chaundy
John Harris
Wilf Mace
Jon Ogborn

The Nuffield–Chelsea Curriculum Trust is grateful to the authors and editors of the first edition:

Organizers: P. J. Black, Jon Ogborn;
Contributors: W. Bolton, R. W. Fairbrother, G. E. Foxcroft, Martin Harrap, John Harris, A. L. Mansell, A. W. Trotter.

PHYSICS TEACHERS' GUIDE 2

UNITS H to L

REVISED NUFFIELD ADVANCED SCIENCE

Published for the Nuffield–Chelsea Curriculum Trust
by Longman Group Limited

Longman Group Limited
Longman House, Burnt Mill, Harlow, Essex CM20 2JE, England
and Associated Companies throughout the World

First published 1971
Revised edition first published 1986

Design and art direction by Ivan Dodd
Illustrations by Oxford Illustrators Limited

Filmset in Times Roman and Univers
Made and printed in Great Britain
by William Clowes Limited, Beccles and London

ISBN 0 582 35418 8

CONTENTS

FOREWORD

When the Nuffield Advanced Science series first appeared on the market in 1970, they were rapidly accepted as a notable contribution to the choices for the sixth-form science curriculum. These courses were devised by experienced teachers working in consultation with the universities and examination boards, and subjected to extensive trials in schools before publication and they introduced a new element of intellectual excitement into the work of A-level students. Though the period since publication has seen many debates on the sixth-form curriculum, it is now clear that the Advanced Level framework of education will be with us for some years in its established form. That period saw various proposals for change in structure which were not accepted but the debate to which we contributed encouraged us to start looking at the scope and aims of our A-level courses and at the ways they were being used in schools. Much of value was learned during those investigations and has been extremely useful in the planning of the present revision.

The revision of the physics course under the general editorship of John Harris has been conducted with the help of a committee under the chairmanship of K. F. Smith, Professor of Physics, University of Sussex. We are grateful to him and to the committee, whose other members were W. F. Archenhold, J. Bausor, Professor P. J. Black, Professor R. Chambers, A. E. De Barr, Roger Hackett, John Harris, Wilf Mace, Robert Northage, Professor Jon Ogborn, A. J. Parker, and Maurice Tebbutt. We also owe a considerable debt to the Oxford and Cambridge Schools Examinations Board which for many years has been responsible for the special Nuffield examinations in physics and to the Assistant Secretary of the Board, Mrs B. G. Fraser, who has been an invaluable adviser.

The Nuffield–Chelsea Curriculum Trust is also grateful for the advice and recommendations received from its Advisory Committee, a body containing representatives from the teaching profession, the Association for Science Education, Her Majesty's Inspectorate, universities, and local authority advisers; the committee is under the chairmanship of Professor P. J. Black, educational consultant to the Trust.

Our appreciation also goes to the editors and authors of the first edition of Nuffield Advanced Physics, who worked with Jon Ogborn and P. J. Black, the project organizers. Their team of editors and writers included: W. Bolton, R. W. Fairbrother, G. E. Foxcroft, Martin Harrap, John Harris, A. L. Mansell, and A. W. Trotter. Much of their original work has been preserved in the new edition.

I particularly wish to record our gratitude to the General Editor of the revision, John Harris, Lecturer at the Centre for Science and Mathematics Education, Chelsea College, and a member of the team responsible for the first edition. To him, to E. J. Wenham, Consultant Editor of the revision, and to the editors of the Units in the revised course – all teachers with a wide experience of the needs of students and of the current state of physics education – Roger Hackett, Nigel Wallis, David Grace, Mark Ellse, Charles Milward, Trevor Sandford, Paul Jordan, Peter Harvey, Maurice Tebbutt, David Chaundy, Wilf Mace, Stephen Borthwick, Peter Bullett, and Jon Ogborn, we offer our most sincere thanks.

I would also like to acknowledge the work of William Anderson, publications manager to the Trust, his colleagues, and our publishers, the Longman Group, for their assistance in the publication of these books. The editorial and publishing skills they contribute are essential to effective curriculum development.

K. W. Keohane,
Chairman, Nuffield–Chelsea Curriculum Trust

INTRODUCTION

This is the *Teachers' guide* for the second year of the Revised Nuffield Advanced Physics course. Like *Students' guide 2* it covers Units H to L, the five Units of the second year of the course.

A full discussion of the nature of the revised course and its relationship to the original course, the content and aims of the course, suggested teaching methods and so on, is given in the Introduction to *Teachers' guide 1*.

ACKNOWLEDGEMENTS

One of the pleasantest aspects of the development of *Revised Nuffield Advanced Physics* has been the willing way in which so many people have contributed and become involved in the work. Above all, teachers have helped in many ways, and the very number who have done so makes it impossible to acknowledge the contribution of each individual. Many have offered suggestions at meetings or have written in with ideas for questions, demonstrations, and so on. We have tried to consider carefully all the suggestions put forward and, inevitably, it is impossible to give proper credit to the source or origin of every idea we have used. One who has made a particularly valuable contribution in this way is Colin Price. To him and the many others whose contributions go unacknowledged, we offer our sincere thanks.

Other teachers have helped by conducting trials of some of the more radically changed parts of the course, and of a major innovation – the 'Dynamic modelling system'. The trial schools are: Aylesbury Grammar School; Beechen Cliff School, Bath; Bexley–Erith Technical High School, Bexley; Bishop Hedley High School, Merthyr Tydfil; Cheltenham College; Esher College; Forest Hill School, London; Godolphin and Latymer School, London; The Grammar School, Batley; The Greenhill School, Tenby; Haverstock School, London; Heathland School, Hounslow; Henbury School, Bristol; Highfield School, Wolverhampton; Howell's School, Llandaff; King Edward VI College, Nuneaton; Kingsbridge School; Lady Margaret High School, Cardiff; Malvern College; Marlborough College; Netherhall School, Cambridge; North London Collegiate School; Northgate High School, Ipswich; Oulder Hill Community School, Rochdale; Richmond-upon-Thames College; Royal Grammar School, High Wycombe; Rugby School; Samuel Ward Upper School, Haverhill; and Sutton Manor High School.

We are grateful to the Inner London Education Authority for trying some of our material on electronics in their 1983 Summer School for sixth-form students at the North London Science Centre.

Mark Ellse has read and commented on much of the draft material, and has made particularly useful suggestions about the up-dating of some experiments and pieces of equipment.

Thanks are due to a group of teachers, convened by Bob Fairbrother, who met several times to discuss assessment. Their suggestions led to some changes in the structure of the examination.

Others, as well as teachers, have helped, of course. While he was working as a technician at the Centre for Science and Mathematics

Education, Chelsea College, Phil Webb found time in a busy schedule to try out ideas for demonstrations and experiments, and to suggest ideas for new apparatus.

CLEAPSE School Science Service reviewed all the suggested experiments and demonstrations and made useful suggestions on the safety aspects of some of them.

Industry has helped too, and, among others, we are indebted to Rank Xerox, Amersham International P.L.C., and the CEGB for technical help and information.

Examination questions in the *Students' guide* are reprinted by permission of the Oxford and Cambridge Schools Examinations Board. All are taken from Oxford and Cambridge Nuffield A-level Physics papers. Where guide lines for answers to examination questions are provided it must be understood that these are not the Examination Board's responsibility.

The Consultative Committee have, I believe, been asked to work harder and contribute more than is usually expected of such a group. As well as attending many meetings they have read and commented in detail on draft manuscripts – sometimes in a far from ideal state – and they have done all this most willingly.

It is a pleasure to acknowledge E. J. Wenham's help and sound advice. Much of what is written in these books has benefited from his knowledge and experience as teacher and author.

All of us who have contributed to these books owe a great debt of gratitude to Nina Konrad and her colleagues in the Publications office of the Nuffield–Chelsea Curriculum Trust for their thorough and painstaking work in preparing our manuscripts for the printers and our sometimes quite inadequate drawings for the artists.

Finally, I would like to express my sincere thanks to Paul Black and Jon Ogborn. Their help and support has been invaluable. During a period when both have been particularly busy, they have still found time to give advice both on general matters and on points of detail. They were, of course, the chief architects of the original Nuffield Advanced Physics course. Their willingness to be involved with what must at times have seemed like a severe distortion of their original plans, says much about their generosity of spirit.

John Harris

Unit H
MAGNETIC FIELDS AND A.C.

David Chaundy
Malvern College

Suggested time allocation: five weeks

PLAN OF THE UNIT

Section H1
Magnetic fields

ideas of field from GCSE and from Unit E, 'Field and potential' ► **Forces on currents**

$F = BIl$
$F = BQv$
Hall effect

circular motion: $F = mv^2/r$
energy: $\frac{1}{2}mv^2 = eV$ ► **Measurement of e/m; accelerators and mass spectrometers**

Fields near currents: straight wire and solenoid ► Unit J, 'Electromagnetic waves'

Section H2
Electromagnetic induction

e.m.f. in a moving wire

Kirchhoff's Law (Unit B) ► **d.c. generators and motors**

Faraday's Law

	Effect of iron	
electric circuits; resistance (Unit B) ▶	**Reluctance**	
	Mutual inductance	▶ Section H3: transformers
	Self inductance	▶ Section H3: electrical oscillations
	Section H3 Alternating current	
	Transformers	
	Eddy currents	
	r.m.s. value	
capacitors (Unit B) ▶	**a.c. in a capacitor**	
	a.c. in an inductor	
Unit D, 'Oscillations and waves' ▶	**Electrical oscillations**	

INTRODUCTION

More perhaps than most others, this Unit is a compromise between the many good things which can be got out of the study of its subject matter. Electromagnetism can be used to illustrate a wide variety of the features of physics and of practical engineering. It can be used to show the power of the tools of mathematical analysis at work in calculating fields. It can be used to show how concepts can be built up into complex logical structures and to illustrate the interplay of definition and fact. It can be used simply to develop an understanding of principles, like the principle of electromagnetic induction, which are needed elsewhere. It can be used as a way of showing the importance and use of the field concept, and, in particular, of the way in which vector fields are handled. It can also be used to illustrate how a host of practical machines may be devised, based on a few principles only.

What is certain is that not all these things can be attempted at once. We have chosen to lay emphasis on those principles of electromagnetism which are the bases of so many applications involving the creative ingenuity and flexibility of mind that characterize the engineer.

The work handles some of the ideas in a way that engineers will recognize as akin to their own methods, especially on those occasions when they want answers quickly and choose not to worry too much about deeper theoretical matters. The flux from currents in wires is treated as if it were like the flow of water or of electric current, the amount of flow being decided by the 'push' it is given (ampere-turns) and the 'resistance' to its flow (reluctance). We suggest this approach because, in the context of the work being developed, it is the *easiest* way, and also because it is the *practical* way, relating closely to such things as iron-cored transformers. Reluctance is a concept of more use to an engineer than to a physicist. We make no apology for abandoning the physicist's point of view in some parts of the course; it is well enough represented elsewhere, and to look at things in only that one way would be to ignore the variety of interests among students.

Electromagnetism is complicated. Were it not so, we would have preferred to pass as rapidly as possible to its uses. Because electromagnetism *is* complicated, Section H1 has been given over to a simple preliminary approach to the magnetic field via the force on a current. This has the advantages that a B-field is actually measured in the first lesson or so, and that the discussion can rapidly move to a set of applications of magnetic fields – accelerators and mass spectrometers – which are simple and useful (though they do not illustrate the engineering virtues that the later uses reveal).

Because to want everything, and to want it now, is as fruitless in teaching as it is elsewhere, other aims, usually embodied in the teaching of this subject matter, have been sacrificed to a greater or lesser extent. The integration of an equation for the field of a current element, to give the fields of coils or wires, is a fine opportunity to show the power of mathematics. We have not taken that opportunity. Instead, students make measurements of fields near currents and compare their results with the theoretical relationships. This is no less valuable an exercise than using the tools of analysis, and has the additional virtue of emphasizing that the magnetic field is a measurable thing.

Much of the Unit, however, follows well trodden paths. Section H2 contains a careful development of electromagnetic induction, which, while being treated with as practical a slant as possible and emphasizing ideas which will be of value in Section H3, deals mainly (as it must) with rates of change, turns, flux, and the idea of linking. Here the engineer's and the physicist's interests largely coincide.

The work on electromagnetic induction is presented as a series of demonstrations, because the interlocking of the relevant factors is complex enough for it to be hard to see how to expect results from students left entirely to themselves. But the demonstrations are, we suggest, best done by *students*, not by the teacher. This will not necessarily improve the quality of demonstrations, but it may involve each student in the work, in a way that sitting back and watching the teacher working hard is less likely to do.

Section H3 begins with students' investigations of transformer action, followed by a consideration of power losses in a.c. circuits. Both are matters of direct practical importance in the transmission of electric power. The Unit ends with a brief look at oscillations in reactive circuits, a topic which provides both an opportunity to bring out the formal and mathematical similarity between electrical and mechanical oscillations, and also to point to the usefulness of such circuits in a variety of applications.

Of the two vector fields, B and H, only B is employed in this Unit.

Of course, other lines of attack are possible. The design of this Unit, perhaps more than most, should be regarded merely as one attempt to put the ideas together in a sensible sequence.

Some of the problems of teaching electromagnetism are discussed in Appendix I on page 420.

THE PLACE OF THE UNIT IN THE COURSE

Some of the ideas presented in this Unit are difficult – magnetic fields are more complicated than electric or gravitational fields; and electromagnetic induction is not a subject which students find easy to understand. For these reasons this Unit is in the second year of the course; by now students already have some familiarity with the simpler fields and will have had considerable experience of d.c. electric circuits, including the behaviour of capacitors. Ideas about mechanical oscillations from Unit D, 'Oscillations and waves', will be useful in the discussion of electrical oscillations at the end of this Unit.

Ideas from this Unit may find practical application in Unit I, 'Linear electronics, feedback and control', and they are extended theoretically in Unit J, 'Electromagnetic waves'.

LIST OF SUGGESTED EXPERIMENTS AND DEMONSTRATIONS

SECTION H1
MAGNETIC FIELDS

The magnetic effects of currents are quite astonishing and are enormously useful. That there are magnetic forces between charges, but only when those charges *move*, is a strange phenomenon, especially if one recalls that to be moving or not is a relative matter. Faraday discovered that a current in one wire can cause a current to flow in a nearby wire, but only when the first current is *changing*. He also persuaded people to take the idea of a field as a physical 'something' quite seriously. Accepting this idea, Maxwell was able to predict the existence of radio waves. Einstein was puzzled by Maxwell's theory and by the dependence of the electromagnetic effects on *motion*; the outcome was the theory of relativity.

Evidently, the theoretical significance of these phenomena is very great. But so is their usefulness; our civilization depends to a very large degree on generators, motors, and transformers as well as on the use of electromagnetic waves.

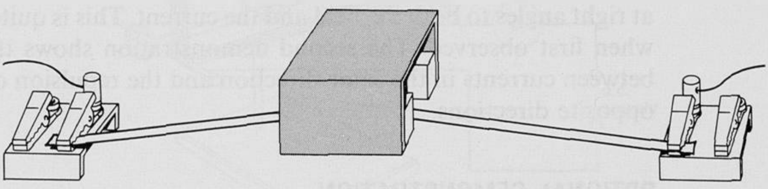
This Unit will not go deeply into the theory of electromagnetic fields as was done earlier in the course with electric and gravitational fields. Instead, differences between those fields and magnetic fields will be considered as will the engineering problems associated with the construction of efficient generators, motors, and transformers. This will clearly demand a consideration of electromagnetic induction and this will lead to a study of inductance and to work on a.c., showing why it is used so extensively and what it can do that d.c. cannot do.

We suggest starting with some demonstrations to show something of the strange nature of electromagnetism. Individual teachers may prefer other demonstrations to those described here; the key thing is that discussion of these first demonstrations will reveal the amount of revision which may be necessary. Students should have a picture of the fields between two unlike poles, near to a current in a straight wire, inside a coil, and inside a solenoid. If they are unsure of the effect of combining the uniform field between unlike poles and the circular field around a straight conductor, the catapult field should certainly be demonstrated.

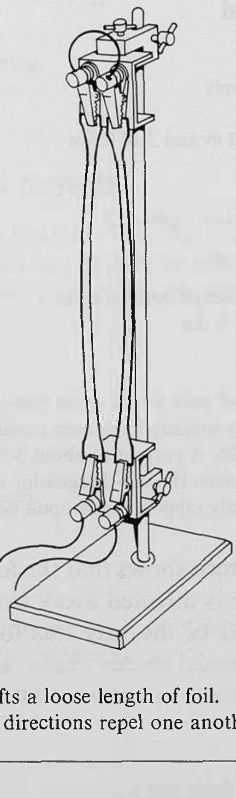
DEMONSTRATION

H1 Forces on currents; forces between currents

ITEM NO.	ITEM
59	l.t. variable voltage supply
92B	2 Magnadur magnets
92I	mild steel yoke
1153	aluminium cooking foil
1040	2 clip component holders
503,4	retort stand base and rod
505	2 bosses
1508	demonstration meter, 10 A d.c.
529	scissors
1000	leads



(a)



(b)

Figure H1
(a) Force on a current lifts a loose length of foil.
(b) Currents in opposite directions repel one another.

H1a Forces on currents

Cut a length of foil about 10 mm wide and 1 m long, securing the ends to two clip component holders. Pass a current of about 2 A. Hold the magnet assembly over the centre of the foil strip to make it rise – figure H1(a).

H1b Forces between currents

Cut two lengths of foil about 10 mm wide and 500 mm long and suspend them side by side about 10 mm apart – figure H1(b). It is important to get the tension in the foils just right and it may help to concertina the ends a little. The current may need to be as high as 8 A. Try the two conductors *i* in series and *ii* in parallel.

The first of these demonstrations shows that the force on the current is at right angles to both the field and the current. This is quite unexpected when first observed. The second demonstration shows the attraction between currents in the same direction and the repulsion of currents in opposite directions.

OPTIONAL DEMONSTRATION

H2 The catapult field

ITEM NO.	ITEM
59	l.t. variable voltage supply
503	retort stand base
504	retort stand rods: 1 of 1 m and 2 of 0.5 m
1153	4 nails, 15 cm long
505	7 bosses
	piece of card with hole (see figure H2)
92B	2 Magnadur magnets
92W	iron filings
1501	reel of 0.45 mm PVC-covered copper wire
1508	demonstration meter, 5 A d.c.
1000	leads

Wind a large, square coil with ten or more turns of the insulated wire. Such a coil, about 0.2 m square, may already have been constructed for REVISED NUFFIELD PHYSICS *Year 3*, demonstration 96. A current of about 3 A will provide a field which is comparable to that between the two Magnadur magnets. When the card is sprinkled with iron filings and gently tapped, the catapult field should show clearly. (See figure H2.)

The iron filings pattern shows that the force experienced by the current-carrying conductor is directed away from the strongest field. Students should be reminded of the way this force is used in the moving-coil ammeter and the model motor made in Year 3 of the Nuffield O-level Physics course (see also REVISED NUFFIELD PHYSICS *Pupils' Text Year 5*, page 106). If students have not previously made these two items and if time permits, they should be given the chance to make them now.

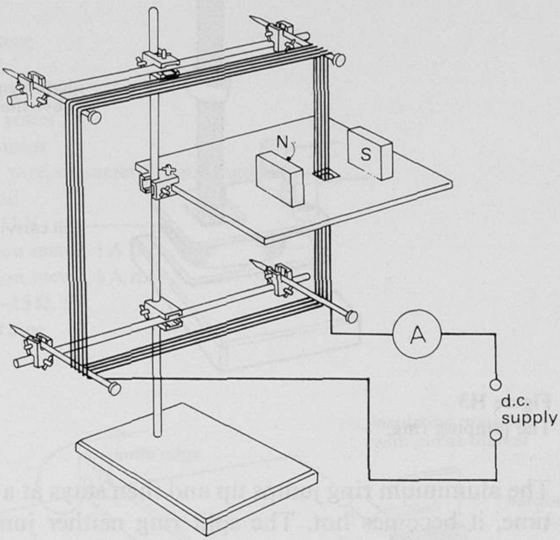


Figure H2
Catapult field demonstration.

DEMONSTRATION
H3 Forces on induced currents

This demonstration not only amuses students but also contains much of the physics of this Unit. It may be offered as a puzzle which they can attempt to solve now, but which should be clearly understood by the end of the Unit.

ITEM NO.	ITEM
59	l.t. variable voltage supply
1058	coil with 120 + 120 turns
503	retort stand base
504	retort stand rod (mild steel)
92P	aluminium ring
92Q	split aluminium ring
1000	leads

Adjust the supply voltage to provide a safe maximum alternating current of about 5 A in the coil and then switch on briefly.

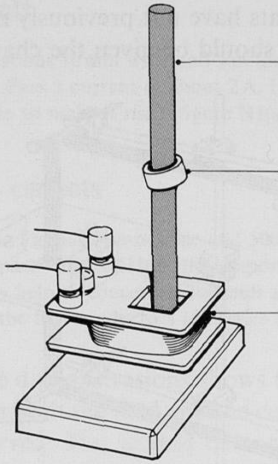


Figure H3
The jumping ring.

The aluminium ring jumps up and then stays at a certain height. After a time, it becomes hot. The split ring neither jumps nor becomes hot. Students may suggest that currents are flowing in the ring and they may see that it is behaving like a single-turn secondary of a transformer; but a full explanation must wait until the end of Section H2.

Applications of electromagnetism include the cycle dynamo, motor car ignition system, the gramophone motor, and the magnetic deflection system of a television picture tube. If posters or wall charts of large generators and motors, electric engines, or particle accelerators are available, these could be used to stimulate interest in some of the applications of the basic physics covered in the Unit.

MEASURING MAGNETIC FIELDS

An electric field is measured by finding the force experienced by unit charge placed in the field. A gravitational field is measured by finding the force experienced by unit mass placed in the field. In both cases the force and the field are in the same direction.

A magnetic field can be measured by finding the force experienced by unit length of a conductor which is carrying unit current. In this case the force and the field are not in the same direction.

DEMONSTRATION

H4a The directions of forces in magnetic fields

ITEM NO.	ITEM
1136	current balance
1079	flat solenoid
92B	4 Magnadur magnets
92I	2 mild steel yokes
92D	plotting compass
1501	bare copper wire, diameter 2 mm, length 1 m
1030	1100-turn coil
176	2 batteries, 12 V
1508	demonstration meter, 1 A d.c.
1508	demonstration meter, 5 A d.c.
541/1	rheostat, 10–15 Ω, 5 A
108/3	roll of tickertape
529	scissors
1000	leads

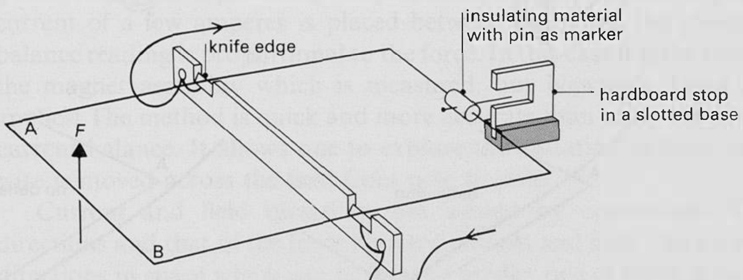


Figure H4
The wide current balance.

A number of current balances is now available and suitable for some, if not all, of the experiments suggested below. All detect vertical forces on an arm through which a known current flows. The current balance (figure H4) rests either on a pair of razor blades or on brass strips through which currents of up to 3 A may be supplied. Paper or wire riders are used to level the balance before use and to measure the force when a current flows. If pivots made from razor blades are used, it is advisable to push the wire frame firmly along the blades to cut a small groove in the wire and to blunt the blades a little. Such a balance is more stable but less sensitive if it is bent a little about the pivot.

More convenient for demonstration is a current balance which works in conjunction with a top-loading balance – figure H5(h). Such a current balance may be designed so that the force in newtons is shown directly by the change in reading of the top-loading balance.

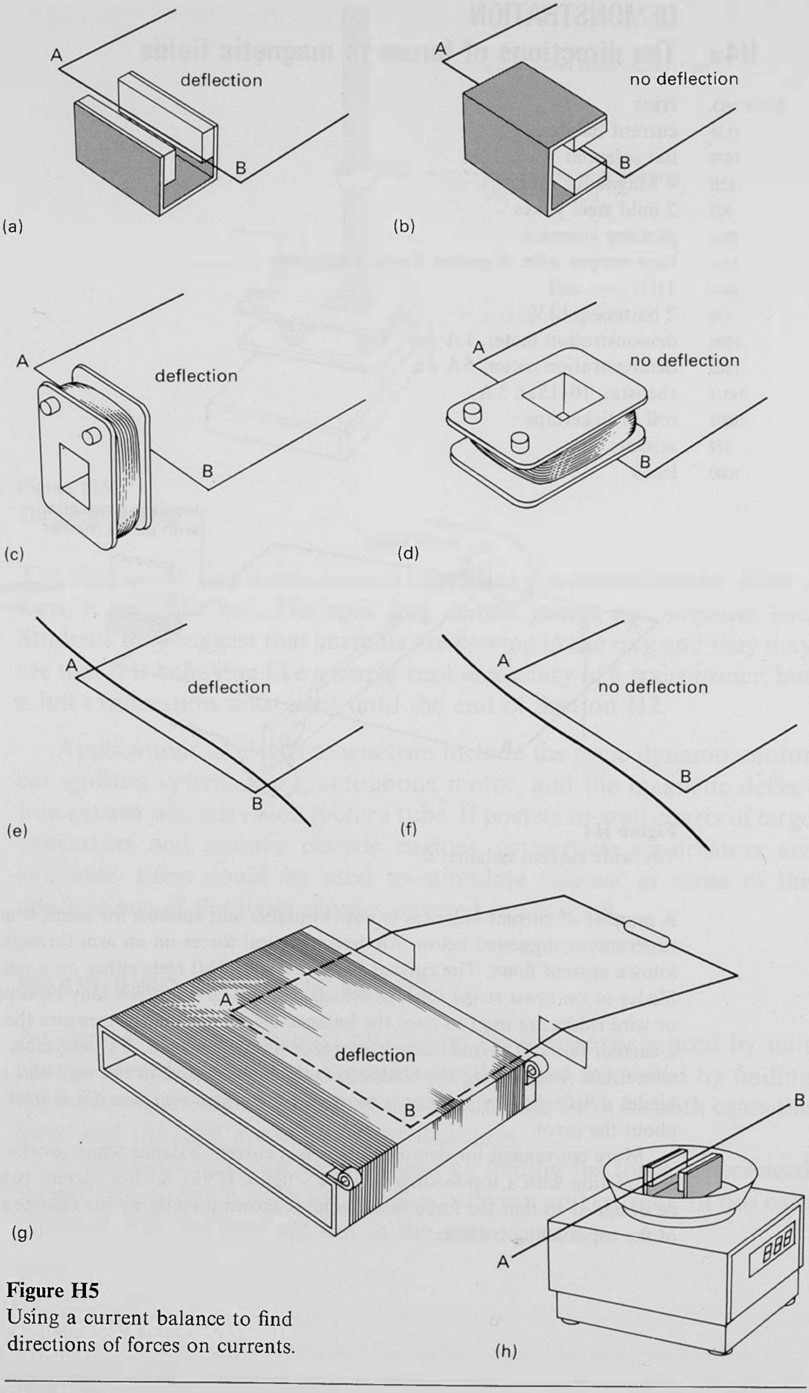


Figure H5
Using a current balance to find directions of forces on currents.

Figure H5 shows some configurations which might be examined. In all cases, AB is the current-carrying conductor of the balance.

In figures H5(a) and (b) the field direction between the two Magnadur magnets may be found with a plotting compass and it is easy to show that the force is in a direction which is at right angles both to the field and the current.

In (c) and (d) the 1100-turn coil needs between 1 and 2 A and the field decreases with distance from the coil.

In (e) and (f) the straight wire should be shorted momentarily across a secondary cell. The arrangement should not be used to get a steady deflection.

In (g), the flat solenoid requires at least 2 A. The field inside it is fairly uniform and will be used again in calibrating a Hall probe.

In (h), a sensitive top-loading balance is used to give a direct reading of the force between a current and a field. A pair of magnets mounted on a mild steel yoke is placed on the pan of the balance. A wire carrying a current of a few amperes is placed between the poles; the change in balance reading is proportional to the force. In this case it is the force on the magnet assembly which is measured, but Newton's Third Law applies! The method is quick and more accurate than using the simpler current balance. It allows one to explore the variation in force as the wire is moved across the field from pole to pole.

Current and field directions are agreed by convention. These directions and that of the force between current and field make a set of directions in space which can be used to predict one of them, given the other two. Three adjacent edges of a box may be used to illustrate this as well as the familiar left hand or motor rule (in the UK and elsewhere where motorcars drive on the left). Nevertheless, remembering the actual directions is not essential; in practical cases 'try it and see' is the usual rule!

Figure H6(a) (page 16) illustrates the familiar left hand rule. An alternative which some may find simpler is shown in figure H6(b). The parallel fingers of the *right* hand point in the direction of the field, the thumb shows the direction of the current (or movement of positive charge). The force is in the direction in which the palm would push. (Most people push with their *right* hands.)

But, as stated above, the important thing to remember is that F is perpendicular to both B and I, not whether it is up or down.

The directions remind one that 'the seashore crab is a politician. Threatened with danger from above, he looks straight ahead and runs away sideways'.

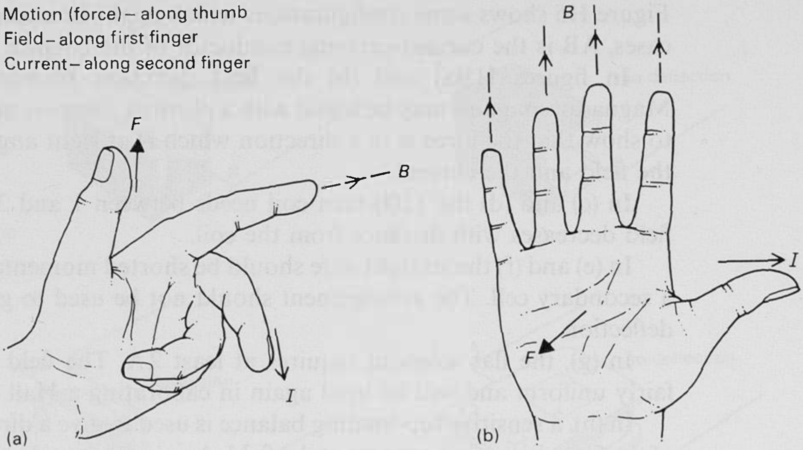


Figure H6
Force direction rules.

DEMONSTRATION
H4b Measuring a magnetic field

The arrangement shown in figure H4 (page 13) may be used to measure a magnetic field. The current in the wire AB is measured by a 5 A d.c. ammeter. This current should be in the direction to raise the wire between the magnets oriented as in figure H5(a). The magnetic force is then counterbalanced by the weight of a small rider of paper tape (it is worth weighing a few metres of paper tape and keeping it for making riders of known masses).

With the current off, adjust the balance so that the marker pin is in the middle of the stop; a small paper rider may be useful in doing this. Then put on the wire one of the paper riders which is known to balance the force due to a current of about 1 A; the arm AB falls. Switch on and increase the current until balance is restored. Note the current. Place a second equal paper rider on the wire and increase the current until balance is again achieved. If these currents are not too high, a third paper rider may be added.

Next add a second magnet assembly to extend the length of the field. Ideally one should first check that the two magnet assemblies are of equal strength. Keep the current constant and show that with two magnet assemblies two paper riders are needed to achieve balance, where the weight of one paper rider balances the force with only one pair of magnets. The second pair of magnets doubles the length of the field but does not change its strength. (If a wide current balance, *i.e.* about 30 cm, is available, two pairs of Magnadur magnets on separate yokes can be used. With a current balance that is only 10 cm wide it is necessary to place two magnets on each side of a single yoke, with the length of the magnets vertical rather than horizontal.)

Students may have met the equation $F = BIl$ before and may feel that these demonstrations that $F \propto I$ and $F \propto l$ are unnecessary. They may point out that the field between two Magnadur magnets is not very uniform and that it is therefore difficult to know what length to use for l.

A simple but rough method would be to assume that the field was uniform across the width of the magnet and zero outside it. Nevertheless, the point is that

$F \propto I \times l$

One can object that the ammeter which is used to measure the current is itself a form of current balance and that the argument is a circular one. This is discussed further in Appendix I, page 420.

Confining ourselves to cases in which B and I are perpendicular to one another, the proportionalities allow us to write

$F = BIl$

where B is a measure of the strength of the magnetic field.

It follows that $B = F/Il$ so that the units of B are $\mathrm{N\,A^{-1}\,m^{-1}}$.

The strength of the field in the gap between the pairs of Magnadur magnets may now be obtained; it is about $3 \times 10^{-2}\,\mathrm{N\,A^{-1}\,m^{-1}}$. The B-field inside a coil will usually be much less than this, while between the poles of a powerful magnet B might be as much as $1\,\mathrm{N\,A^{-1}\,m^{-1}}$.

The current balance can now be used to measure the field inside a flat solenoid with at least two different currents flowing in the solenoid and a suitable fixed current in the current balance (figure H7). This will confirm that $F \propto I$.

The strength of the field in the middle of the flat solenoid through which a current of 1 A is flowing should be noted. It will be required later for the calibration of field probes.

In this course, no one who calls B the magnetic field strength will run into any difficulty, but this is not its usual name. For reasons which will appear later, it is known as the *magnetic flux density* (and sometimes as the *magnetic induction*).

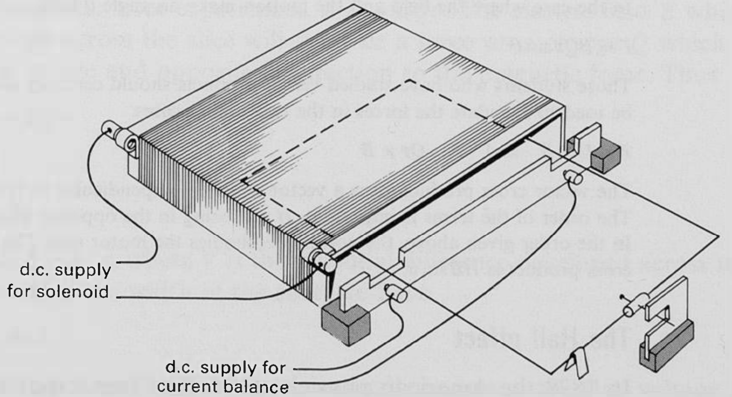


Figure H7
Measuring the field in a solenoid.

The unit $N\,A^{-1}\,m^{-1}$ reminds one that B is measured by the force on unit length of a conductor carrying unit current. This unit has the special name *tesla* (T). It is named after the engineer who pioneered the distribution of a.c. and who invented one form of induction motor. Tesla, who was born in Dalmatia (which is now part of Yugoslavia), spent much of his working life in the USA.

It will be noted that B has strange properties not shared by gravity or electricity. The force is there only when the charges involved are moving; it depends on the number of charges and on how fast they move (*i.e.* the size of the current). This force is at right angles to the field direction, not along it. Why has the magnetic force these strange properties? Any new law may raise such a question, though the oddity of the rules for magnetic forces suggests it very forcibly. It happens that these facts can be explained and linked with the electrostatic field by relativity – but this is not part of the course.

Questions

Questions 1 to 5 are about forces on currents.

THE FORCE ON A MOVING CHARGE

Currents consist of moving charges, and so the force on a current in a wire is the sum of the forces on the individual moving charges. Question 7 in the *Students' guide* derives the expression $F = BQv$ from $F = BIl$. (Question 6 is an introductory question about circular motion.)

Vector notation

In the case where the field and the motion make an angle θ with each other,

$F = BQv \sin\theta$

Those students who have studied vector products should certainly see how they can be used to calculate the forces in the two similar cases:

$\mathbf{F} = I\mathbf{l} \times \mathbf{B}$ and $\mathbf{F} = Q\mathbf{v} \times \mathbf{B}$

The vector cross product gives a vector which is perpendicular to both $I\mathbf{l}$ and to $\mathbf{B}$. The order of the terms is important, $I\mathbf{l} \times \mathbf{B}$ being in the opposite direction to $\mathbf{B} \times I\mathbf{l}$. In the order given above, the notation embodies the motor rule. The magnitude of the cross product is $IlB \sin\theta$.

The Hall effect

In 1879 the American physicist E. H. Hall found that when a current was flowing in a conductor, a magnetic field at right angles to the current caused a very small p.d. across the conductor. If semiconductors are used instead of ordinary conductors, there is a much larger p.d. The Hall effect will prove very useful in measuring magnetic fields later and is widely used in industry for that purpose. The effect is first demonstrated in small slices of semiconductor.

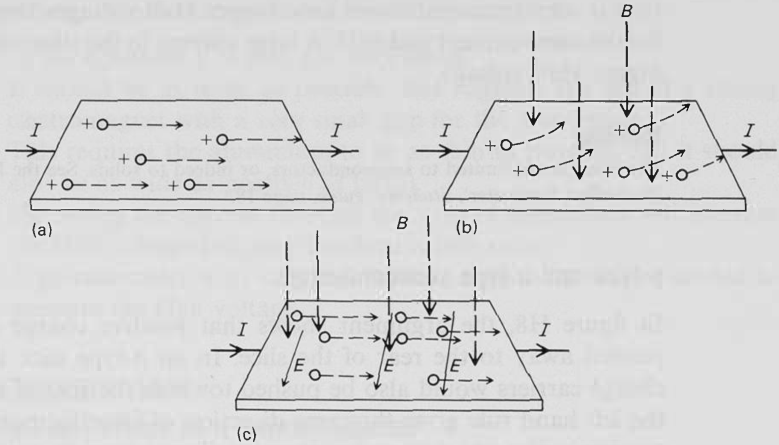


Figure H8
Forces on positive charges moving in a conductor with a transverse magnetic field.

Positive charges moving across a slice, as in figure H8(a), are pushed sideways by the magnetic field B, as shown in figure H8(b). In consequence, positive charges build up towards the rear of the slice, leaving the front edge with a negative charge, until the electric field which develops is large enough to prevent further sideways movement of charge – figure H8(c).

If the charge carriers each have charge Q and, on average, velocity v, the magnetic force experienced is $F = BQv$. The electric field E which develops across the slice will produce a force on a charge Q which is equal in size and opposite in direction to the magnetic force. Thus

$$EQ = BQv$$

and

$$E = Bv$$

Since $E = V/d$, where V is the potential difference developed across the slice and d the width of the slice, we have

$$V = Bvd$$

This potential difference, V, is usually known as the Hall voltage. It would be indicated by a sensitive galvanometer connected across the slice. Since this potential difference across the specimen is proportional to B, the effect may be used for measuring magnetic fields. The Hall voltage can be made large by having large carrier speeds. Work in Unit B, 'Currents, circuits, and charge', showed that large carrier speeds are linked with small numbers of charge carriers for the same current; that is why semiconductors have bigger Hall voltages than do metals for the same current and field. A large current in the slice will also give a bigger Hall voltage.

Reading

The effect is not limited to semiconductors, or indeed to solids. See the Reading 'Hall-effect flowmeter', *Students' guide*, page 18.

p-type and n-type semiconductors

In figure H8, the argument shows that positive charge carriers are pushed away to the rear of the slice. In an n-type slice the negative charge carriers would also be pushed towards the rear of the slice (for the left hand rule gives the same direction of force for positive charge carriers going one way as for negative charge carriers going the other way). It follows that the p.d. is reversed in sign. The motor rule can be used to decide which type of semiconductor is being used.

Students may wish to have a little information about the two sorts of doped semiconductor. Pure silicon has four valence electrons, which are normally closely bound in the crystal lattice. The addition of about one part in a million of an impurity such as arsenic with five valence electrons makes an n-type semiconductor. In this case there is one loosely bound electron for each arsenic atom and the conductivity increases about one million times. The addition of an impurity such as aluminium with only three valence electrons leaves a vacancy for an electron. This vacancy or 'hole' can be filled by a neighbouring electron; electrons moving in one direction to fill holes are equivalent to positive holes moving in the opposite direction.

It may be helpful to compare the two types of conduction to two types of bucket line for conveying water from a water tank to a fire. In the n-type bucket line, the person next to the water tank fills a bucket which is passed down the line. In the p-type bucket line, all the people in the line hold full buckets of water. The person fighting the fire takes a full bucket from the end of the line and in turn each person without a full bucket takes one from his neighbour. An observer will now see the 'lack of a bucket', or a hole, passing from the fire to the water tank.

The account offered above is greatly simplified. Although we do not expect students to have any more detailed knowledge of p-type and n-type semiconductors, those who would like further information could be referred to BENNET, *Electricity and modern physics*, or to MORGAN and HOWES, *Solid state electronic devices.*

Some students may wish to measure the Hall effect in a good conductor such as aluminium. The measurement raises some interesting design problems which can be summarized thus:

1 In the equation $V = Bvd$, v is very small.
2 B should be as large as possible; this suggests the use of a strong electromagnet with a very small gap for the aluminium.
3 This requires the aluminium to be as thin as possible; but it should also be as wide as possible (d large).
4 Increasing the current through the strip of aluminium will increase the Hall voltage but may overheat a thin strip.
5 A galvanometer with very high voltage sensitivity will be needed to measure the Hall voltage.

DEMONSTRATION
H5a The Hall effect in a semiconductor

ITEM NO.	ITEM
1012	semiconductor Hall effect demonstration
1507	milliammeter, 100 mA d.c.
1033	cell holder with four cells
1101	sensitive galvanometer
92B	2 Magnadur magnets
541/1	rheostat, 10–15 Ω
1000	leads

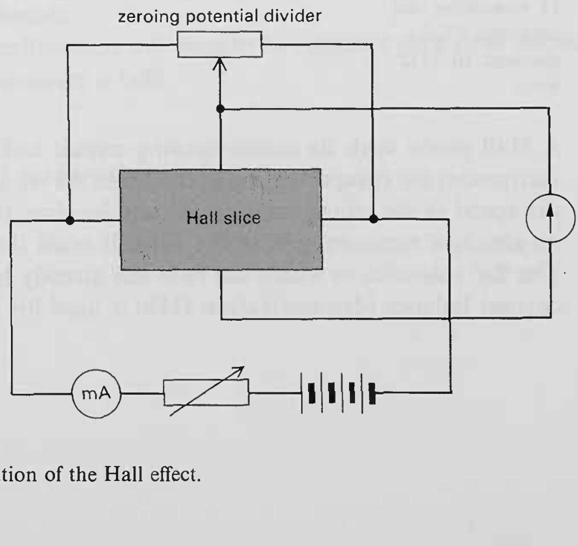


Figure H9
Demonstration of the Hall effect.

The Hall effect kit contains a slice of semiconductor material with connections at each end. Leads are also connected to the two sides of the slice. These are used in the measurement of any p.d. between the two sides. In practice it is not possible to get the side leads exactly opposite one another and so a potential divider is included to offset any initial p.d. between these side connections.

Connect the p-type slice as shown in figure H9 and adjust the rheostat to give a current which is less than the maximum permissible (usually about 50 mA). Adjust the potential divider to give zero current in the galvanometer. Bring a magnet near to the slice so that its field is perpendicular to the plane of the slice. A deflection will be observed on the galvanometer, suggesting that a potential difference has been developed between the two sides of the slice. This deflection increases as the field increases and also as the current through the slice increases. It changes direction if the field is reversed. Repeat the experiment with the n-type slice; with the same pole of the magnet approaching, the galvanometer deflection is in the opposite direction. If sufficient apparatus is available the demonstration may be more effective if the two slices are connected in series, each with a separate galvanometer.

DEMONSTRATION

H5b Calibrating a Hall probe

ITEM NO.	ITEM
1138	Hall probe and circuit box (with suitable meter)
1079	flat solenoid
59	l.t. variable voltage supply
1064	l.t. smoothing unit
1507	ammeter, 5 A d.c.
541/1	rheostat, 10–15 Ω
1000	leads

A Hall probe with its accompanying circuit box makes a convenient instrument for comparing magnetic fields. As we are unable to measure the speed of the charge carriers within the slice, the probe cannot give an absolute measurement of the field. It must therefore be calibrated. The flat solenoid, in which the field has already been measured with a current balance (demonstration H4b) is used for this.

The circuit box connected to the probe requires a battery to drive current through the Hall slice and to power an amplifier contained in the box. A meter connected to the box measures the amplified Hall voltage. Before carrying out the calibration, the potential divider on the circuit box must be adjusted to give zero output when there is no magnetic field operating (this implies that the probe should be held in the plane of the Earth's magnetic field; see figure H10).

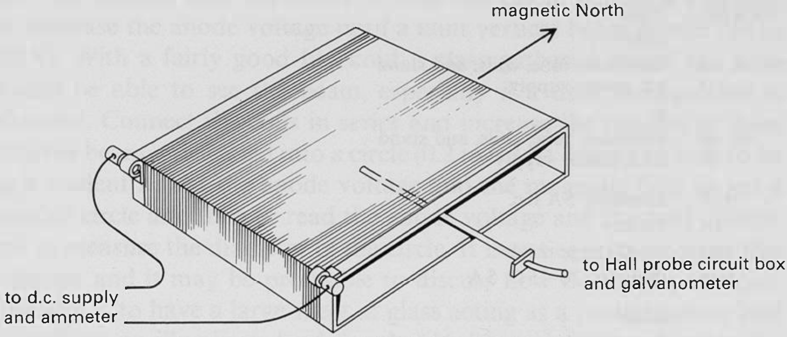


Figure H10
Calibrating a Hall probe.

A direct current of 1 A is passed through the solenoid. The magnetic field strength is known from demonstration H4b. It is in the order of 10^{-3} T.

The current balance experiment showed that the field in the solenoid was proportional to the current in it. It is worth checking to see whether the Hall voltage generated in the probe is also proportional to the current in the solenoid.

This calibration will need to be repeated each time the Hall probe is used to measure a field.

Questions

Questions 8 to 10 are about the Hall effect.

DEMONSTRATION

H6 Measuring the specific charge for electrons

This experiment is fully described in REVISED NUFFIELD PHYSICS *Teachers' guide Year 5*, pages 31–37.

ITEM NO.	ITEM
	either
61, 62 (or 235, 140, 139)	fine-beam tube, coils, and stand
15	h.t. power supply
	or
138–140	deflection tube, coils, and stand
14	e.h.t. power supply
1507	ammeter, 5 A d.c.
176	battery
1005	multi-range meter
541/1	rheostat, 10–15 Ω, 5 A
	either
1136	current balance to fit within the coils
	or
1138	Hall probe with circuit box and suitable meter
30	2 slotted bases
501	metre rule
1153	sheet of glass, about 20 cm × 30 cm
108/3	tickertape
529	scissors
1000	leads

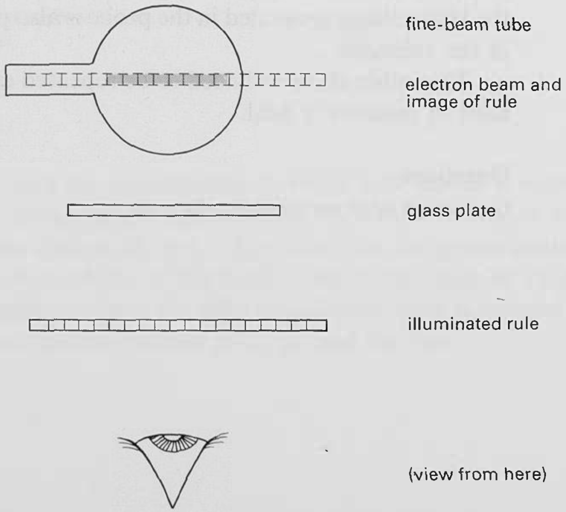


Figure H11
Plan view of arrangement for measuring diameter of electron beam.

Connect the fine-beam tube following the maker's instructions. H.t. supplies which can give up to 60 mA at 300 V can give fatal electric shocks. They must be used with care, and shrouded 4 mm plugs should be used.

Note that whereas some older tubes may require a heater voltage of 12 V, the Teltron tube requires 6 V. After the heater has been switched on, increase the anode voltage until a faint vertical beam is seen (80 to 120 V). With a fairly good blackout a class gathered round the tube should be able to see the beam, especially if a dark background is provided. Connect the coils in series and increase the current in them until the beam is deflected into a circle (0.2 to 0.3 A). It may be best to let each student adjust the anode voltage and the magnetic field to get a suitable circle and then to read the anode voltage and the field current and to measure the diameter of the circle. It is not easy to measure this diameter and it may be profitable to discuss how this might be done. One way is to have a large sheet of glass acting as a partial mirror and to position an illuminated rule so that its image is in the plane of the circle of electrons (figure H11). Even with this arrangement it is difficult to illuminate the rule effectively and it may be worth making a 'light rule'–a Perspex rod with a torch bulb at the bottom end. Small grooves are filed across the rod at centimetre intervals and light from the bulb is emitted from each groove. Alternatively, a transparent plastic ruler may be illuminated in the same way.

The magnetic field in the plane of the beam is measured with a Hall probe or with a current balance which will fit between the coils after the tube has been removed (see REVISED NUFFIELD PHYSICS *Teachers' guide Year 5*, pages 34–35).

The derivation given in question 11 of the *Students' guide* gives the specific charge for electrons as

$$e/m = 2V/B^2r^2$$

where e/m is the specific charge, V the accelerating potential difference, B the field, and r the radius of the circle.
For electrons,

$$e/m = 1.76 \times 10^{11}\,\mathrm{C\,kg^{-1}}$$

Since the charge on an electron is $e = 1.6 \times 10^{-19}$ C, the mass of an electron is $m = 9.1 \times 10^{-31}$ kg. The smallness of this mass has important consequences; in Unit L, 'Waves, particles, and atoms', it will be seen how it helps to determine the size of atoms.

Note that the uncertainties involved in this experiment are considerable. Values of B and r can be no more than approximations. These values are then squared and the product B^2r^2 found!

Mass spectrometers and accelerators: more reasons for making charges travel in circles

If the only use for charges going round in circles in a magnetic field were the measurement of the specific charge for the electron, it might not be worth learning about. But the idea has many other uses.

One of these is the measurement of Q/m and so of m for ions and fragments of molecules by adaptations of the method used for electrons. Machines for determining Q/m are called mass spectrometers; they are widely used in research and in industry for analysing new substances, monitoring processes such as steel-making, and for many other purposes. Students can get the feel of using a mass spectrometer and appreciate some of the problems encountered by using the computer simulation called 'Mass spectrometer' from the Computers in the Curriculum Project published by Longman Micro Software.

Another use is in nuclear research and the nuclear industry, where it is useful to be able to accelerate particles artificially so as not to be limited to the bombarding particles provided free by nature (as were Rutherford and his colleagues in the alpha-scattering experiments already described in Unit F, 'Radioactivity and the nuclear atom'). But this is not easy because the particles have to reach such very high speeds – up to 300 m per microsecond. They can be accelerated from rest to this kind of speed in a few microseconds but as, in this time, they have travelled several thousand metres the accelerating force has to be provided over very large distances. This presents difficulties. Rather than build long accelerating tubes across the landscape, it is possible to bend the path of the particles into circles or spirals using magnetic fields. Although the resulting accelerators are relatively compact, the greater energies required for high energy experiments today lead to larger and larger accelerators. The Super Proton Synchrotron at CERN has a circumference of 7 km, and the proposed Large Electron–Positron storage ring (not strictly an accelerator) will have a circumference of 27 km. These huge distances mean that linear accelerators (like the 3000 m installation at Stanford, U.S.A.) are no longer built for very high energy experimentation.

Reading tasks

This is one of the occasions in the course on which students can be given individual reading tasks and teachers who feel that their students might benefit from such an exercise could assign one at this point. (But such a task may not be well timed for students who have recently undertaken one of the 'Energy options' in Unit G, 'Energy sources'.)

Many books and articles give plenty of interesting information. Each student, or pair of students, could have one particular task and be asked to tell the rest about it, or to write a summary to be handed round for the rest to read. A brief report will suffice: the precise details of the working of mass spectrometers and accelerators of various kinds are not important – what matters are the principles they have in common.

References

BOLTON, *Patterns in physics.* Chapter 13.
CARO, MCDONNELL, and SPICER, *Modern physics.* 3rd edn. Chapter 13.
STAFFORD, *The use of high energy machines in particle physics.*
WENHAM, *et al., Physics: concepts and models.* 2nd edn. Chapter 48.
WILSON and LITTAUER, *Accelerators: machines of nuclear physics.*

Questions

Questions 12 to 16 are about the forces on charged particles.

MEASURING FIELDS NEAR TO CURRENTS

Students have seen iron filings patterns of magnetic fields in earlier courses and they should now know how to measure the field inside a solenoid. They also have, in the Hall probe, a useful tool for comparing magnetic fields. Now they can investigate in detail the fields around currents flowing in straight wires, in coils, and in solenoids.

It is possible to find rules analogous to (but more complicated than) the inverse-square law for the electric field of a point charge. These rules allow one to calculate the field of any coil by adding up the contributions of all the bits of the coil. Indeed, books on the subject are full of such calculations. But the calculations are only reasonably easy for the simplest arrangements – long straight wires, circular coils. Although such calculations play a part in designing, say, a new sort of motor, there is often no substitute for measuring the field, especially if iron is used. So we prefer measuring a field to calculating it.

The experiments which follow require the field of a current-carrying wire or coil to be measured rather carefully; varying the distance, or the number of turns, or the cross-sectional area, etc., and seeing how closely the measurements agree with the simple formulae for such simple configurations as long wires and solenoids.

Hall probe and search coil

It is convenient to measure fields with Hall probes as these give readings which are directly proportional to the field. Some may prefer to use small search coils but these give readings which are proportional to the rate of change of the field. In this case it would seem better to leave this work until the induction of alternating e.m.f. has been dealt with. However, that upsets the suggested logical development. In addition, it is less confusing to many students if the effect used depends on the field rather than on the rate of change of the field. Using a search coil to measure a field is not unlike using a speedometer to measure distances. Nevertheless, a search coil could be used as a 'black-box' device which can be shown to give a proportionately bigger reading in a larger field, without first dealing with electromagnetic induction. A fixed frequency must, of course, be used.

Teachers might like to try connecting a square-wave generator to a solenoid and to insert a Hall probe and a search coil side by side. These are connected to the two Y inputs of a double-beam oscilloscope. The display brings out the differences between the two detectors very neatly.

EXPERIMENT
H7 Fields near electric currents

ITEM NO.	ITEM
	Conductors
	either
101	large Slinky
30	2 slotted bases
501	2 wooden strips (*e.g.* rulers) to support Slinky
52K	2 crocodile clips
	or
1037	set of solenoids
	or
1042	magnetic field board
1501	reel of 0.45 mm PVC-covered wire
	or
1058	coil with 120 + 120 turns

ITEM NO.	ITEM
	Sources of current
	either
176	12 V battery
541/1	rheostat, 10–15 Ω, 5 A
1507	ammeter, 10 A d.c.
	or
27	transformer
541/1	rheostat, 10–15 Ω, 5 A
1507	ammeter, 10 A a.c.
	or
1109	signal generator
1507	ammeter, 1 A a.c.
1000	leads

ITEM NO.	ITEM
	Field-measuring devices
	either
1138	Hall probe and circuit box
1101	sensitive galvanometer
	or
1039/1	axial search coil
1039/2	lateral search coil
1511	oscilloscope

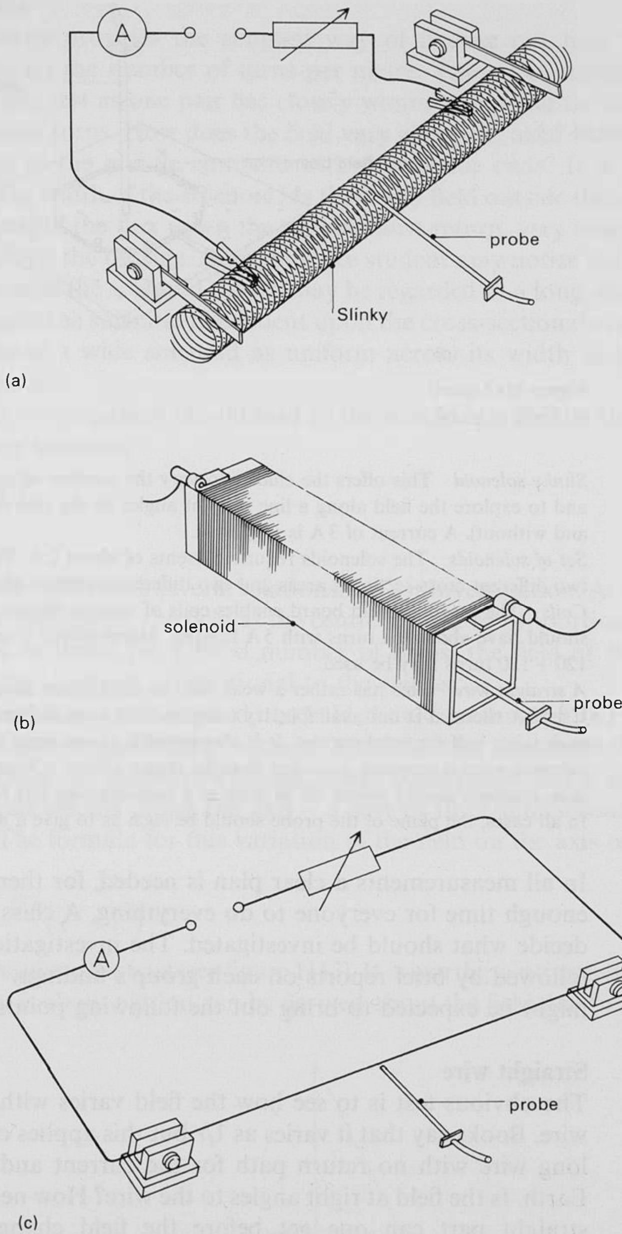
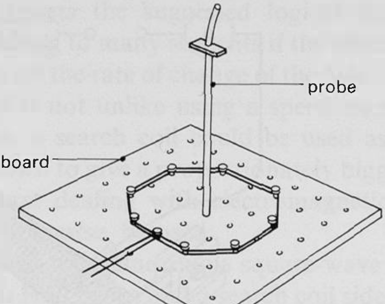


Figure H12 (part)
(a) Field in a Slinky solenoid.
(b) Field in a solenoid.
(c) Field near a long straight wire.



(d)

Figure H12 (part)
(d) Field of a coil.

Slinky solenoid This offers the chance to vary the number of turns per unit length and to explore the field along a line at right angles to the axis of the coil (both within and without). A current of 3 A is suggested.
Set of solenoids The solenoids require currents of about 2 A. The set of four includes two different cross-sectional areas and two different numbers of turns per unit length.
Coils The magnetic field board enables coils of various shapes to be made. Each should have about ten turns with 5 A flowing. Alternatively, a coil such as that with 120 + 120 turns may be used.
A straight wire This has rather a weak field so the current should be 10 A if possible. If a 10 A rheostat is not available, try using two 5 A ones in parallel, ensuring that each takes half the total current. It is also possible to use several parallel wires each carrying several amperes, provided that the return circuit is kept well away. In any case, students should switch off as soon as a measurement has been made.
In all cases, the plane of the probe should be such as to give a maximum reading.

In all measurements a clear plan is needed, for there will not be enough time for everyone to do everything. A class discussion could decide what should be investigated. The investigations should be followed by brief reports on each group's findings. These reports might be expected to bring out the following points.

Straight wire

The obvious test is to see how the field varies with distance from the wire. Books say that it varies as $1/r$ but this applies only to an infinitely long wire with no return path for the current and no field from the Earth. Is the field at right angles to the wire? How near to the end of the straight part can one get before the field changes appreciably in magnitude or direction? Is the field the same on the two opposite sides of the wire? If it is not, it is likely that the Earth's field is making a difference and it would be wise to change the orientation of the wire.

Solenoids

The Slinky provides the simplest way of finding out how the field depends on the number of turns per metre. The set of solenoids also allows this test as one pair has closely wound turns and the other pair has spaced turns. How does the field vary along the axis? How does its strength at the middle compare with that at the ends? Is it constant across the width of the solenoid? Is there any field outside the solenoid? (Presumably the flux down the middle must return, very much spread out, around the outside. A wide-awake student may notice that there is a field *round* the solenoid, which may be regarded as a long wire.) Is the field inside the solenoid dependent upon the cross-sectional area? Is the field inside a wide solenoid as uniform across its width as that in a narrow one?

This investigation should lead to the conclusion that, at the middle of a long solenoid

$$B \propto NI/l$$

Coils

A circular coil offers several relationships for investigation. At constant current and radius, the field at the centre should be proportional to the number of turns; for a fixed number of turns, the field at the centre should be inversely proportional to the radius.

Everywhere on the board, the field should go through the board at right angles to the surface – does it? The field near the centre should be a minimum for sideways displacements of the probe, but a maximum for displacements along the axis – a sort of three-dimensional saddle point. The formula for this variation of the field on the axis of the coil

$$B = \frac{\mu_0}{2} NI \frac{r^2}{(x^2 + r^2)^{3/2}}$$

could be given for trial (see figure H13). It is worth pointing out that this formula is about all that can be derived about the field due to a current

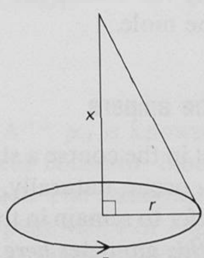


Figure H13

in a coil by simple integration. It is difficult to calculate the field at a point away from the axis, but easy to measure it.

Over what area, or volume, is the field constant to within ten per cent? This result may be linked to the use of a pair of circular coils (a Helmholtz pair) in demonstration H6.

The question of the field which may be due to the currents in the leads to the coil could be discussed. Do such leads contribute to the field if they are close together? How close is close? It can be shown that there is no field near two close parallel wires carrying equal currents in opposite directions – figure H14(a). If one of the wires is kinked after the fashion shown in figure H14(b) the net field remains zero. This is related to the fact that current elements can be resolved into components.

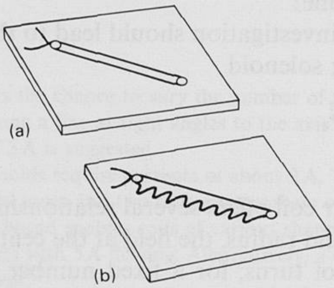


Figure H14

The field of a long wire

Starting from the empirical facts that

$B \propto I$ and $B \propto 1/r$

we have

$B = kI/r$

The constant, k, depends on the system of units used. In SI the ampere is the unit of current. Indeed, it is one of the seven base units of the system, the others being the metre, the kilogram, the second, the kelvin, the candela, and the mole.

Definition of the ampere

Up to this point in the course a student might be forgiven for supposing that ammeters occur naturally, already carrying reliable markings. Some might prefer to remain in this state. Of course it would be possible to decide that *this* ammeter *here* is the 'standard' one and to compare others with it. But ammeters include magnets which will vary in strength as time goes by, so the standard current could alter without anyone being the wiser. Something more reliable is needed.

The chosen standard is quite arbitrary, settled upon for the convenience of physicists who have to make accurate current measurements. In SI *the ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible cross-section, and placed 1 metre apart in a vacuum, would produce between these conductors a force equal to* 2×10^{-7} *newton per metre of length.*

Students need to *recognize* the definition but examination questions would not ask for it to be set down, though a question might well state all or part of the definition and ask for it to be used in the discussion of a problem. Similar remarks apply to formulae like those for the field in a solenoid or near a long wire.

This definition gives a value to the constant k in the formula for the field due to a current in a long wire. A second wire one metre away would experience a force of 2×10^{-7} N per metre length if the current in each wire were one ampere.

$$F = BIl = 2 \times 10^{-7}\,\mathrm{N}$$

But

$$B = kI/r$$

So

$$kI^2l/r = 2 \times 10^{-7}\,\mathrm{N}$$

and since $I = 1$ A and $l = r = 1$ metre,

$$k = 2 \times 10^{-7}\,\mathrm{N\,A^{-2}}$$

Since the field in a plane at right angles to the conductor is circular in form, the magnetic effect of the current at a distance r from the conductor is 'spread' around a circle of circumference $2\pi r$. It tidies things up a little if $2\pi r$ is used instead of r in the denominator and an additional 2π appears in the numerator. The formula for the field near a long straight wire now becomes

$$B = \frac{\mu_0 I}{2\pi r}$$

where μ_0 is $4\pi \times 10^{-7}\,\mathrm{N\,A^{-2}}$. μ_0 is known as the *permeability of free space* or the magnetic field constant. Since it is defined rather than measured, its value is exact. The relationship between μ_0, ε_0, and c is discussed in Appendix I, page 420. For experiments in air, μ_0 may be used since the relative permeability of air is very close to unity.

SI units lead to formulae containing π for the fields with circular (or spherical) symmetry, for example the circular field around a long straight wire. π is absent from formulae for uniform fields, for example inside a solenoid or a toroid.

If students wonder why a number such as 2×10^{-7} appears in the definition, it may be helpful to point out that the ampere has not always been defined in this way and that the value of the force in the present definition was chosen so that the newly defined ampere was as close as possible in size to those in earlier definitions.

Question

Question 17 deals with the definitions of the ampere and the constant μ_0.

In practice, no sensible person would set up two extremely long wires one metre apart and measure the force between them when 1 ampere was flowing, because 2×10^{-7} N is much too small a force to measure accurately in such a situation. Since the force between the two is proportional to $1/r$, the wires might be brought closer together; that would increase the force and students can be reminded of demonstration H4 – figure H5(e) – where a measurable force can be obtained with large currents. If the same current I were flowing in both wires it could be calculated from $F = BIl$ and $B = \mu_0 I/2\pi r$. These lead to the equation $I^2 = 2\pi r F/\mu_0 l$. The quantities F, r, and l need only metre rules, masses, and clocks for their measurement (the clock is used in measuring a force from the acceleration given to a known mass). The current I can be found from these non-electrical measurements. Such a measurement is said to be an *absolute* one as no ammeter is involved.

The quality of a measurement with the apparatus used in demonstration H4 would be poor, though the principle is worth showing. The deflection of the current balance could be measured by the use of a counterpoise but its very crudeness points to a need to do better.

The main need is for a stronger B-field, so a solenoid might serve. The flat solenoid would be better than one of the square ones because a longer current-balance wire can be put into it. Before attempting to make such a measurement, the formula for the field inside a flat solenoid must be derived from the formula for the field of a straight wire. This is treated below as an optional extra piece of work for those interested. These students would then be in a position to make an absolute, but not very accurate measurement of a current. However, the majority may prefer to examine the photograph of the National Physical Laboratory's current balance which is used periodically to determine the ampere. Each determination takes many months of careful measurement. This photograph appears on page 6 of the *Students' guide.*

Questions

Questions 18 to 22 are about magnetic fields near currents.

Optional extra The field of a solenoid

The field of a flat solenoid is the same as the field of a square solenoid, provided that it has the same current flowing and the same number of turns per metre. In figure H15 it can be seen that the adjacent sides of the four square solenoids can be ignored as they carry equal currents in opposite directions. We need only find the field due to two carpets of parallel wires carrying currents in opposite directions. The short, vertical edges of the solenoid contribute just what is needed to make up for the fact that the carpets of wires come to an end.

The field of such a carpet of wires (figure H16) can be calculated from the field of one wire, by adding up the effects of many wires. The field B at a point such as P in figure H16 will be calculated. The carpet has n wires per unit length, each carrying a current I.

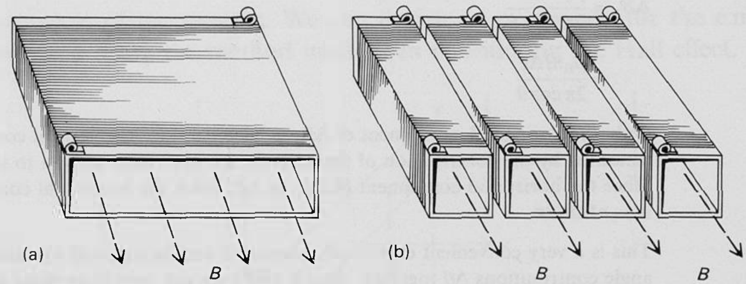


Figure H15
A flat solenoid and four comparable square ones.

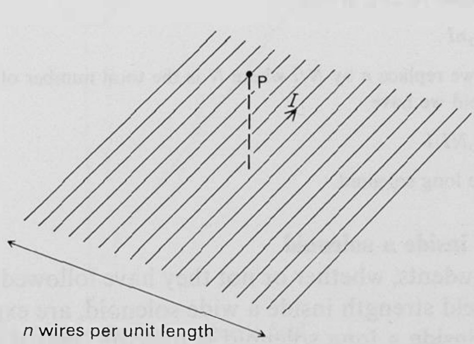


Figure H16
A single carpet of wires.

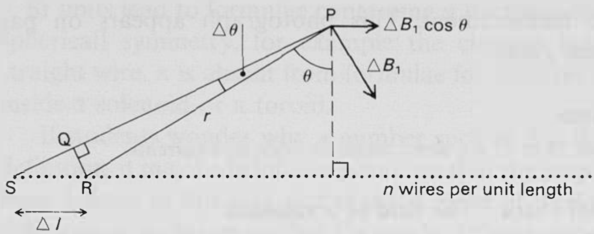


Figure H17

Figure H17 shows the arrangement, looking at the ends of the wires. If $\Delta\theta$ is small, the wires in the section RS of length Δl contribute a field ΔB_1 at P where

$$\Delta B_1 = \frac{\mu_0 n \Delta l I}{2\pi r}$$

Angle QRS approaches θ as $\Delta\theta$ approaches zero, so, in the limit, RS becomes QR/cos θ. Also, QR approaches $r\Delta\theta$ as $\Delta\theta$ approaches zero, so, by making $\Delta\theta$ sufficiently small, hardly any error is made in saying that $\Delta l = r\Delta\theta/\cos\theta$. Hence

$$\Delta B_1 = \frac{\mu_0 n I r \Delta\theta}{2\pi r \cos\theta}$$

$$= \frac{\mu_0 n I \Delta\theta}{2\pi \cos\theta}$$

Only the horizontal component of ΔB_1 is effective since the vertical component will be cancelled by the contribution of the wires in the equivalent section to the right of P. Since the horizontal component of ΔB_1 is $\Delta B_1 \cos\theta$, the horizontal contribution to B_1 is $\mu_0 n I \Delta\theta / 2\pi$.

This is a very convenient expression, because it can be summed by taking all the small angle contributions $\Delta\theta$ together, since n and I do not vary from place to place. If P is very close to the carpet, then the total contribution to $\Delta\theta$ of such a carpet is π, so $B_1 = \frac{1}{2}\mu_0 n I$. The other carpet above P will also contribute π, and so the total field at P is equal to $2B_1$, giving

$$B = \mu_0 n I$$

If we replace n by N/l where N is the total number of turns and l the length of the solenoid we have

$$B = \mu_0 N I / l$$

for the long solenoid.

Field inside a solenoid

All students, whether or not they have followed the above derivation of the field strength inside a wide solenoid, are expected to know that the field inside a long solenoid is uniform, that it depends on the current and the number of turns per unit length. While they are not expected to memorize the formula $B = \mu_0 N I / l$, they are expected to be able to use it in problems.

SECTION H2

ELECTROMAGNETIC INDUCTION

The concept of electromagnetic induction is one which students may find difficult to grasp. The treatment here has therefore been split into two parts. The first of these deals with the events associated with a wire cutting a magnetic field, and the second with the concepts of flux and of rate of change of flux linkage.

Some initial revision of earlier work may be needed. A galvanometer connected to a wire which is moving between the poles of two Magnadur magnets on a yoke will show a current, but the response of the galvanometer is too slow to show proportionality. Moving a bar magnet into a coil gives a bigger effect. If displayed on an oscilloscope, the effect can be seen to be roughly proportional to the rate of movement of the magnet. We can derive an expression for the e.m.f. generated using the method used when considering the Hall effect.

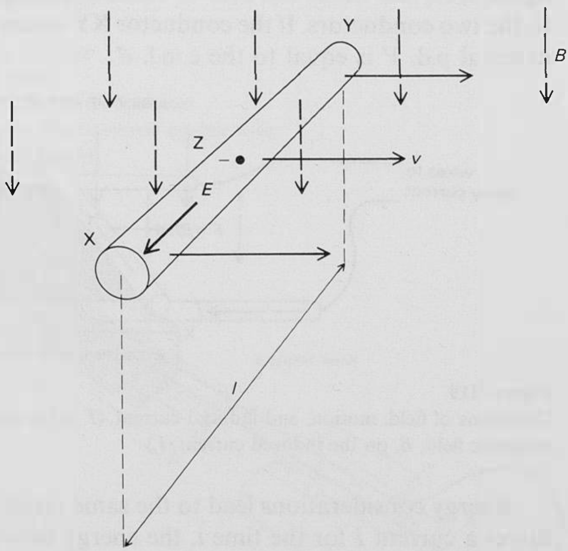


Figure H18
A conducting bar moving across a magnetic field.

A straight conductor in the form of a bar XY of length l moves across a magnetic field B with speed v (see figure H18). An electron in the bar at point Z will experience a force of magnitude Bev in a direction such that it will tend to move towards the end X. This redistribution of charge produces an electric field E within the conductor. The motion of electrons will cease when the electric force upon them is equal and opposite to the magnetic force. Then $Ee = Bev$ and so $E = Bv$.

The electric field intensity E is the potential difference V between X and Y per unit length of XY (as we have seen in Unit E, 'Field and potential')

$E = V/l$

Combining these two expressions for E we have

$V/l = Bv$

and

$V = Bvl$

The energy needed to separate the charges and to establish the electric field derives from the energy put into the system as the wire is moved against the magnetic force. It is usual to say that an e.m.f. $\mathscr{E}$ has been induced in the conductor XY.

If the conductor XY moves along a pair of conductors as in figure H19, this e.m.f. will drive a current through a resistor connected to the two conductors. If the conductor XY has negligible resistance, the external p.d. V is equal to the e.m.f. $\mathscr{E}$.

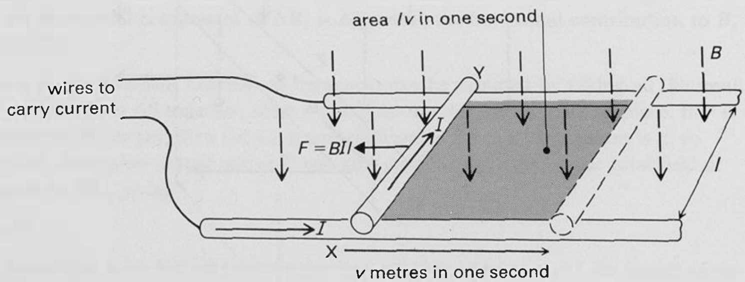


Figure H19
Directions of field, motion, and induced current. (F is the force exerted by the magnetic field, B, on the induced current, I.)

Energy considerations lead to the same result. If an induced e.m.f. $\mathscr{E}$ drives a current I for the time t, the energy delivered is $\mathscr{E}It$. This must come from the force needed to push the conductor against the magnetic force BIl. In time t this transforms mechanical energy $BIlvt$. So

$BIlvt = \mathscr{E}It$

and we have

$\mathscr{E} = Bvl$ as before.

Since the motion of the conductor is in the opposite direction to the magnetic force, the rule relating field direction, current direction, and motion direction will be a right hand one.

It should help students if we always talk of the e.m.f. induced *in* a wire or circuit, and not across a circuit. Certainly we measure the p.d. across the terminals of a transformer but that e.m.f. is induced in the separate turns of the winding, all connected in series.

The two demonstrations which follow test the relationship between the induced e.m.f. and the rate at which the conductor is moving relative to the field.

DEMONSTRATION
H8 The e.m.f. induced in a moving wire

A wire is moved through a magnetic field at a steady speed; the e.m.f. induced is proportional to that speed. The field can be moved past the wire with the same result.

ITEM NO.	ITEM
1101	sensitive galvanometer
106/1	2 dynamics trolleys
92B	10 Magnadur magnets
92I	5 mild steel yokes
1501	bare copper wire, 2 mm diameter
1501	reel of 0.45 mm PVC-covered copper wire
1040	clip component holder
1153	adhesive tape
1000	leads

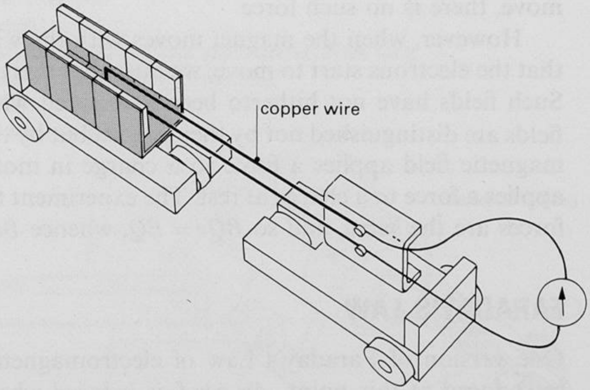


Figure H20
Moving wire or moving magnet.

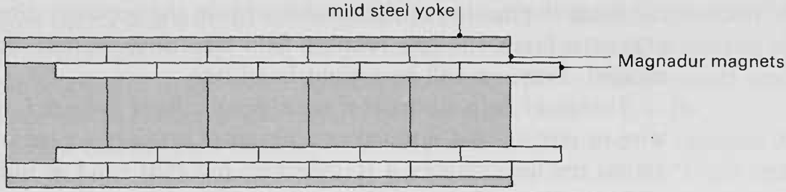


Figure H21
Detail of magnet assembly.

Put two sets of magnets on the yokes as shown in figure H21. This arrangement helps to keep the assembly stable as well as improving the field. Secure the assembly to one trolley with tape. Bend the copper wire into a loop about 30 cm long and 2 cm across, and mount it on a component holder secured to the other trolley. The vertical part of the loop should fall nicely into the magnetic field of the permanent magnet assembly. Connect the loop to the galvanometer by long flexible leads.

When the magnet trolley is held still and the vertical wire moved through the field at a steady speed, a steady deflection is seen on the galvanometer. Doubling the speed doubles this deflection. The surface on which the trolleys move should be as smooth as possible since vibrations in the wire loop may make it difficult to achieve steady deflections.

Moving the wire one way cannot be distinguished from moving the magnet the other; it is only the relative motion which matters.

Some students may appreciate that E-fields and B-fields are closely connected. When the wire moves, the electrons are carried along and the reason for thinking that a B-field is present is that there is a force on the moving charges which will be equal to BQv. If the electrons do not move, there is no such force.

However, when the magnet moves with the wire stationary, we see that the electrons start to move; we infer that there is an E-field present. Such fields have not hitherto been associated with magnets. The two fields are distinguished not by their origin, but by their effect. Whereas a magnetic field applies a force to a charge in motion, an electric field applies a force to a charge at rest. The experiment tells us that these two forces are the same, and so $BQv = EQ$, whence $Bv = E$.

FARADAY'S LAW

One version of Faraday's Law of electromagnetic induction may be introduced at this point. *An e.m.f. is induced when a conductor cuts a magnetic field and it is proportional to the rate at which the conductor cuts the field.*

Lenz's rule

The direction of this induced e.m.f. is given by Lenz's rule. *The e.m.f. induced in a conductor is always in a direction which tends to oppose the motion or change inducing it.*

This can be checked by working out the directions carefully. It is a consequence of the conservation of energy. If the current were in a direction which aided the motion of the conductor, a small push given to the conductor would enable it to increase its motion indefinitely; this does not happen!

Questions

Questions 23 to 25 deal with e.m.f.s induced in moving conductors.

DEMONSTRATION
H9 Inducing e.m.f.s in a motor/dynamo

This demonstration utilizes a motor which is driven as a dynamo. The induced e.m.f.s are larger than those in demonstration H8 and this demonstration shows that they are proportional to the rate of rotation. The demonstration also serves as an introduction to the behaviour of a motor.

ITEM NO.	ITEM
150	fractional horsepower motor
176	12 V battery
1508	demonstration meter, 1 A d.c.
1508	demonstration meter, 5 V d.c.
541/1	2 rheostats, 10–15 Ω, 5 A
1155	geared hand drill
1155	rubber pressure tubing to fit 5 mm shaft, about 10 cm long
1155	steel rod, 5 mm diameter, about 10 cm long
177	lamp, 12 V, 6 W
74	lampholder, s.b.c., on base
1000	leads

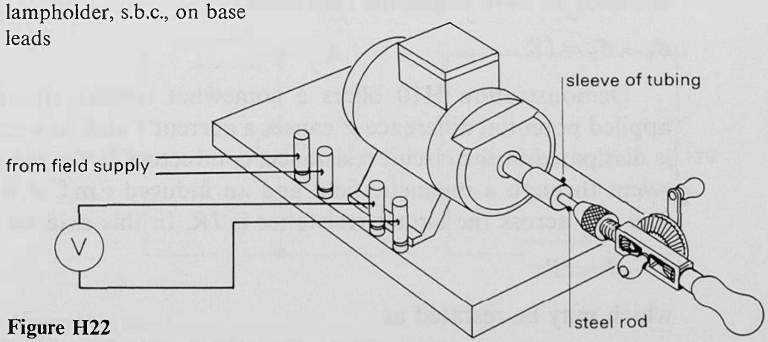


Figure H22
Fractional horsepower motor as a dynamo.

To use the f.h.p. motor as a dynamo, set the field current at 0.5 A and turn the rotor (armature) at a steady speed with the hand drill (see figure H22).

Turning the rotor sweeps many conductors in series through a strong magnetic field. The e.m.f. induced can be shown to be proportional to the rate at which the drill is turned. (The e.m.f. can also be shown to increase as the field current is increased but it is unlikely that direct proportionality will be observed, especially as the iron core of the rotor approaches saturation. This method might be used to investigate the magnetic properties of the iron in the armature.)

If the dynamo is connected to an external circuit, a lamp say, there will be a current in the rotor coils and therefore a force (BIl) on them as they move through the magnetic field. The dynamo should be harder to turn than when it is on open circuit. This effect can be felt. Once again, the conservation of energy applies. The rotor should be turned quite fast with a high field current to make the demonstration effective.

These demonstrations show that a motor may be used as a dynamo and that, when a dynamo is supplying current there is a force tending to retard the motion of the rotor. That fact that dynamo and motor effects occur together whichever machine is used makes it difficult to understand them. The motor has been chosen for a fuller treatment, partly for its own sake and partly because it can help in the understanding of circuits containing inductors, which also have e.m.f.s induced in them.

The motor (item 150) includes a commutator and, when used as suggested above, produces direct current for the external circuit. But, essentially, it is generating an alternating e.m.f. Such e.m.f.s will be considered later.

Before the next demonstration it will be useful to remind students of Kirchhoff's Second Law as it applies to a circuit with more than one source of e.m.f. For the circuit of figure H23 in which the cells are assumed to have negligible resistance,

$$\mathscr{E}_1 - \mathscr{E}_2 = IR$$

Demonstration H10 offers a somewhat similar circuit. Here an applied potential difference V causes a current I and, as a result, energy is dissipated in the circuit resistance; conductors in the rotor are being swept through a magnetic field and an induced e.m.f. $\mathscr{E}$ is generated. The p.d. across the circuit resistance is IR. In this case we may write

$$V - \mathscr{E} = IR$$

which may be restated as

$$V = \mathscr{E} + IR$$

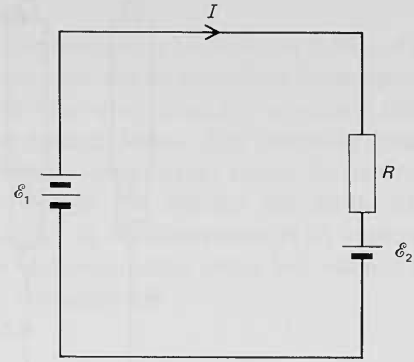


Figure H23

DEMONSTRATION
H10 Behaviour of a fractional horsepower motor

This demonstration introduces a simple treatment of the d.c. motor and deals with the theory which will be required in the experiment which follows.

ITEM NO.	ITEM
150	fractional horsepower motor with pulley wheel
176	2 batteries, 12 V
1508	demonstration meter, 1 A d.c.
1508	demonstration meter, 15 V d.c.
1507	ammeter, 10 A d.c.
541/1	2 rheostats, 10–15 Ω, 5 A
1153	piece of wood to provide a friction load
1153	string
81	2 newton spring balances, 10 N
503–6	2 retort stand bases, rods, bosses, and clamps
44/1	2 G-clamps
1000	leads

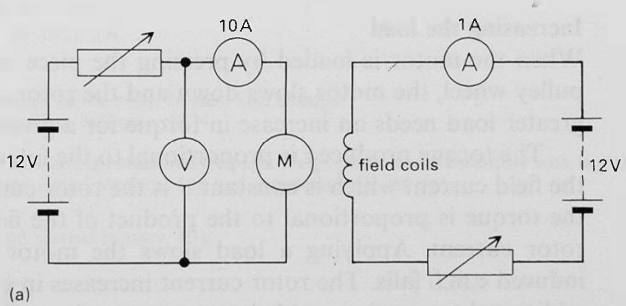


Figure H24 (part)
(a) Circuit for investigation with a motor.

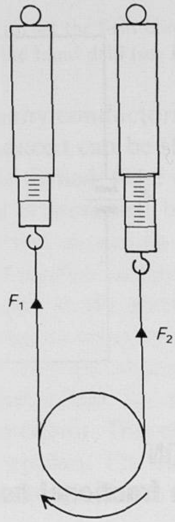


Figure H24 (part)
(b) Band brake on pulley.

Rotor resistance
With the field circuit disconnected, adjust the rheostat in the rotor circuit to give a current, I, of about 5 A. Measure V, the p.d. applied across the rotor, and use $V = IR$ to find the resistance R (which will be about 0.5 Ω). Then connect the field circuit; as the rotor speeds up, the rotor current drops. This suggests the growth of an induced e.m.f. in the rotor, as was seen in demonstration H9. Calling this induced e.m.f. $\mathscr{E}$, the equation becomes $V - \mathscr{E} = IR$. This can be compared with the expression of Kirchhoff's Law for a circuit with opposing e.m.f.s (figure H23). The induced e.m.f. will be found to be almost as great as the applied p.d. V.

Increasing the load
When the motor is loaded by pressing the piece of wood against the pulley wheel, the motor slows down and the rotor current increases. A greater load needs an increase in torque for a given speed of rotation.

The torque produced is proportional to the force BIl. B depends on the field current which is constant. I is the rotor current. Consequently the torque is proportional to the product of the field current and the rotor current. Applying a load slows the motor down and so the induced e.m.f. falls. The rotor current increases in consequence and so additional torque is provided.

Power in the motor

The total power transformed is IV which is $I(IR + \mathscr{E})$ or $I^2R + I\mathscr{E}$. The first of these two terms is the power which heats up the rotor windings; the second provides the work done by the motor, the losses in the iron core, and the mechanical losses. The magnetic losses are difficult to measure; they include eddy-current losses (to be dealt with later) and hysteresis losses (which are beyond the scope of this course). In measuring the torque in demonstration H10, students will be able to estimate the iron and mechanical losses but will not be able to confirm these directly by measurement.

The band brake

Two securely supported 10 N spring balances joined by a string which passes under the pulley wheel on the motor axis constitute a simple band brake – figure H24(b). When the motor is stationary the readings on the two balances are the same – they are pulling against each other. When the pulley wheel rotates, one balance pulls against the other. The difference in their readings $(F_2 - F_1)$ is the force provided by friction between the string and the circumference of the wheel. The frictional force may be increased by raising the spring balance supports. Any point on the circumference of the wheel travels a distance $2\pi r$ in each revolution. The mechanical power delivered by the motor is therefore $2\pi rn(F_2 - F_1)$, where n is the rotational frequency of the rotor.

OPTIONAL EXPERIMENT
H11 The efficiency of a small d.c. motor

ITEM NO.	ITEM
9B or 150	small d.c. motor with pulley wheel
176	12 V battery
1507	ammeter, 1 A d.c.
1507	voltmeter, 10 V d.c.
541/1	rheostat, 10–15 Ω, 5 A
	either
503-6	2 retort stand bases, rods, bosses, and clamps
81	2 newton spring balances, 10 N
1153	string
130/2, 1151	means of measuring rotational frequency, *e.g.*, photodiode assembly with light source and oscilloscope (hand stroboscope and stopwatch would do)
	or
31/1	hanger and slotted masses, 10 g
1153	string
507	stopwatch
501	metre rule
1000	leads

Safety note: If a xenon flasher is used to measure rotational frequency, teachers should be aware that frequencies around 7 Hz have been known to cause epileptic fits with certain people.

This experiment is not so much an investigation as a straightforward measurement of some characteristics of a d.c. motor. The theory has been outlined above and the students now apply it to measure the efficiency of the motor.

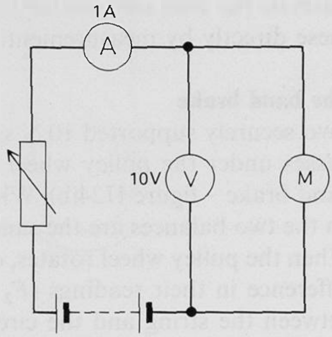


Figure H25
Energy input to a d.c. motor.

Any small permanent-magnet motor can be used and the meters chosen to suit.

Students should measure the resistance of the rotor, the electrical power supplied (from current and p.d. measurements), and the mechanical power delivered at one rotor speed. This can be done by allowing the motor to raise a small load steadily, making measurements of vertical distance moved in a given time.

Students should calculate the efficiency and the electrical and mechanical losses. Some might continue and plot graphs of rotational frequency against load and of rotor current against load. This extension would require the students to find the rotational frequency by viewing a small black disk with a single white radial line through a hand stroboscope, or by using a photodiode, light source, and oscilloscope.

Questions

Questions 26 and 27 are about motors.

FLUX AND INDUCED E.M.F.

The e.m.f. induced in a wire when it sweeps through a magnetic field is $\mathscr{E} = Bvl$. Now the product vl is the area swept out per second and so

$$\mathscr{E} = B \times \text{area swept out per second.}$$

The product of B and the area (at right angles to B) through which it passes is called the *magnetic flux*, Φ. It will be recalled that B has units $\mathrm{N\,A^{-1}\,m^{-1}}$. The units of flux are therefore $\mathrm{N\,A^{-1}\,m^{-1}\,m^{2}}$ or $\mathrm{N\,A^{-1}\,m}$. This unit is known as the *weber* (Wb).

The use of the word 'flux', with its suggestion of flow, will be discussed later in the section.

The faster a conductor cuts magnetic flux, the bigger the induced e.m.f.

An e.m.f. can also be induced in a circuit by changing the flux through it and keeping all the wires still. Electrons are not 'swept through' a field and yet there is still a force on them. The rules for the two distinct effects can be put into exactly the same form, which is itself an alternative form of Faraday's Law:

Induced e.m.f. $\mathscr{E}$ = rate of change of flux linked with the circuit.

The flux is calculated by multiplying the field strength B by the area, perpendicular to B, through which it passes. If the flux links with N turns in series, the induced e.m.f. is N times larger than if it links with one turn.

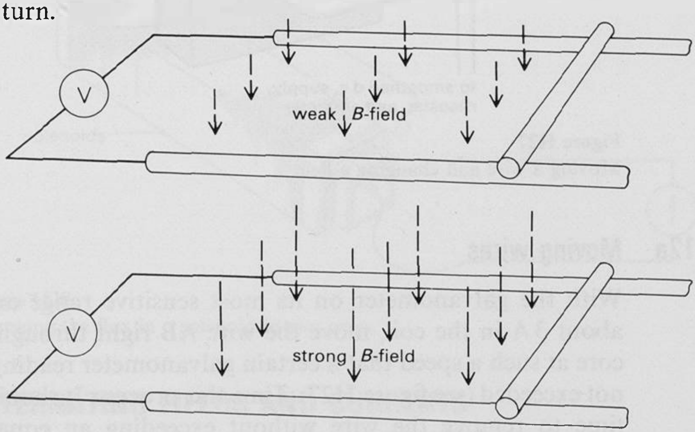


Figure H26
Inducing an e.m.f. without moving any conductor. There is an e.m.f. whenever B changes.

Questions

Questions 28 and 29 are about flux and flux density.

The demonstrations which follow show that an e.m.f. is induced if flux is cut or if the flux linked with a circuit changes. The magnitude of the induced e.m.f. depends on the rate of flux cutting or on the rate of change of flux linked. If the rates are equal, then the induced e.m.f.s are equal.

DEMONSTRATION

H12 Moving wires and changing flux

ITEM NO.	ITEM
147	demountable transformer kit (use 300-turn coil)
1037	set of solenoids
59	l.t. variable voltage supply
1064	low-voltage smoothing unit
541/1	rheostat, 10–15 Ω, 5 A
1508	demonstration meter, 5 A d.c.
1101	sensitive galvanometer
507	stopclock
1000	leads

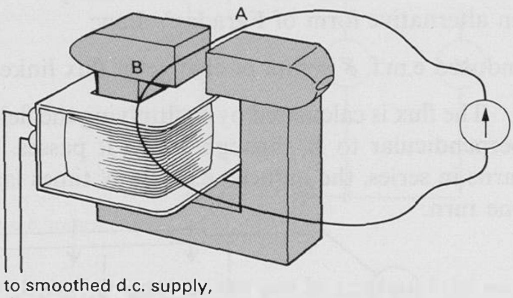


Figure H27
Moving a wire and changing a field.

H12a Moving wires

With the galvanometer on its most sensitive range and a current of about 3 A in the coil, move the wire AB right through the gap in the core at such a speed that a certain galvanometer reading (say 50 mm) is not exceeded (see figure H27). Time this process. It should take the same time to remove the wire without exceeding an equal but opposite galvanometer reading. In a second experiment, link the wire with the iron core so that it lies within the U of the electromagnet and raise the current in the 300-turn coil from zero to 3 A at a rate such that the same galvanometer reading is maintained. The time needed to do this will again be the same.

It is interesting to repeat the demonstration with the wire folded into a long narrow loop. Passing the end of the loop into the magnet will not induce an e.m.f., for the wire is not linked with the flux. It is also worth using two or three turns of wire to show what 'flux-linkage' means.

H12b Changing flux

Two solenoids are required, both close-wound, one of large and one of small cross-section. Connect the smaller solenoid to the galvanometer set on its most voltage-sensitive range and place it inside the larger solenoid. Connect the larger solenoid to the smoothed d.c. supply by way of a rheostat and an ammeter, and pass a current of about 3 A. Pull the smaller solenoid out of the larger one slowly, keeping the galvanometer indication as steady as possible at about 20 mm. Note the time needed to remove the solenoid completely. Then, with the inner solenoid replaced, reduce the current in the outer one to zero, using the control of the variable voltage supply, at such a rate as to keep the galvanometer indication steady at the value used previously. Record this time; it will be close to the earlier time.

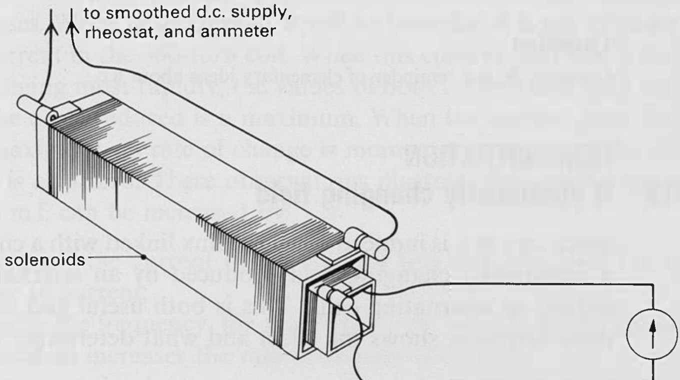


Figure H28
Changing the flux in a solenoid in two ways.

ALTERNATING FIELDS AND CURRENTS

It now becomes important for students to know something of the salient features of a.c. Those who have followed the Nuffield O-level physics course will have important basic knowledge (see REVISED NUFFIELD PHYSICS *Pupils' Text Year 5*, Chapter 7). Others may need to renew their acquaintance with the behaviour of alternating currents. The points to be made include the following:

1 Mains electricity is distributed from generating stations as sinusoidal a.c. with a frequency (in the UK) of 50 Hz and at a range of voltages; in the home the voltage is 240 V.

2 If, say, a 60 W lamp is connected to the supply and one attempts to measure the current with an ordinary moving coil ammeter, the reading will be zero. This is, of course, the average current. (Any demonstration should use a low voltage supply.)
3 Nevertheless, an a.c. ammeter will show a current of around 0.25 A. This raises the question of precisely what it is reading.
4 An alternating current flowing in a coil (*e.g.* of a ticker-timer or the coil used in demonstration H3) produces an 'alternating' magnetic field.
5 Alternating voltages can be increased or decreased by transformers – specially constructed iron rings with a primary and a secondary coil or coils of wire.
6 Search coils which some students may have used can be thought of as the secondaries of air-cored transformers.

Question

Question 30 is a reminder of elementary ideas about a.c.

DEMONSTRATION
H13 A continually changing field

Since an e.m.f. is induced when the flux linked with a circuit is changing, a continually changing field produced by an alternating current will induce an alternating e.m.f. This is both useful and complicated. This demonstration shows the effect and what determines its magnitude.

ITEM NO.	ITEM
147	demountable transformer kit (use 300-turn coil)
1109	signal generator
1501	0.45 mm PVC-covered copper wire
541/1	rheostat, 10–15 Ω, 5 A
1511	double-beam oscilloscope
1508	demonstration meter, 1 A a.c.
1155	wire strippers
1000	leads

Connect the low-impedance output of the signal generator to the 300-turn coil (see figure H29) and select a frequency of about 100 Hz. Connect one input of the oscilloscope across the rheostat so that the deflection of one beam depends on the current through the coil. Connect the second input to a length of wire (about 4 m) which is to be wound round one open limb of the transformer core. As each turn is added, the induced e.m.f. increases. Finally place a laminated bar across the open limbs with the coil in place.

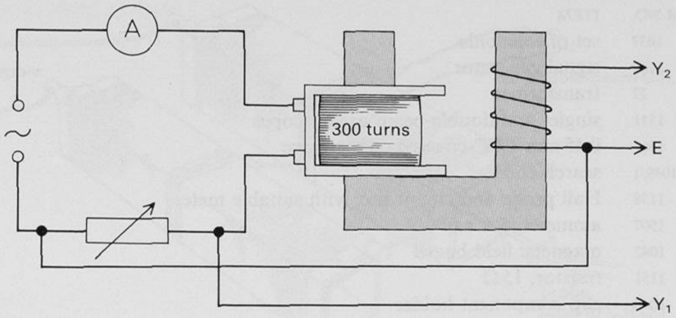


Figure H29
Winding a transformer turn by turn.

As turns are added to the open limb of the core, the alternating e.m.f. increases. When large enough, it will be seen that it is not in phase with the current in the 300-turn coil. When this current (and hence the field) is changing most rapidly, the values of both current and field are zero, but the e.m.f. induced is a maximum. When the current (and the field) is a maximum, its rate of change is momentarily zero and the induced e.m.f. is also zero. These observations illustrate Faraday's Law nicely. The e.m.f. can be increased by:

i increasing the current in the coil, for this also increases the rate of change of current;
ii raising the frequency, for this reduces the time in which the current rises and so increases the rate of change of current;
iii placing a laminated iron yoke on the U-core, for this greatly increases the flux produced by a given current (it may be difficult to maintain the same current in the coil because, as we shall see later, the inductance increases).

These changes are neatly summarized by Faraday's Law that the e.m.f. induced is equal to the rate of change of flux linkage.

STUDENT DEMONSTRATION
H14 Induction using a.c.

These experiments, which students may set up and then demonstrate to each other, show how various factors affect the alternating e.m.f. induced. Each demonstration should reinforce the understanding of the concept of flux and provide a further illustration of Faraday's Law.

ITEM NO.	ITEM
1037	set of solenoids
1109	signal generator
27	transformer
1511	single- and double-beam oscilloscopes
1501	0.45 mm PVC-covered copper wire
1039/1	search coil
1138	Hall probe and circuit box with suitable meter
1507	ammeter, 5 A a.c.
1042	magnetic field board
1151	resistor, 15 Ω
1040	clip component holder
1000	leads

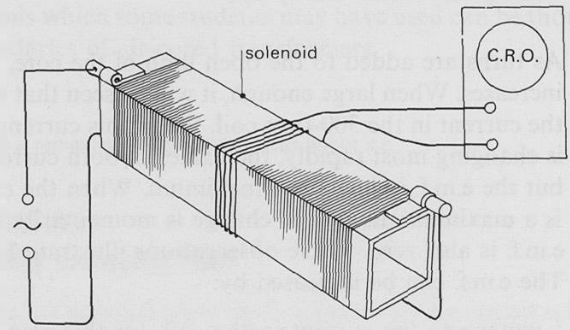


Figure H30
Effect of number of turns on induced e.m.f.

H14a Number of turns

See figure H30. The larger close-wound solenoid is fed with a.c. from the low-impedance output of the signal generator at about 2 kHz or, less effectively, from a transformer at 50 Hz. A 3 to 4 m length of covered wire is wrapped round the centre of the solenoid, turn by turn, with its ends connected to the input of the oscilloscope, whose sensitivity is set to 0.1 V cm^{-1}. For good quantitative results, the wires leading to the oscilloscope should be twisted together – this will require disconnection from the oscilloscope whilst each additional turn is added. As the number of turns is increased, the induced e.m.f. indicated by the oscilloscope rises. At mains frequency, 10 to 20 turns will be needed, but at 2 kHz, 2 to 10 turns will suffice.

H14b Area

See figure H31. Two close-wound solenoids, one large and one small, are connected in series. They are kept well apart and parallel to one

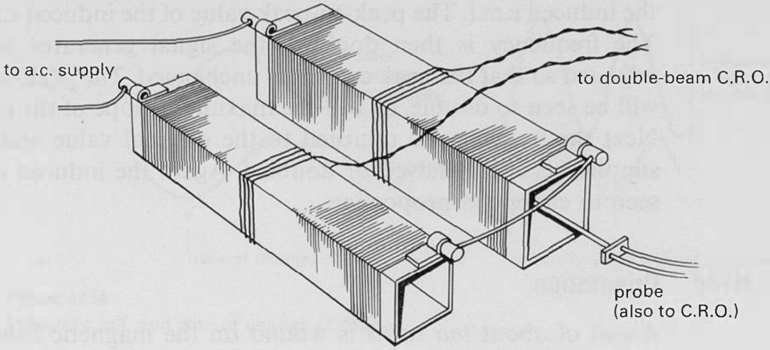


Figure H31
Effect of area on induced e.m.f.

another so that the field inside one is little affected by the field of the other. They can be fed from a 12 V transformer and the fields inside them shown to be equal, using either a Hall probe or a search coil. Ten turns of the covered wire are wrapped tightly round the middle of each solenoid; an oscilloscope (preferably double-beam) is used to compare the induced e.m.f.s.

H14c Field and rate of change

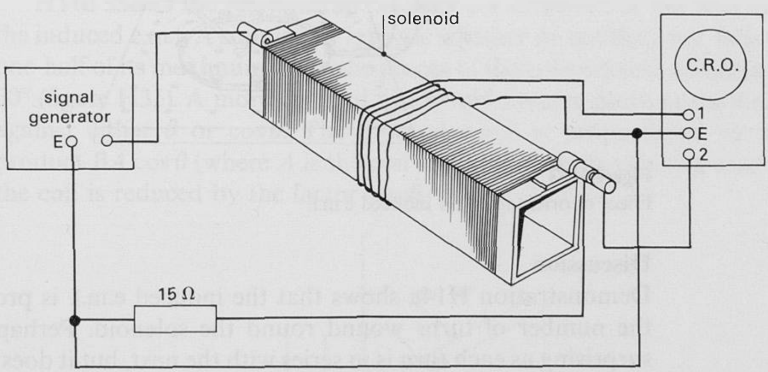


Figure H32
Effect of rate of change of flux on induced e.m.f.

A double-beam oscilloscope is essential for this demonstration (see figure H32). One input is connected across the 15 Ω resistor and indicates the current in the close-wound, fatter solenoid which comes from the low-impedance output of a signal generator set at about 1 kHz. This beam is used to trigger the time-base. The other input is connected to a ten-turn coil wrapped around the centre of the solenoid and shows the induced e.m.f. The peak-to-peak value of the induced e.m.f. is noted. The frequency is then doubled, the signal generator output being adjusted so that the peak current is unchanged. The peak-to-peak e.m.f. will be seen to double, as will the maximum slope of the current trace. Next the frequency is returned to the original value and the current supplied is either halved or doubled. Again the induced e.m.f. will be seen to change in proportion.

H14d Orientation

A coil of about ten turns is wound on the magnetic field board (see figure H33). It may be roughly circular or even square. It is supplied with current of about 5 A from a transformer. A search coil connected to an oscilloscope is placed at the centre of the coil with its plane parallel to the coil. The induced e.m.f. is noted. The search coil is then tilted until its plane is perpendicular to the coil whilst the change in the e.m.f. is observed.

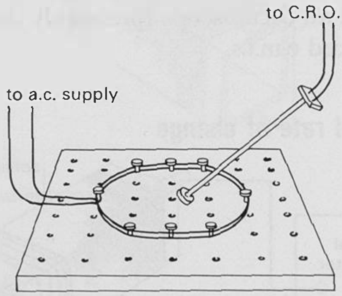


Figure H33
Effect of orientation on induced e.m.f.

Discussion
Demonstration H14a shows that the induced e.m.f. is proportional to the number of turns wound round the solenoid. Perhaps this is not surprising as each turn is in series with the next, but it does illustrate the idea of flux linking a coil. The flux linkage can also be increased by increasing the current in the solenoid.

H14b shows that, whereas the *B*-field inside each solenoid is the same, more flux links the larger coil because it has a larger area. The cross-sectional area of the large solenoid is twice that of the smaller and the ratio of the induced e.m.f.s is also 2:1.

H14c shows that the induced e.m.f. is proportional to the rate of change of flux linked with the coil. When the frequency is doubled with the amplitude kept constant, the maximum rate of change is also

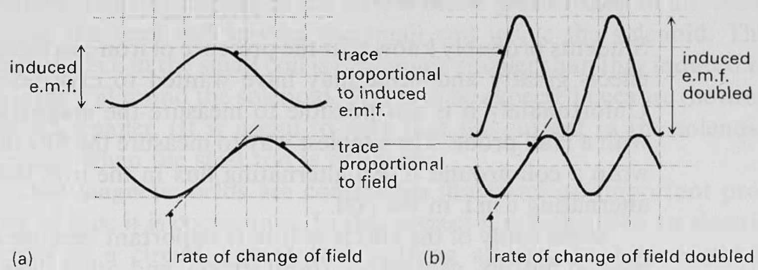


Figure H34
Induced e.m.f. and rate of change of field.

doubled since the current, and hence the field, must change by the same amount in half the time (figure H34); so the induced e.m.f. is doubled. The peak e.m.f. is also shown to be directly proportional to the current in the solenoid. Students should use the oscilloscope calibrations to measure the peak e.m.f. induced and to compare these with the result obtained from the equations

$$B = \mu_0 N_s I / l \qquad \text{and} \qquad \mathscr{E} = N_c \mathrm{d}\Phi / \mathrm{d}t$$

where N_s and N_c are the numbers of turns on the solenoid and the external coil.

H14d shows the relationship between the direction of the field and the induced e.m.f. A simple test is to see whether or not the e.m.f. falls to one-half of its maximum when the planes of the coils enclose an angle of 60° (figure H35). A more detailed test would involve plotting the e.m.f. against either θ or $\cos\theta$. The induced e.m.f. is proportional to the product $BA\cos\theta$ (where A is the coil area) because the effective area of the coil is reduced by the factor $\cos\theta$.

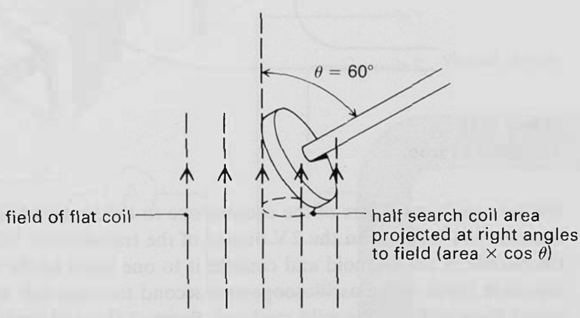


Figure H35
The $\cos\theta$ rule.

THE EFFECT OF IRON

Students probably know that the presence of iron can increase magnetic effects greatly and some may have wanted to experiment with iron. Unfortunately it is not possible to measure the magnetic flux in iron with a Hall probe. The simplest way to measure the flux in the iron is to wrap a coil around it. An alternating flux in the iron then induces an alternating e.m.f. in the coil.

Some study of the effects of iron is important because iron is widely used in motors, generators, transformers, and other devices.

DEMONSTRATION
H15 The effect of iron in a solenoid

An iron rod about 12 mm in diameter is put inside a solenoid. When a current flows in the solenoid the total flux increases somewhat; the flux density is much greater in the iron than in the air.

ITEM NO.	ITEM
504	retort stand rod of mild steel (not stainless)
27	transformer
1511	double-beam oscilloscope
1037	large close-wound solenoid
1501	0.45 mm PVC-covered copper wire
1507	ammeter, 1 A a.c.
1000	leads

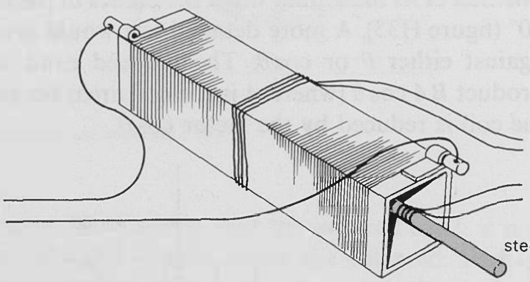


Figure H36
The effect of iron.

Set the two Y-amplifiers of the oscilloscope to the same sensitivity of $0.5\ \mathrm{V\,cm^{-1}}$. Connect the solenoid to the 2 V output of the transformer. Wind a ten-turn coil round the middle of the solenoid and connect it to one input of the oscilloscope. Connect the other input of the oscilloscope to a second ten-turn coil which is closely wound round the middle of the mild steel rod. Support this rod centrally inside the solenoid.

The induced e.m.f. in the small coil around the steel rod is found to be almost as great as the e.m.f. induced in the large coil around the solenoid. The flux density in the steel is much greater than in air. Now remove the steel rod leaving the small coil inside the solenoid. The induced e.m.f. in the small coil is now much smaller than that induced in the large coil round the solenoid. This is to be expected because the area is much smaller. (It is useful to note that the current in the solenoid increases when the steel rod is removed.)

That magnetic fields are continuous illustrates an important property of flux; it is continuous. In this respect it is rather like an electric current in a circuit, except that nothing appears to flow round a 'magnetic circuit'. In the demonstration, the 'magnetic circuit' which includes the iron rod also includes a great deal of air and this accounts for the rather small effect on the total flux through the solenoid.

EXPERIMENT
H16 Increasing the iron in a magnetic circuit

Students can experiment with different iron circuits to see how they affect the total flux.

ITEM NO.	ITEM
92G	2 C-cores, preferably matched, with clip
1153	pieces of card and paper
529	scissors
27	transformer
1501	0.45 mm PVC-covered copper wire
1511	oscilloscope
1507	ammeter, 1 A a.c.
1507	voltmeter, 5 V a.c.
541/1	rheostat, 10–15 Ω, 5 A
1155	micrometer screw gauge
1000	leads

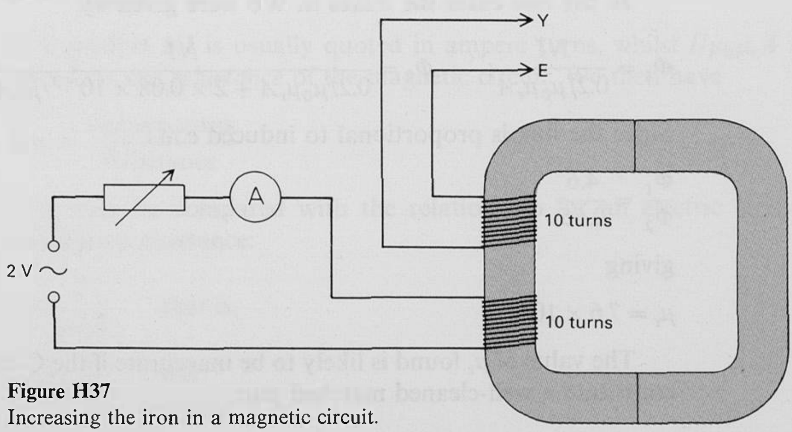


Figure H37
Increasing the iron in a magnetic circuit.

A coil of ten turns, wound on a C-core, is fed with 1 A a.c. from the transformer. A second coil with ten turns is wound round the core and is connected to the oscilloscope to display the e.m.f. induced in it, and hence to indicate the flux. It should be possible to slide this second coil along the core. The magnetic circuit can be completed with the second C-core whilst pieces of paper or card can be put between the cores to make small 'air' gaps.

With a single C-core the flux linked with the coil is seen to be fairly uniform as it is slid along the core, but to decrease rapidly in the air as the flux spreads out at the end of the core. With two C-cores making good contact, the flux is much greater and is fairly uniform round the magnetic circuit. With a piece of paper or card cut to fit the end of a C-core to make a small gap in the magnetic circuit, the flux is considerably reduced but remains fairly uniform round the circuit.

Measuring the relative permeability

Students can replace the 10-turn secondary coil by a 120-turn coil and connect this coil to an a.c. voltmeter. They should measure the secondary voltage when the iron circuit is complete and also when it is interrupted by one or more thicknesses of paper between the cores. The primary current should be kept constant at, say, 1.0 A. The secondary voltages can then be used to compare the reluctances of the circuit before and after adding the paper (which is assumed to have a relative permeability of unity). From these measurements μ_r for the iron can be calculated.

In a typical experiment, the primary current was kept constant at 1.0 A. With no gap between the cores, the secondary voltage was 4.6 V; with one thickness of paper in each gap of 0.08 mm, the secondary voltage was 0.65 V. The length of the magnetic circuit was 0.20 m, measured along the centre of the C-cores.

In the two cases the fluxes in Wb were given by

$$\Phi_1 = \frac{NI}{0.2/\mu_0\mu_r A}; \qquad \Phi_2 = \frac{NI}{0.2/\mu_0\mu_r A + 2 \times 0.08 \times 10^{-3}/\mu_0 A}$$

Since the flux is proportional to induced e.m.f.

$$\frac{\Phi_1}{\Phi_2} = \frac{4.6}{0.65}$$

giving

$$\mu_r = 7.6 \times 10^3$$

The value of μ_r found is likely to be inaccurate if the C-cores do not constitute a well-cleaned matched pair.

Reluctance

Where a solenoid is wound in a circle (see figure H38), the field strength inside is everywhere the same as it is in the *centre* of a long straight solenoid:

$$B = \mu_0 NI/l$$

and the flux is

$$\Phi = \mu_0 NIA/l$$

where A is the cross-sectional area of the solenoid.

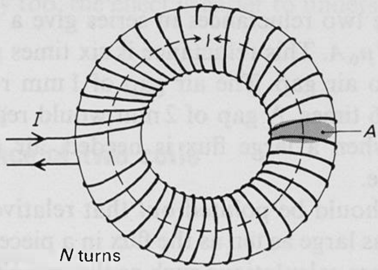


Figure H38 *N* turns

When the solenoid is formed round a cylinder of iron, the flux is increased by a factor μ_r, which is known as the relative permeability of the iron. (It will be recalled that μ_0 is the permeability of free space.) It follows that the flux in the iron is

$$\Phi = \mu_0 \mu_r NIA/l$$

This may be rearranged as

$$\Phi = \frac{NI}{l/\mu_0 \mu_r A}$$

The product NI is usually quoted in ampere turns, whilst $l/\mu_0 \mu_r A$ is known as the *reluctance* of the magnetic circuit. We then have

$$\text{flux} = \frac{\text{ampere turns}}{\text{reluctance}}$$

This may be compared with the relationship for an electric circuit containing resistance:

$$I = \frac{V}{\rho l/A} \qquad \text{that is,}$$

$$\text{current} = \frac{\text{p.d.}}{\text{resistance}}$$

The two relationships are similar in form, with the permeability $\mu_0\mu_r$ in the former corresponding to $1/\rho$ in the latter (ρ is the resistivity).

In the same way that resistances in series are added to find the total resistance, reluctances in series are added to find the total reluctance. A simple calculation shows why even a small air gap in a magnetic circuit causes a large reduction in flux.

Suppose that an iron circuit has a 0.2 m length of iron with a relative permeability of 1000. The reluctance in $\mathrm{A\,Wb^{-1}}$ is $0.2/1000\,\mu_0 A$, that is $0.0002/\mu_0 A$.

Introducing an air gap of only 1 mm (0.001 m) will give an added reluctance in $\mathrm{A\,Wb^{-1}}$ of $0.001/\mu_0 A$ (assuming μ_0 for air to be unity).

The two reluctances in series give a total reluctance in $\mathrm{A\,Wb^{-1}}$ of $0.0012/\mu_0 A$. This reluctance is six times greater than that of the circuit with no air gap. The air gap of 1 mm reduces the flux by a factor of about 6 times. A gap of 2 mm would reduce it about 11 times. This is why, when a large flux is needed, air gaps are eliminated as far as possible.

It should be pointed out that relative permeabilities can vary by a factor as large as ten as the flux in a piece of magnetic material changes. Therefore calculations such as the one above are not nearly as accurate as similar calculations with resistances, which do not usually vary very much as the current changes. The REVISED NUFFIELD ADVANCED SCIENCE *Book of data* gives initial and maximum relative permeabilities for some magnetic materials.

It is worth stressing the importance of reluctance in practical engineering. A car starter motor, generator, or alternator (obtainable from a breaker's yard), or a vacuum cleaner motor is a useful exhibit. Contrast the construction – the number of coils, the amount of iron present, and the way in which it fills up most of the space – with the model motors made in school physics courses. Another example has to do with instrumentation. The change in reluctance of a magnetic circuit with a variable air gap leads to a change in flux. This can be made the basis of a proximity meter or a vibration monitor. Several examples of applications to measurement and control systems are given in BARCLAY and GIBBON *Physics principles at work.*

Reading

One of the electromagnetic flowmeters described in the Reading on page 16 of the *Students' guide* uses several of the ideas from this section of the Unit (induced e.m.f., reluctance, ...).

Questions

Questions 31 to 35 are about flux, the effect of iron, and reluctance.

INDUCTANCE

Although mutual inductance is little used in this course, it makes a good introduction to the subject of inductance in general. It is comparatively easy to understand how a current change in one coil can induce an e.m.f. in another coil; indeed demonstration H17 shows just that. That a current change in a coil can induce an e.m.f. in the same coil and hence modify the current is rather more difficult to comprehend, especially as the induced e.m.f. varies. This is why the operation of the d.c. motor was examined in some detail (see experiments H10 and H11). In a motor an induced e.m.f. also modifies the current flowing but, as, at steady speed, this e.m.f. is steady too, the effect is easier to understand than it is in an inductor.

DEMONSTRATION

H17 Mutual inductance of two coils

ITEM NO.	ITEM
1109	signal generator
1511	double-beam oscilloscope
1151	resistor, 1 Ω
1058	2 120 + 120 turn coils
1000	leads

The oscilloscope trace Y_1 monitors the p.d. across the 1 Ω resistor – see figure H39(a) (page 62) – and hence displays directly the current in the primary coil. Trace Y_2 measures the e.m.f. induced in the secondary coil, which should be placed adjacent to the primary (note that no iron core is used). Set the signal generator to 1 kHz sine wave, and use the low-impedance output. Adjust the amplitude until the peak p.d. across the 1 Ω resistor is about 400 mV, corresponding to a peak current of 400 mA. Increase the oscilloscope time-base speed to $10\,\mu\text{s}\,\text{cm}^{-1}$ and the Y_1 sensitivity to $50\,\text{mV}\,\text{cm}^{-1}$, adjusting the trigger control to examine the part of the Y_1 waveform which rises nearly linearly. (Note that the graticules supplied with some oscilloscopes use grid squares which are not 1 cm × 1 cm.)

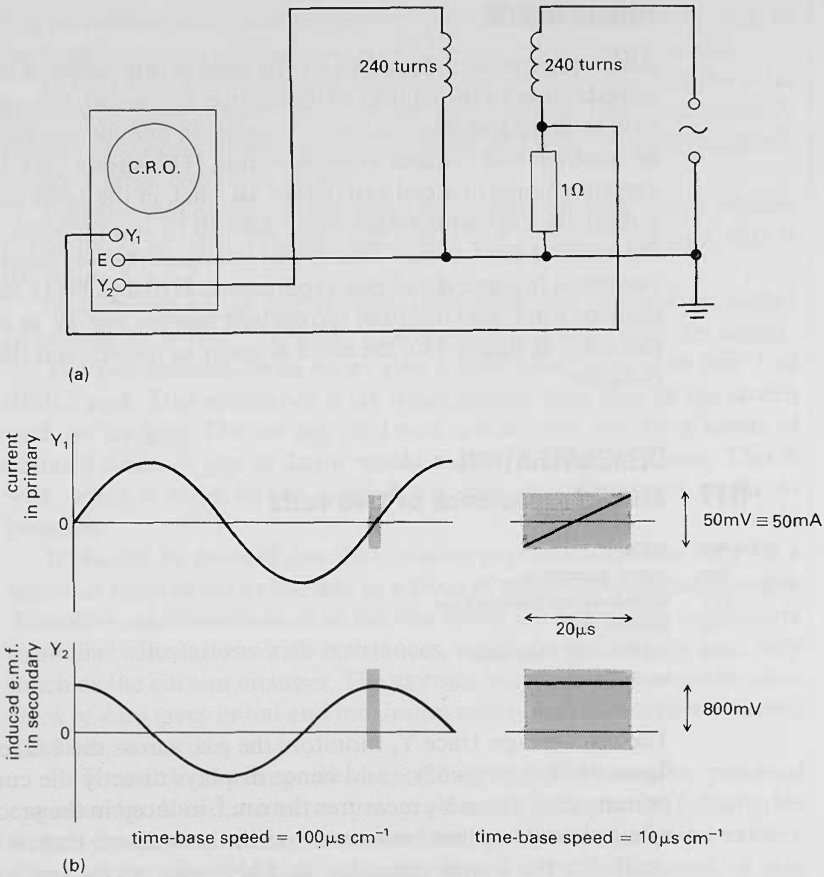


Figure H39
(a) Mutual inductance.
(b) Measuring material inductance – oscilloscope display.

Measuring the mutual inductance

Slight adjustments to the trigger and position controls may be needed to obtain a display similar to that of figure H39(b). The amplitude of the signal generator output may be adjusted to give a convenient measurement for dI/dt.

The induced e.m.f. in the secondary coil, $\mathscr{E}$, is found to be proportional to the rate of change of current in the primary coil. The ratio between them is called the *mutual inductance*, M,

$$M = \frac{\mathscr{E}}{\mathrm{d}I/\mathrm{d}t}$$

The unit of inductance is $\mathrm{V\,s\,A^{-1}}$. It is called the *henry* (H), after the American physicist who discovered self induction in 1832.

From the values given in figure H39(b),

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{50\,\mathrm{mA}}{20\,\mu\mathrm{s}} = 2.5 \times 10^3\,\mathrm{A\,s^{-1}}$$

$$\mathscr{E} = 800\,\mathrm{mV}$$

$$M = \frac{\mathscr{E}}{\mathrm{d}I/\mathrm{d}t} = \frac{800\,\mathrm{mV}}{2.5 \times 10^3\,\mathrm{A\,s^{-1}}} = 320\,\mu\mathrm{V\,s\,A^{-1}}$$

$$= 320\,\mu\mathrm{H}$$

Changing the numbers of turns on the primary (N_p) and secondary (N_s) windings will demonstrate that the mutual inductance is proportional to the product N_pN_s. M also depends on the material of the core and, because the permeability is variable, it also depends on the current in the primary coil.

Question

Question 36 is about the determination of the value of M from measurements of $\mathscr{E}$ and $\mathrm{d}I/\mathrm{d}t$.

DEMONSTRATION
H18 Self induction

Students may have wondered about self inductance already. An alternating current in a coil is reduced when an iron core is put into the coil, as has been noted in demonstration H15. Two more demonstrations follow.

ITEM NO.	ITEM
1030	high-inductance coil
92G	double C-core with clip
92R	2 m.e.s. lamps, 2.5 V, 0.3 A
92S	m.e.s. neon lamp
92T	3 m.e.s. lampholders
1033	cell holder with two cells
52L	mounted bell push (or switch)
541/1	rheostat, 10–15 Ω, 5 A
1000	leads

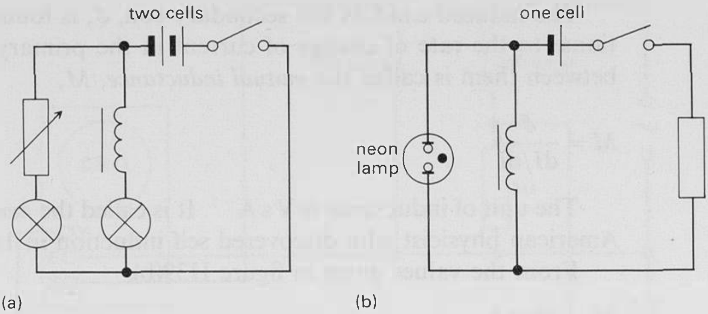


Figure H40
Simple inductor experiments.

Adjust the rheostat – figure H40(a) – until the two lamps are equally bright, then switch the current off. If possible the switch should make an audible sound as it is closed. This makes it easier for the students to observe that, when the switch is closed again, the lamp in series with the inductor comes on after the lamp in series with the resistor.

The delay is a consequence of the induction of an e.m.f. in the inductor as the current in it grows from zero. This is a clear application of Lenz's rule: as the flux linking the circuit grows, the induced e.m.f. tends to oppose that growth.

The inductive effect that accompanies switching the current off is unlikely to be observed because of the resistance in the circuit containing the two lamps.

This effect can be shown in a striking way by using a small neon lamp as in figure H40(b). Opening the switch will induce an e.m.f. high enough to cause the lamp to start to glow (which needs around 70 V).

If a 20 000 turn coil with its massive C-core is available, the slow rise of a current may be shown very easily. In this case the neon lamp will remain lit for quite some time after switching off; but beware of shocks!

Self inductance

In demonstration H10 the current in the rotor coils was reduced because an e.m.f. was induced in them when the rotor turned in the magnetic field. The relationship $V = IR$ for the stationary rotor became $V - \mathscr{E} = IR$ for the rotating one. It is useful to show a similar effect in a coil. A 240-turn coil on a pair of C-cores will pass about 1.5 A when a d.c. potential difference of 2 V is applied. Evidently its resistance is about 1 Ω. When a 50 Hz a.c. potential difference of 2 V is applied, the current is only about 0.02 A. The changing flux due to the changing current in the coil induces an e.m.f. in the coil itself. This induced e.m.f.

reduces the current (as it must by Lenz's rule, or, more fundamentally, by the conservation of energy).

The e.m.f. induced in an inductor L is $L\,\mathrm{d}I/\mathrm{d}t$ and the equation given above can be written $V = IR + L\,\mathrm{d}I/\mathrm{d}t$ and interpreted as follows. In a circuit containing resistance and inductance the applied p.d. is equal to the sum of the p.d. across the resistor (IR) and the p.d. across the inductor ($L\,\mathrm{d}I/\mathrm{d}t$). The p.d. across a resistor depends on its resistance and on the current; the p.d. across an inductor depends on its inductance and on the rate of change of current.

Note to teachers: p.d. and e.m.f.

From now on we shall write equations in the form $V = IR + L\,\mathrm{d}I/\mathrm{d}t$ and we shall speak of the p.d. across an inductor rather than the e.m.f. induced in it. If students find this change of viewpoint confusing, it may help to consider again the circuit shown in figure H23 (page 43) in which R is the total circuit resistance (including the resistances of the cells). We can consider both batteries as *sources* of energy and write Kirchhoff's Law as $\mathscr{E}_1 - \mathscr{E}_2 = IR$.

Suppose, however, that battery 2 in figure H23 is a storage battery (*e.g.* a lead–acid accumulator). In the circuit of figure H23, this battery is being charged – it is acting as a *sink*. In a similar way, an inductor can be regarded as a sink rather than a source. When the current grows in the inductor, energy is needed to increase the magnetic field. The magnetic field stores this energy, just as energy is stored in the storage battery. Analogous cases include the storage of energy in the electric field between the plates of a charged capacitor, and the storage of energy when a mass is lifted in a gravitational field. When the current in the inductor is reduced, energy is returned from the magnetic field to the electric circuit just as the storage battery or the charged capacitor will return their stored energy to a circuit.

How much of this discussion it is useful to pass on to students is of course a matter for the teacher to judge; little or none would probably be better than too much.

DEMONSTRATION

H19 Measurement of self inductance

The rate of change of current in a coil is measured when a potential difference is applied to it, first with d.c. and then with a.c. The resistance of the coil does not matter very much because both measurements are taken when the current in the coil is very small.

ITEM NO.	ITEM
1030	high-inductance coil
92G	double C-core and clip
1033	cell holder with one cell
59	l.t. variable voltage supply
1511	double-beam oscilloscope
1017	resistance substitution box
52L	mounted bell push (or switch)
1000	leads

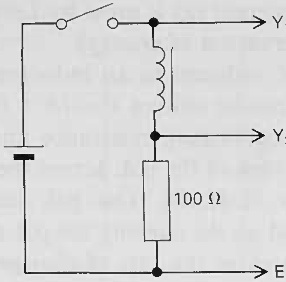


Figure H41
Circuit for measurement of self inductance.

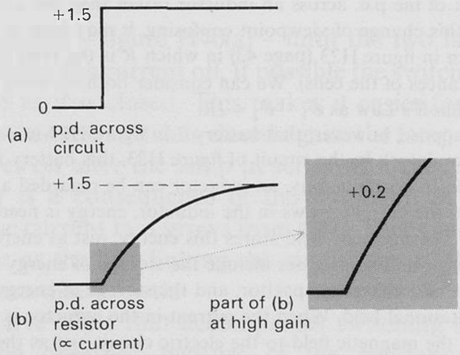


Figure H42
Measurement of self inductance, oscilloscope traces.

The circuit (see figure H41) contains both inductance and resistance, and when the switch is closed the current rises slowly. The Y_1 trace shows the sudden rise of p.d. across the inductor (the exact voltage does not matter but it is best to use only a single cell so that there is little risk of the iron in the inductor saturating). The Y_2 trace shows the p.d. across the 100 Ω resistor and so indicates the current in the inductor. Show the rise of current (with Y_2 set to 0.2 V per division and the time-base set to 50 ms per division). It is now necessary to concentrate on the very beginning of the rise of current. Set the controls of the oscilloscope to give an almost diagonal line across the screen, see figure H42(b). Suitable settings are 20 mV per division for Y_2 and 1 ms per division for the time-base. Trigger the time-base from Y_1 and set the trigger-level control so that the time-base triggers each time the switch is closed. The initial rate of rise of current can be found from the initial rate of rise of the Y_2 trace, which is $R\,dI/dt$.

When the current is zero, there is no p.d. across the resistor. It follows that at that instant the p.d. across the inductor is equal to V, the p.d. applied to the circuit. Thus,

$$V = L\frac{\mathrm{d}I}{\mathrm{d}t}$$

The gain and the time-base speed may then be reduced to show the full trace of the rise of the current in the circuit. The highest current reached is given by

$$V = IR$$

These two equations are both special cases of the general equation

$$V = IR + L\frac{\mathrm{d}I}{\mathrm{d}t}$$

The initial rate of rise of current depends only on the inductance and has nothing to do with the resistance of the circuit, whereas the final current depends only on the resistance of the circuit and has nothing to do with the inductance.

Taking the iron core out of the coil reduces the inductance and so the current rises much more quickly; but the final current is the same as before.

The behaviour of the inductor when a.c. is supplied is important. The coil will allow less current to pass than with a d.c. supply of comparable p.d.; with the iron core in place, the current is considerably less. Why? The a.c. is continually changing. As we have seen in demonstration H14, the peak voltage determines the maximum rate of change of current; the period determines the time for which the current can rise (or fall); so the current is limited even if the resistance of the circuit is negligible.

We may repeat demonstration H19 with 50 Hz a.c. from a variable voltage supply in order to examine this further, making more accurate measurements of the steady trace which is obtained (see figure H43).

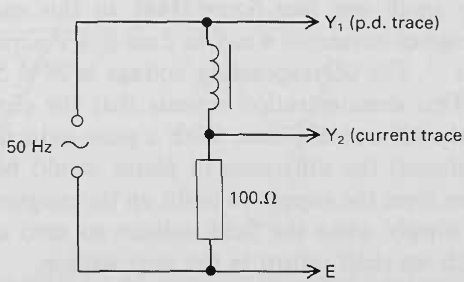


Figure H43
Measurement of self inductance using a.c.

Set the sensitivity of the oscilloscope to $0.2\,\mathrm{V\,cm^{-1}}$ for one input and arrange for the sweep to trigger from a negative-going signal at this input. Set the time-base at $2\,\mathrm{ms\,cm^{-1}}$, so that one cycle of the 'current' trace taken across the resistor (Y_2) fills the screen. Adjust the X and Y shift controls and the supply voltage so that the steepest part of this trace is centred on the screen and the slope adjusted to a conveniently measurable value (see figure H44). The 'voltage' trace (Y_1) requires a sensitivity of about $5\,\mathrm{V\,cm^{-1}}$.

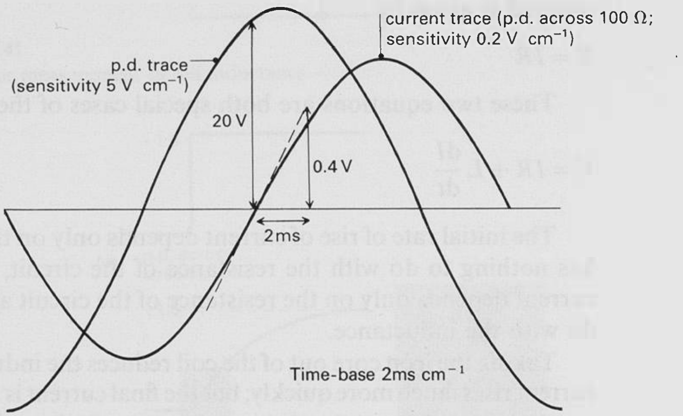


Figure H44
Measurement of self inductance using a.c.

The current through the $100\,\Omega$ resistor also flows through the inductor and so the Y_2 voltage can be used to calculate the current through the inductor. At the moment when the current is zero, the p.d. measured by Y_1 gives the induced e.m.f. in the inductor.

The inductance is the ratio of the voltage when the current is zero to the maximum rate of change of current (obtained from the slope of the 'current' trace). The error involved in using the peak voltage (rather than the voltage when the rate of change of current is a maximum) is a very small one (see figure H44). In this case, the maximum rate of change of current is 4 mA in 2 ms (0.4 V across $100\,\Omega$ in 2 ms) which is $2\,\mathrm{A\,s^{-1}}$. The corresponding voltage is 20 V. So the inductance is 10 H.

This demonstration reveals that the current and the voltage are nearly 90° out of phase. With a pure inductor (*i.e.* one with no ohmic resistance) the difference in phase would be precisely $\pi/2$. Energy is taken from the supply to build up the magnetic field and is returned to the supply when the field reduces to zero again. This is a matter to which we shall return in the next section.

Questions

Questions 37 to 41 are about self inductance.

The jumping ring

Earlier in this Unit, demonstration H3 was presented as an intriguing puzzle to be explained later. Students should now be able to construct an explanation along the following lines. Alternating current in the 240-turn coil produces a changing magnetic flux in the iron rod. Changing flux in the rod induces an alternating e.m.f. in the aluminium ring. This ring has a very low resistance and so large currents circulate in it (causing it to get quite warm). In short, the system is a transformer, and since $N_s \ll N_p$, $I_s \gg I_p$. (See Section H3.)

The ring is pushed upwards, away from the coil. This implies an interaction between the current induced in the ring and the current in the primary 240-turn coil. Since the force is repulsive, the two currents must be in opposite senses for more than half of each cycle of a.c. The induced e.m.f. in the ring reaches its extreme values when $d\Phi/dt$ has its extreme values, *i.e.* when the current in the primary is changing most rapidly. That is, when the primary current is momentarily zero. If the current in the ring were in phase with the induced e.m.f., then the primary and secondary currents would be one-quarter of a cycle out of phase; they would be in the same sense and in the opposite sense for equal fractions of each cycle. But because the ring has a very low resistance the relationship between induced e.m.f. and current is determined largely by the inductance of the ring. So the current lags the induced e.m.f. (see figure H45).

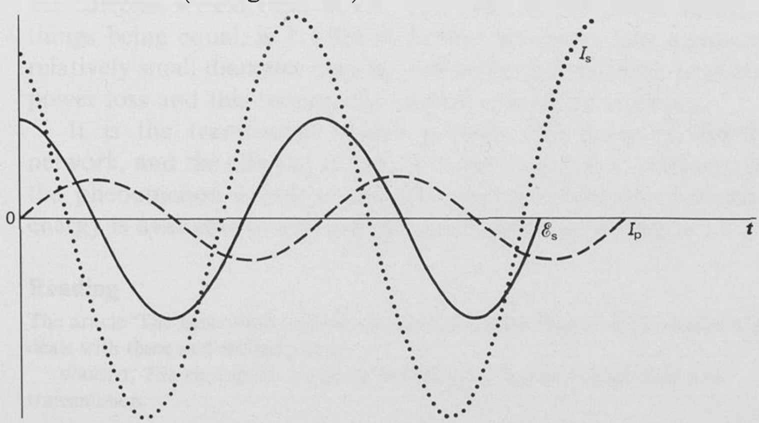


Figure H45
Current in 240-turn coil (I_p), e.m.f. induced in the ring ($\mathscr{E}_s$), and current in the ring (I_s) against time.

A quantitative version of the argument is possible and may offer a useful challenge to keen students. They will need to make estimates and to apply many of the ideas studied in this Unit. Estimating the phase lag for the current in the ring will take them beyond the material covered in the course.

SECTION H3
ALTERNATING CURRENT

TRANSFORMERS

It was necessary briefly to recall the basic facts about a.c. in the previous section. Now the study must be taken further. It will be useful to refer back to demonstration H9, in which a d.c. motor was driven as a generator. In this device a commutator is used to maintain a d.c. output but the rotor carries a set of coils which are rotated in a stationary magnetic field and the e.m.f.s induced in the coils are alternating in nature. In commercial generators, a rotating magnetic field sweeps past three sets of stationary coils, inducing an alternating e.m.f. in each set. In this course we shall concentrate on single-phase a.c. since this is the form in which the energy is conveyed to our homes; however, a brief reference to three-phase a.c. may be useful in order that students may appreciate why there are three high-tension cables on the transmission lines of the National Grid. This extends over some 14 000 km, distributing energy at voltages of 400, 275, 132, 66, and 33 kV, with smaller lines operating at 11 and 6.6 kV.

It may be necessary to remind students that the high voltages are required in order to attain high efficiency in the distribution network (about 93 % overall – see Unit G, 'Energy sources'). Since electrical heating of the transmission line is dependent on I^2, it is important to keep I small. To transmit the 660 MW output of a large generating station at 400 kV requires currents totalling 1.65 kA. But at, say, 11 kV, the currents would total 60 kA. The ratio of the power losses, other things being equal, is 1:1300. A further benefit is that conductors of relatively small diameter may be used without exceeding an acceptable power loss and this reduces the capital cost of the network.

It is the transformer which permits this form of distribution network, and the efficient design of transformers is a necessary part of the phenomenon everyone usually takes for granted: that electrical energy is available in every home at the touch of a switch.

Reading

The article 'The generation and transmission of electric power' in the *Students' guide* deals with these and related points.

WRIGHT, *The vital spark* is a good introduction to power generation and transmission.

EXPERIMENT

H20 Investigation of transformer action

ITEM NO.	ITEM
27	transformer to provide 6 and 12 V a.c.
1058	2 coils with 120 + 120 turns
92G	double C-core and clip
1507	2 ammeters, 5 A a.c.
177	2 s.b.c. lamps, 12 V, 6 W
74	2 s.b.c. lampholders on bases
504	retort stand rod, mild steel
541/1	rheostat, 10–15 Ω, 5 A
1511	oscilloscope
1000	leads

A single set of this equipment may be used to answer a number of questions.

H20a What is the effect of iron in the circuit?

When the steel rod is inserted in the coil, the lamp dims as the current decreases. When the iron circuit is closed with the two C-cores clipped together – figure H46(b) – the current falls to a low value and the lamp ceases to glow.

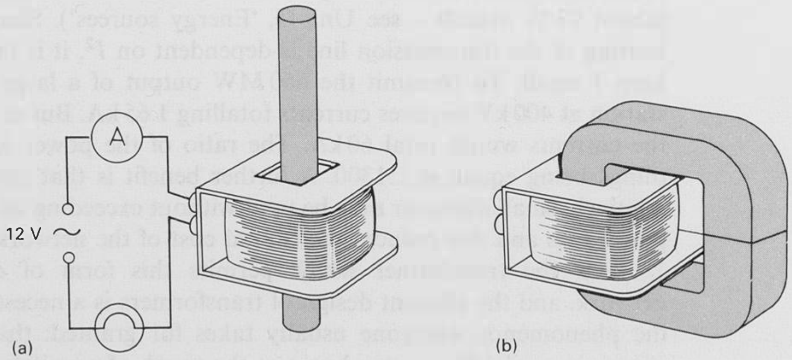


Figure H46

Discussion should bring out the facts that, firstly there is much more flux in the complete iron core and hence larger e.m.f.s are induced in the coil, and secondly the coil with the iron core in place has a large inductance. These are of course equivalent ways of saying the same thing.

H20b What is the effect in a second coil on the same core?

Students add a second 120 + 120 turn coil to the C-core assembly and use the oscilloscope to compare the p.d. applied to the first coil with the e.m.f. induced in the second using turns ratios of 240:240, 240:120, 120:240, 120:120. It will be found that the voltage ratios are very nearly equal to the turns ratios.

What happens to these voltages if a 12 V, 6 W lamp is connected across the second coil using the 240:240 turns ratio?

The device is now a transformer with a turns ratio of 1:1. It may be worth noting that such a transformer is useful for isolating a circuit supply from the mains supply.

Students return to the turns ratio of 120:240 and supply the primary coil of 120 turns with 6 V a.c. What happens to the lamp? This raises the question of what may be happening to the currents in the two coils and to the input and output power.

H20c How does the current in one coil depend on the current in the other?

Students use the rheostat to control the resistance in the secondary circuit and note corresponding values of the currents in both primary and secondary circuits. They will find that the current ratios are very nearly in the inverse ratio of the numbers of turns.

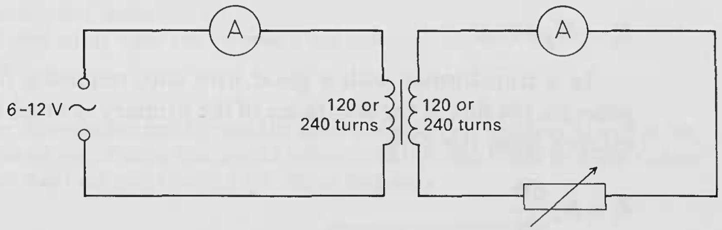


Figure H47

H20d How do the input and output powers compare? (*Optional extension*)

This investigation may well be left until after peak and r.m.s. values have been considered.

Students can use the same circuit as in experiment H20c to measure the input and output powers for a given turns ratio as the resistance of the secondary circuit is changed. The currents measured by the ammeters will be r.m.s. values. If an oscilloscope is used to measure the p.d.s across the primary and secondary coils, either peak or half-peak values will be measured. To find the r.m.s. value students will need to know that $V_{\text{r.m.s.}} = V_0/\sqrt{2}$. (See pages 78 and 79 of this *Teachers' guide*.)

Students should attempt to account for any discrepancy between input and output powers by calculating I^2R losses in the primary and secondary coils and any losses in the meters. There are also the so-called iron losses (due to eddy currents in the iron core and magnetic hysteresis). See demonstration H21.

Summary: The action of a transformer

An alternating e.m.f. is induced in the secondary because there is an alternating flux through it. The e.m.f. induced in one secondary turn is equal to $\mathrm{d}\Phi/\mathrm{d}t$. If the secondary has N_s turns in series, the induced e.m.f. $\mathscr{E}_\mathrm{s}$ is given by

$$\mathscr{E}_\mathrm{s} = N_\mathrm{s}\,\mathrm{d}\Phi/\mathrm{d}t$$

If the iron core is a complete ring, almost all the flux goes through both coils, and hardly any escapes round the outside. The p.d. across the primary drives a current through its resistance and maintains the changing magnetic flux through the coils ($V_\mathrm{p} = I_\mathrm{p}R + N_\mathrm{p}\,\mathrm{d}\Phi/\mathrm{d}t$). If the primary has many turns and if the iron core is both fat and short, a small current will produce a large flux. Almost all the applied p.d., V_p, maintains a changing flux $\mathrm{d}\Phi/\mathrm{d}t$ through N_p turns of the primary, and

$$V_\mathrm{p} \approx N_\mathrm{p}\,\mathrm{d}\Phi/\mathrm{d}t$$

In a transformer with a good iron core requiring little current to generate the flux, if the resistance of the primary is small and if little flux escapes from the core,

$$\mathscr{E}_\mathrm{s} = N_\mathrm{s}\frac{\mathrm{d}\Phi}{\mathrm{d}t}$$

$$\mathscr{E}_\mathrm{s} \approx N_\mathrm{s}\frac{V_\mathrm{p}}{N_\mathrm{p}}$$

$$\Rightarrow \frac{\mathscr{E}_\mathrm{s}}{V_\mathrm{p}} \approx \frac{N_\mathrm{s}}{N_\mathrm{p}}$$

The transformer itself contains no source of energy; indeed, some energy transformed by it must go to warming it up, because the coils have resistance and because the iron core becomes warm as its magnetization is switched to and fro. The electrical power out of the secondary cannot exceed that into the primary. If the currents are I_s and I_p, then

$$I_p V_p > I_s \mathscr{E}_s$$

and

$$I_s/I_p < N_p/N_s$$

For good transformers in which the losses are very modest

$$I_p V_p \approx I_s \mathscr{E}_s$$

and

$$\frac{I_s}{I_p} \approx \frac{N_p}{N_s}$$

Questions

Questions 42 to 47 are about transformers; question 48 is about the transmission of electric power.

DEMONSTRATION
H21 Eddy currents

ITEM NO.	ITEM
27	transformer to provide 12 V a.c.
147	demountable transformer kit
1508	demonstration meter, 5 A a.c.
504	mild steel retort stand rod (or even a flat file)
1000	leads

Some demountable transformer kits include a solid iron armature as well as the laminated one. Failing this, a solid armature can be improvised by using a short retort stand rod (mild steel), a flat file, or long nails.

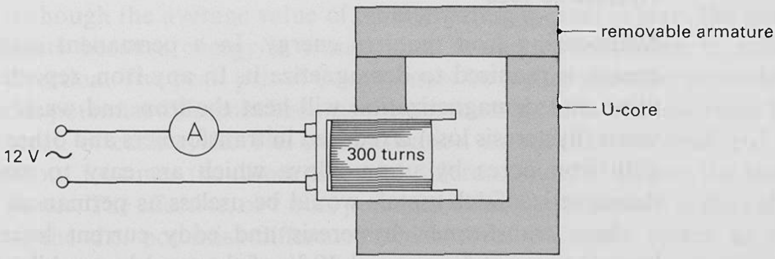


Figure H48
Eddy current heating.

With a laminated iron 'armature' closing the magnetic circuit (figure H48), the current through the coil will be found to be quite small. When the laminated armature is replaced by a solid iron one, the current will rise appreciably. After a minute or so the solid iron armature will be quite warm although the laminated U-core remains cool.

The alternating magnetic flux produced by the current in the coil induces an e.m.f. in any conductor with which it is linked. The iron core is such a conductor. It provides a series of conducting paths parallel to the turns in the coil, and currents flow in these conducting paths. These are 'eddy currents'; they have the effect of warming the core and so they waste energy. Constructing the core of laminations which are laid parallel to the flux and insulated from one another reduces the eddy currents and hence the energy losses.

Eddy current loss becomes more serious at higher frequencies. As we have seen, induced e.m.f.s are proportional to the frequency and, since the power loss is V^2/R, it follows that the power loss is proportional to the square of the frequency. At high radio frequencies, eddy current losses might be millions of times higher than at mains frequencies if ordinary transformer steel were used. Such losses are virtually eliminated by the use of ferrite rods. Ferrites are ceramic materials, the molecules of which contain iron. They have high resistivities.

Eddy currents can occur in any iron subject to a changing flux. Consequently the iron in both the rotors and the stators of motors and generators is of laminated construction.

The motor/generator or other commercial machine that was used in the discussion of reluctance and magnetic circuits can be used again to point out the lamination of the iron.

Hysteresis loss

Magnetizing iron requires energy. In a permanent magnet, further energy is required to demagnetize it. In any iron, repeated magnetization and demagnetization will heat the iron and waste energy. This waste (hysteresis loss) is reduced in transformers and other components with iron cores by using alloys which are easy to magnetize and demagnetize. Such metals would be useless as permanent magnets.

In a transformer, hysteresis and eddy current losses (the 'iron losses') account for around 30 % of the total loss, whilst the I^2R loss ('copper loss') accounts for about 70 %. The total loss is itself very small. A very good transformer may have an efficiency as high as 99 %.

ALTERNATING CURRENT IN A RESISTIVE CIRCUIT

DEMONSTRATION
H22 Power in a resistive circuit

This simple demonstration from REVISED NUFFIELD PHYSICS *Teachers' guide Year 5* shows clearly that r.m.s. values are used in a.c. measurements because they are related to the power transferred in resistors.

ITEM NO.	ITEM
1033	cell holder with three cells
27	transformer to provide 2 V a.c.
541/1	rheostat, 10–15 Ω, 5 A
1511	oscilloscope
52L	2 mounted bell pushes (or s.p.d.t. switch)
92R,T	m.e.s. lamp, 2.5 V, 0.3 A, in holder
1000	leads

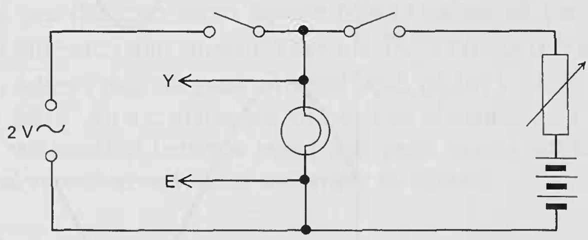


Figure H49
Comparing the brightness of a lamp lit from a.c. and d.c.

Connect the lamp alternately to the transformer and the battery and adjust the rheostat until the lamp is equally bright in both switch positions. Use the oscilloscope, set to d.c., to show the relative values of the steady p.d. and the amplitude of the alternating p.d.

Although the average value of an alternating current is zero, the lamp lights because energy is transformed when current flows in either direction. The peak p.d. of the a.c. must be higher than the equivalent d.c. potential difference because the alternating p.d. is at a low value for much of each cycle. The demonstration shows that the peak p.d. is about 1.4 times the direct potential difference needed to give the same power. The effective, or r.m.s., potential difference of the a.c. is the value of the d.c. potential difference which gives the same power as the alternating p.d.

Root mean square values

For a sinusoidal waveform applied to a resistor, the variations of p.d. and of current with time will have the form shown in figures H50(a) and (b). Since the power dissipated is given by the product of V and I, the graph of power against time will have the form shown in figure H50(c).

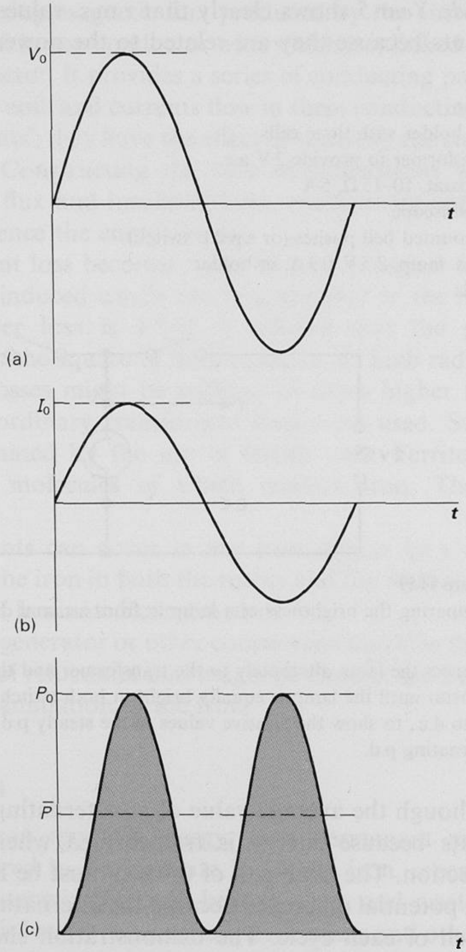


Figure H50
Variation of p.d., current, and power with time for alternating current in a resistor.

The symmetry of this last graph shows that the mean power dissipated, $\bar{P}$, is $\frac{1}{2}P_0$, where P_0 is the peak value.

But $P_0 = V_0 I_0 = V_0{}^2/R$, where V_0 and I_0 are the peak values of p.d. and current and R is the resistance. Thus

$$\bar{P} = \tfrac{1}{2}\frac{V_0{}^2}{R}$$

If a direct p.d. V is applied to the resistor, the power dissipated is V^2/R. For this to be the same as that dissipated by the sinusoidal p.d. we have

$$\frac{V^2}{R} = \tfrac{1}{2}\frac{V_0{}^2}{R}$$

thus $$V = \frac{V_0}{\sqrt{2}}$$

and similarly $$I = \frac{I_0}{\sqrt{2}}$$

These are the root mean square (r.m.s.) values of the sinusoidal potential difference and current. They are $1/\sqrt{2}$ (about 0.7) of the peak values. It follows that the peak value of a p.d. of 240 V a.c. is $\sqrt{2} \times 240$, or about 340 V. An a.c. mains p.d. of 240 V is, of course, an r.m.s. value.

This relationship between mean and peak values applies only to sinusoidal wave forms; it does not apply to others.

Question

Question 49 is about power in a resistive circuit.

Measuring a.c. with moving-coil instruments

Direct current meters are designed for use with completely smooth currents or p.d.s; a.c. instruments are designed for use with sinusoidal waveforms. Difficulties arise if the meters are used for other waveforms.

One is usually interested in the r.m.s. value of an alternating current or p.d., and most a.c. meters are calibrated to read such values. But the calibration does assume that the a.c. is sinusoidal. For other waveforms an a.c. meter will not give the true r.m.s. value and can only be used to make comparisons.

An a.c. attachment for a d.c. meter is also designed so that the meter reading is the r.m.s. value of a sinusoidally varying current or p.d.

A problem can arise if one uses a d.c. meter to measure unsmoothed d.c. The meter reading will indicate the average value of the current, $\bar{I}$. For an unsmoothed direct current derived directly from a full-wave rectifier, $\bar{I} = 2I_0/\pi = 0.637\, I_0$. Since, for a sinusoidal wave form, $I_{\text{r.m.s.}} = 0.707\, I_0$, $I_{\text{r.m.s.}} = 1.11\, \bar{I}$ in this case. The r.m.s. value of this current is therefore greater than the meter reading by a factor of 1.11. And the mean square current, $\overline{I^2}$ – which is what one needs to calculate power for example, or in a current balance experiment where force depends on $(\text{current})^2$ – is 1.23 times larger than $(\bar{I})^2$.

The moral clearly is: use *smoothed* d.c.

CAPACITORS IN A.C. CIRCUITS

Capacitors are dealt with next so that they can be given a relatively full treatment before the behaviour of inductors in a.c. circuits is seen to be, in some degree, the 'opposite to that of capacitors'.

DEMONSTRATION

H23 Slow a.c. in a circuit containing a capacitor

Apart from their application in blocking direct currents, capacitors are frequently used in a.c. circuits because their reactance decreases with increasing frequency. This is examined here, as is the fact that a capacitor dissipates no power. It is easier to see what is happening if very low frequency a.c. is used.

ITEM NO.	ITEM
170	low-frequency a.c. generator
1511	double-beam oscilloscope
1033	cell holder with four cells
1508	demonstration meter, 2.5–0–2.5 mA
1508	demonstration meter, 5 V centre-zero
1017	resistance substitution box
1040	2 clip component holders
1151	2 capacitors, 1000 μF
1151	resistor, 15 Ω
1151	resistor, 1000 Ω
30	slotted base
1000	leads

First show the action of the generator so that the direction of the output p.d. can be related to the position of the drive wheel. Then connect a 1000 Ω resistor and milliammeter in series with the low-frequency a.c. generator and connect a voltmeter across the resistor to display the changing current and p.d.; they are in phase.

Next replace the resistor by a 500 μF capacitor. It is best, with electrolytic capacitors, to use two 1000 μF capacitors in series and back-to-back. The existence of a phase difference between the current in the circuit and the p.d. across the capacitor will be shown by the behaviour of the two meters. As the frequency of the a.c. is increased, the response of the meters becomes less and less helpful. At this point connect a 15 Ω resistor in series with the capacitor and make the connections to the C.R.O. as in figure H51. The Y_2 input is now determined by the current through the 15 Ω resistor and the Y_1 input by the p.d. across the capacitor and the resistor in series with it. That the p.d. across the latter is small may be demonstrated by connecting Y_1 to the 15 Ω resistor.

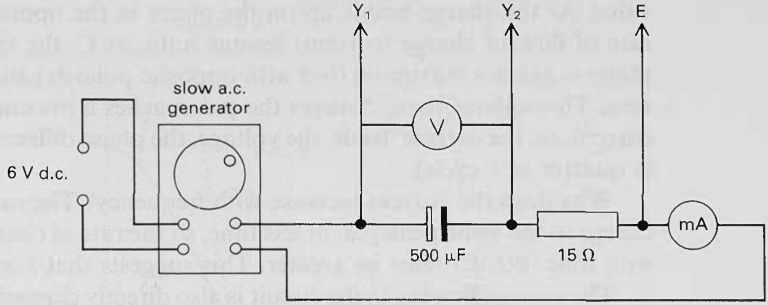


Figure H51

Using meters and the resistor it is clear that the current is in phase with the p.d., and that the meters give smaller and smaller deflections as the frequency is increased. With the capacitor, the current leads the p.d. and increases as the frequency rises but, as before, the meters soon fail to respond. This effect becomes clear when the oscilloscope is used. The fact that current flows in a circuit containing a capacitor, although the capacitor has good insulation between its plates, can now be linked with the charging and discharging of a capacitor considered in Unit B, 'Currents, circuits, and charge'.

Why is there a phase difference? Consider the graphs in figure H52.

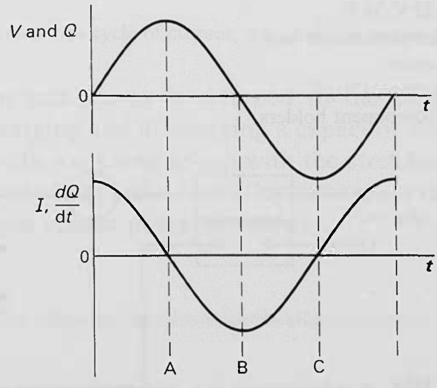


Figure H52
Potential difference and current in a capacitor.

At instant A on the time scale of the graphs, the p.d. across the capacitor, and hence the charge on the capacitor, is a maximum. At this instant, charge has ceased to flow onto the plates and is about to flow from them. Between A and B the p.d. and the charge are decreasing. At B the graph of p.d. against time has its greatest negative slope. At this instant the current (or rate of flow of charge) has its greatest negative value. As the charge builds up on the plates in the opposite sense the rate of flow of charge (current) lessens until, at C, the charge on the plates is again a maximum (but with opposite polarity) and the current zero. Throughout these changes the p.d. reaches a maximum *after* the current, *i.e.* the current 'leads' the voltage, the phase difference being $\pi/2$ (a quarter of a cycle).

Why does the current increase with frequency? The capacitor must charge to the same peak p.d. in less time, so the rate of change of charge with time ($\mathrm{d}Q/\mathrm{d}t$) must be greater. This suggests that $I \propto f$.

The current flowing in the circuit is also directly dependent upon the capacitance C, since $Q \propto C$.

Combining the two, we have $I \propto fC$.

DEMONSTRATION
H24 Power in a capacitor

This demonstration shows that there is no net transfer of energy in a capacitor.

ITEM NO.	ITEM
1518	joulemeter
1508	demonstration meter, 15 V a.c.
1508	demonstration meter, 5 A a.c.
73	lamp, 12 V, 36 W
74	s.b.c. lampholder on base
27	transformer
1151	2 capacitors, 100 μF
1040	2 clip component holders
1000	leads

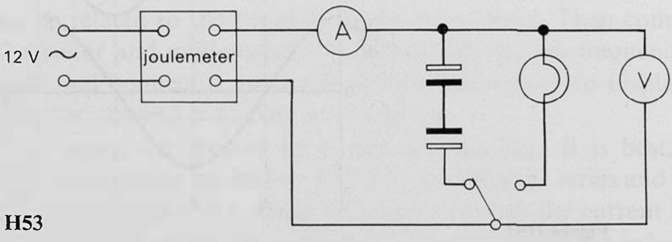


Figure H53

With the lamp as load the joulemeter reading should be approximately equal to the product VI as measured on the two meters. With the capacitors substituted for the lamp, V and I will still be large but the joulemeter reading will be very near to zero. Why is this?

The graphs (figure H54) provide the answer. Between A and C the capacitor is charging from zero to maximum; both I and V are positive and energy is taken from the supply. As V falls to zero (between C and

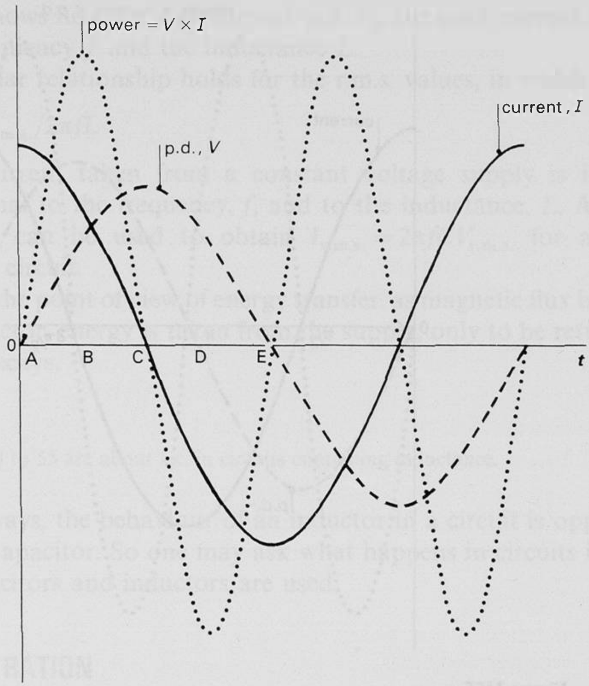


Figure H54
Variation with time over one cycle of current, p.d., and power in a capacitor circuit.

E), I is negative and energy is returned to the supply. This is not different from charging and discharging a capacitor with d.c.; it merely seems stranger with a.c. Comparison with the stretching and releasing of a steel spring may help and it is worth pointing out that both systems can store and then release potential energy.

Questions
Questions 50 to 52 are about a.c. in circuits containing capacitors.

ALTERNATING CURRENT IN AN INDUCTOR

In demonstration H19, it was observed that in an a.c. circuit containing an inductor there is a phase difference between the current and the p.d., with the current lagging. Because any real inductor always has resistance as well as inductance, this phase difference is somewhat less than the one-quarter cycle ($\pi/2$) associated with a pure inductor (see figure H55).

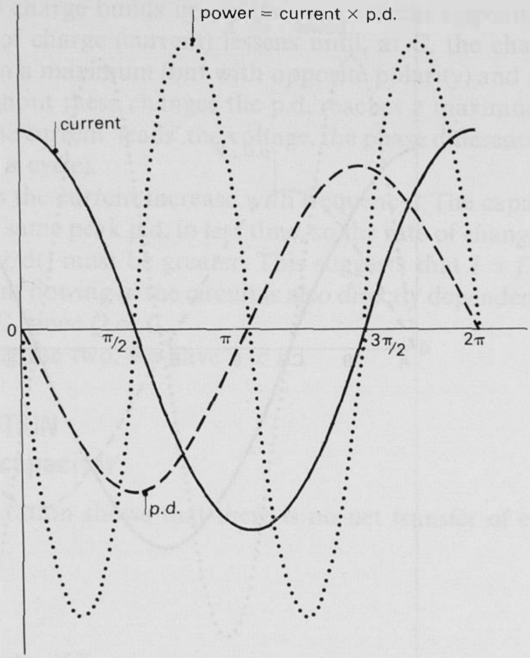


Figure H55
Variation with time over one cycle of current, p.d., and power in an ideal inductor.

Students should have some understanding of the effect of the frequency of the a.c. on the current flowing in a circuit with inductance. The *Students' guide* (pages 13 and 14) gives the outline of the argument for $I \propto 1/fL$. Some students at least may like to see a more complete version of the mathematics.

The p.d. across an inductor required to maintain a changing flux through its coils is given by

$$V = L\,\mathrm{d}I/\mathrm{d}t$$

For sinusoidal a.c.,

$$I = I_0 \sin 2\pi ft$$

where f is the frequency. Hence

$$V = LI_0\, 2\pi f \cos 2\pi ft$$

The maximum p.d. is $V_0 = 2\pi fLI_0$

and the maximum current is $I_0 = V_0/2\pi fL$

This shows how, for a given peak p.d. V_0, the peak current depends on the frequency f and the inductance L.

A similar relationship holds for the r.m.s. values, in which case

$$I_{\mathrm{r.m.s.}} = V_{\mathrm{r.m.s.}}/2\pi fL$$

The current taken from a constant voltage supply is inversely proportional to the frequency, f, and to the inductance, L. A similar derivation can be used to obtain $I_{\mathrm{r.m.s.}} = 2\pi fCV_{\mathrm{r.m.s.}}$ for a purely capacitive circuit.

From the point of view of energy transfer, as magnetic flux builds up in an inductor, energy is taken from the supply, only to be returned as the flux decays.

Questions

Questions 53 to 55 are about a.c. in circuits containing inductance.

In some ways, the behaviour of an inductor in a circuit is opposite to that of a capacitor. So one may ask what happens in circuits in which both capacitors and inductors are used.

DEMONSTRATION

H25 Alternating currents in a circuit containing capacitance and inductance

This demonstration presents a pretty puzzle about two alternating currents adding up to a total current which is less than either.

ITEM NO.	ITEM
1109	signal generator
1081	decade capacitance box, 1 to 10 μF
1058	120 + 120 turn coil
52A	3 matched m.e.s. lamps, 1.25 V, 0.25 A
92T	3 m.e.s. lampholders
1151	2 resistors, 3.9 Ω
1041	2 clip component holders
	either
1507	3 ammeters, 1 A a.c.
	or
1005	1 multirange meter
1511	double-beam oscilloscope
1000	leads

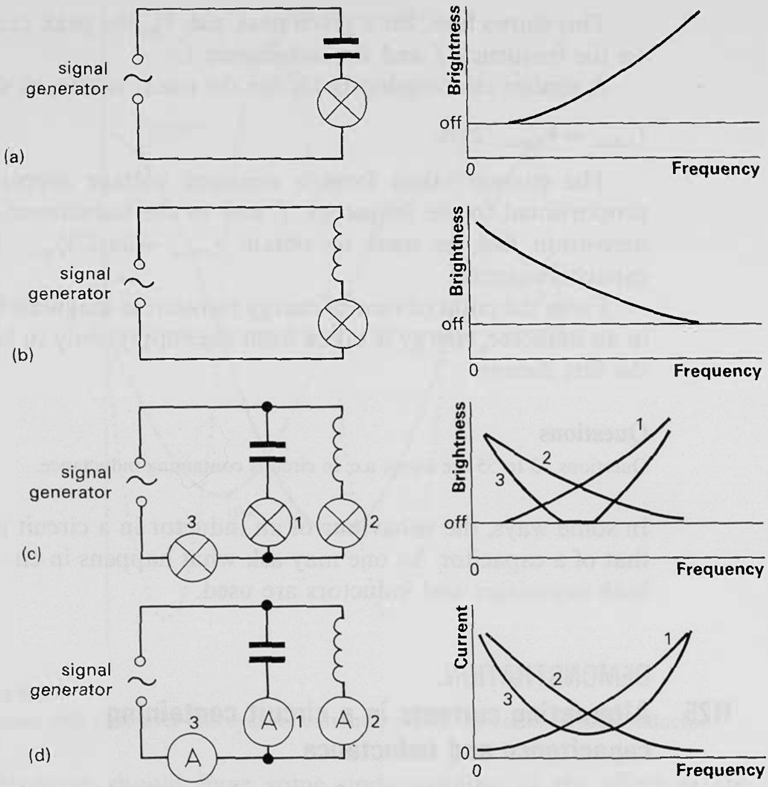


Figure H56
Currents in an *LC* circuit.

It is best to prepare the demonstration in advance and to set up the circuit of figure H56(c), choosing the best capacitor (try 2 to 6 μF) and inductor (try 240 turns) to give resonance in the middle of a range of the signal generator using the low-impedance output. The aim is that the lamp in series with the capacitor should go out at the lowest frequency of the range, and that the lamp in series with the inductor should go out at the top of the range. It may be necessary to connect an oscilloscope across the output of the signal generator to check that the output voltage remains constant. Do not alter the output voltage of the signal generator from now on.

In the demonstration, set up the circuit of figure H56(a) and show that the current increases with frequency. Then try circuit (b) and show that the current falls as the frequency rises. Now connect circuit (c) but with a wire short-circuiting lamp 3. Choose a frequency at which both lamps are equally bright. When the shorting wire is removed lamp 3 might be expected to be 'twice as bright' – but it does not glow at all. The currents do not add up as expected.

Finally replace the lamps by ammeters as in figure H56(d); if the readings are too low, use a multirange meter at points 1, 2, and 3 in turn.

These observations can be explained by recalling that the current in a capacitor leads the p.d. by $\pi/2$, while the current in an inductor lags by $\pi/2$. (The word 'CIVIL' is useful for remembering this: C has I before V; V before I in L.) The two currents in the parallel circuit are in opposition and the resultant current in the main circuit is the difference, not the sum. In fact, the resistances of the lamps and the coil cause the currents to be out of phase by less than π and there is some resultant current. It may be better to display the currents using a double-beam oscilloscope.

Display the currents through the same capacitor and inductor by inserting two low-value resistors (figure H57) and connecting these to the inputs of the oscilloscope as shown. The total current may be examined by connecting a third resistor in the main circuit.

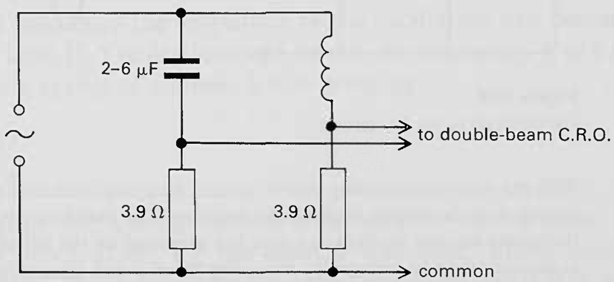


Figure H57
Phase differences in an LC circuit.

As the frequency is altered, the p.d.s across the two resistors, which are proportional to the currents, will vary in amplitude but not in phase difference. The currents will become almost equal and opposite at a frequency which is known as the resonant frequency; at this frequency the total current in the main circuit will be a minimum.

Students may find it curious that resonance, usually associated with maximum amplitudes of oscillation, should give this minimum current. In fact, the oscillating current which is 'sloshing around' the capacitor and inductor circuit is a maximum, while the current taken from the supply to maintain the oscillation is a minimum at resonance.

EXPERIMENT
H26 Oscillations in a parallel *LC* circuit

ITEM NO.	ITEM
1511	capacitors, 550 μF, 250 μF, 100 μF, 50 μF
1030	high-inductance coil
92G	double C-core and clip
1033	cell holder with four cells
1151	potentiometer, 1 kΩ (100 Ω if available)
1511	oscilloscope
	s.p.d.t. switch (if available)
1000	leads

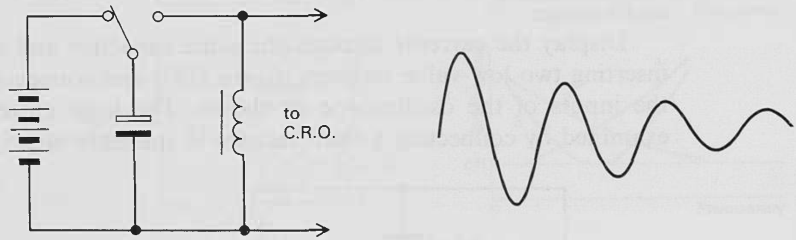


Figure H58
Oscillations in an *LC* circuit.

With the time-base running slowly, several decaying oscillations will be seen when the capacitor is discharged through the inductor. The switchover is best made just after the slowly moving oscilloscope spot has appeared on the left of the screen. The inductance may be reduced by removing the clip and separating the C-cores. The capacitance is easily altered by using other capacitors.

It is worth trying to note the change in the period of oscillation when C is reduced from 500 μF to 100 μF. A shade more than twice as many oscillations may now occupy the same time as previously, since C has changed by a factor of rather more than four and the frequency is inversely proportional to $\sqrt{C}$. The tolerances on electrolytic capacitors make a more precise test hardly worth while.

Students can observe the damping effect of resistance by adding a potentiometer (preferably 100 Ω) in series with the inductor to the *LC* circuit. Increasing the resistance in an electrical circuit makes the oscillations die away more quickly, just as increasing the friction in a mechanical oscillator does. Resistances of 50 to 100 Ω should have a marked effect.

Questions

Questions 56 to 59 are about electrical oscillations.

Mechanical analogue of an LC circuit

There is a sense in which we may say that a capacitor is like a spring and an inductor like a mass.

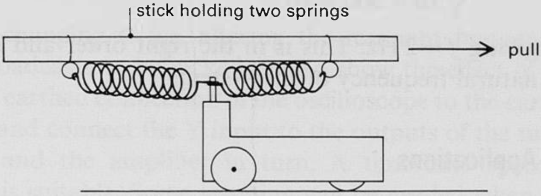


Figure H59
Mechanical analogue of an LC circuit.

A trolley joined by two similar springs to a metre stick, as in figure H59, will give damped oscillations if set moving. With a gentle motion of the metre stick the oscillations can be maintained. At the resonant frequency the amplitude of the oscillation can become quite large. In Unit D, 'Oscillations and waves', the frequency, f, of a mass, m, joined to a spring of stiffness k was given by

$$2\pi f = \sqrt{\frac{k}{m}}$$

Consider the analogy between an inductor and a mass. For the inductor $V = L\,\mathrm{d}I/\mathrm{d}t$; for the mass $F = m\,\mathrm{d}v/\mathrm{d}t$, where $\mathrm{d}v/\mathrm{d}t$ is the acceleration. These equations have the same form and it is in this sense that we may say that inductance and mass are analogous.

We have seen in Unit B, 'Currents, circuits, and charge', that there is a sense in which capacitance and spring stiffness are analogous. For a capacitor $V = (1/C)Q$ and for a spring $F = kx$ where F is the force, k the spring constant, and x the displacement. These equations have the same form and we may think of $1/C$ as the analogue of k.

If we assume that L corresponds to m and $1/C$ to k, the frequency of an LC circuit, by analogy with the equation for the frequency of the mass controlled by a spring, might be given by

$$2\pi f = \sqrt{\frac{1/C}{L}} = \sqrt{\frac{1}{LC}}$$

Is this right? The inductance of the coil used in experiment H26 is about 10 H. With $C = 500\,\mu\text{F}$

$$2\pi f = \sqrt{\frac{1}{10 \times 500 \times 10^{-6}}}$$

giving $f = 2$ Hz. This is in the right order and is roughly equal to the natural frequency of the system.

Applications

Circuits based on electrical resonance find their application over a very wide field – in oscillators and signal generators, in radio tuners and transmitters, in radar, in aerials, in filtering wanted from unwanted signals, and so on.

DEMONSTRATION
H27 A simple radio

ITEM NO.	ITEM
1151	tuning capacitor, 365 pF or 500 pF maximum
1037	set of solenoids (or other coil)
1151	diode
181	general-purpose amplifier
183	loudspeaker (if not in 181)
1511	oscilloscope
1501	reel of 0.45 mm PVC-covered wire (for aerial)
1000	leads

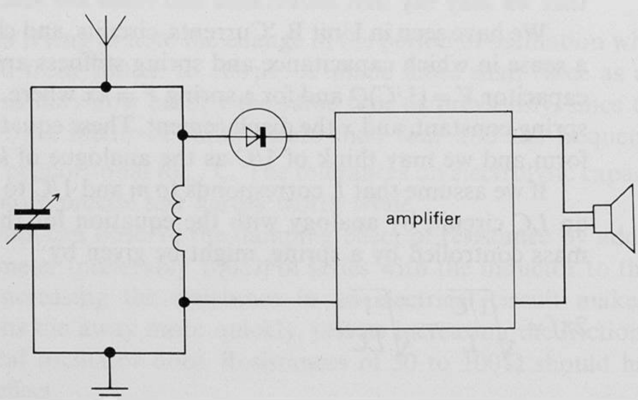


Figure H60
Simple radio circuit.

For long-wave transmissions the closely wound solenoids from item 1037 have suitable inductance. For medium-wave transmission a coil of 40 to 50 turns of 0.56 mm enamelled wire wound on a card or plastic tube about 70 mm in diameter will serve. The tuned circuit should be connected to as long and as high an aerial as possible and to a good earth such as a metal water supply pipe.

Show that changing C (*i.e.* altering the resonant frequency) enables different broadcasts to be picked up. To show the effect of each stage, connect the earthed connection of the oscilloscope to the earthed side of the circuit, and connect the Y input to the outputs of the tuned circuit, the diode, and the amplifier in turn. A time-base speed of about 1 ms cm^{-1} is suitable. Some retuning will be needed when the oscilloscope connection is moved from the output of the tuning circuit to the output of the diode.

The tuning capacitor can be improvised quite effectively from two sheets of cooking foil (about 20 × 20 cm) placed between the pages of a book. A calculation of $C = \varepsilon_0 A/d$ (which may be useful revision) shows that the capacitance of this arrangement is in the right order of magnitude. The value of C can be varied by changing the area of overlap as in a tuning capacitor, or by opening the book, thus changing d. (Note that with this arrangement stray capacitance may play a big role.)

Other examples

It is worth pointing out that when an oscilloscope is switched to a.c., a capacitor (typically about 0.1 μF) is connected in series with the input, so that only a.c. signals are passed to the oscilloscope, the capacitor blocking the passage of d.c.

Another example from the laboratory concerns the smoothing of rectified a.c. (see figure H61).

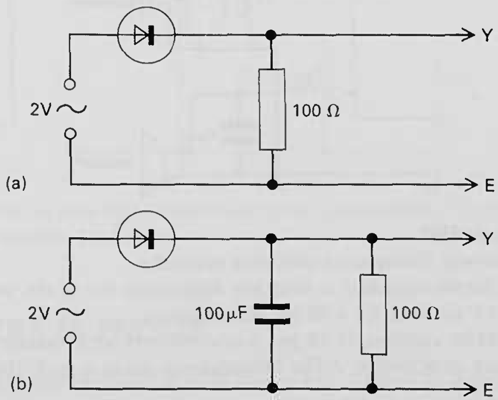


Figure H61 (part)
Smoothing rectified a.c.

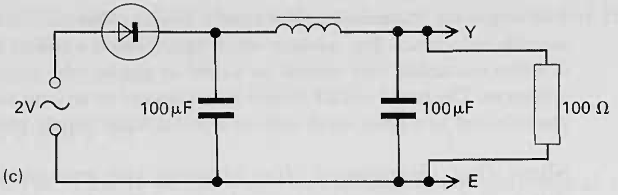


Figure H61 (part)
Smoothing rectified a.c.

Examples of filter circuits used in hi-fi systems are shown in figure H62. Students should be able to explain their action. Similar circuits can be set up in the laboratory; see figures H63 to H65.

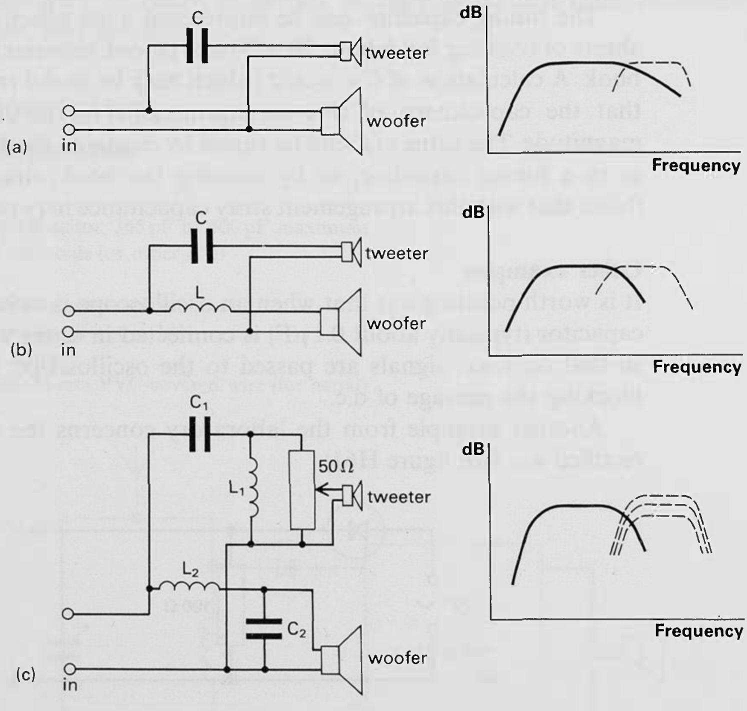


Figure H62
Two-way loudspeaker crossover networks.
(a) Simple capacitor to keep low frequencies out of the 'tweeter'.
(b) *LC* network for 6 dB per octave slopes.
(c) More complex 12 dB per octave network with 'tweeter' attenuator.
Based on BORWICK, J. The Gramophone guide to hi-fi. *David & Charles, 1982.*

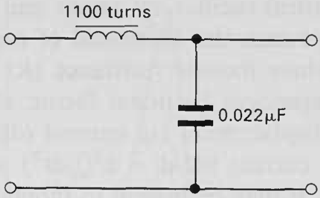


Figure H63
A low pass filter to reduce high-pitched scratching sounds, for example, from a worn disc.

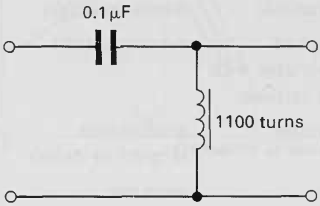


Figure H64
A high pass filter to reduce mains hum and motor rumble from a record player.

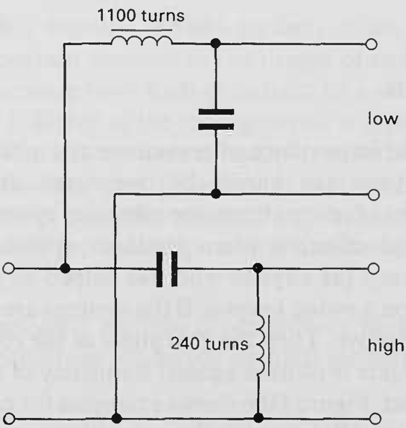


Figure H65
A crossover filter to pass high frequencies to one loudspeaker ('tweeter') and low frequencies to another ('woofer').

SIMILARITIES IN PHYSICS

The examination of oscillations in an *LC* circuit is a particularly good example of the use of analogues or similarities in physics. This example can be taken even further. We observed an analogy between inductance in an electrical oscillatory circuit and mass in a mechanical one, and another between the reciprocal of capacitance and the stiffness of a spring. Others include resistance (R) as the electrical analogue of a velocity-dependent frictional factor; charge (Q) as the electrical analogue of displacement (s); current ($\mathrm{d}Q/\mathrm{d}t$) of velocity ($\mathrm{d}s/\mathrm{d}t$); rate of change of current ($\mathrm{d}I/\mathrm{d}t = \mathrm{d}^2Q/\mathrm{d}t^2$) of acceleration ($\mathrm{d}v/\mathrm{d}t = \mathrm{d}^2s/\mathrm{d}t^2$); and so on. It may be helpful to tabulate some of these, as in table H1.

Characteristic	**Mechanical oscillator**	**Electrical oscillator**
independent variable	time (t)	time (t)
dependent variable	displacement (s)	charge (Q)
rate of change of dependent variable with independent variable	velocity ($\mathrm{d}s/\mathrm{d}t$)	current ($\mathrm{d}Q/\mathrm{d}t$)
second derivative	acceleration ($\mathrm{d}^2s/\mathrm{d}t^2$ or $\mathrm{d}v/\mathrm{d}t$)	rate of change of current ($\mathrm{d}^2Q/\mathrm{d}t^2$ or $\mathrm{d}I/\mathrm{d}t$)
inertia	mass (m)	inductance (L)
stiffness	stiffness (k)	(capacitance)$^{-1}$
resistance	frictional factor	resistance (R)
resonant frequency	$1/2\pi\sqrt{k/m}$	$1/2\pi\sqrt{1/LC}$
period	$2\pi\sqrt{m/k}$	$2\pi\sqrt{LC}$

Table H1

The importance of resonance as a phenomenon in engineering and in physics can hardly be overstated. As an example, consider the transfer of energy from one vibrating system to another. Energy transfer is most effective when the two systems share the same resonant frequency (as anyone who has helped to build up the oscillations of a child on a swing knows). If the systems are damped, the transfer is much less effective. Then the sharpness of the resonance peak obtained when amplitude is plotted against frequency of the driving oscillator is much reduced. Figure H66 shows examples for cases (A) in which there is little damping, (B) in which there is rather more damping, and (C) in which the damping is quite heavy.

The sharpness of the resonance peak is described by its Q-factor (originally a quality factor), mentioned already in Unit D, ‘Oscillations and waves’. This can be defined as the ratio of the resonant frequency to the width of the resonance peak. The width of the peak is taken to be the difference in frequency between points at which (amplitude)^2 is half (resonant amplitude)^2.

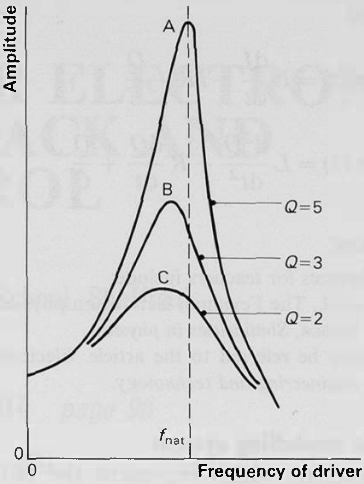


Figure H66
Resonance curves.

The three curves shown in figure H66 have Q-factors of 5, 3, and 2 respectively.

The Q-factor is a measure of the selectivity of the resonant system. This is a very important concept in the design of radio and television receivers: good receivers have high Q-factors. In a medium wave-band radio receiver the Q-factor of the tuning circuit is typically in the order of a few hundred. In u.h.f. receivers as used, for example, in some radars, the resonant cavities may give Q-factors as high as 30 000.

At the other end of the scale, a very low Q-factor, *i.e.* a great deal of damping, is required for buildings which are designed to withstand earthquake shock.

More mathematically inclined students may be shown the formal similarity of the equations for forced oscillations with damping:

Mechanical

$$m\frac{\mathrm{d}^2s}{\mathrm{d}t^2} = F_0\sin(2\pi ft) - r\frac{\mathrm{d}s}{\mathrm{d}t} - ks$$

$$F_0\sin(2\pi ft) = m\frac{\mathrm{d}^2s}{\mathrm{d}t^2} + r\frac{\mathrm{d}s}{\mathrm{d}t} + ks$$

in which F_0 is a force and r is a drag coefficient.

Electrical

$$V_0 \sin(2\pi f t) = L\frac{\mathrm{d}I}{\mathrm{d}t} + IR + \frac{Q}{C}$$

$$V_0 \sin(2\pi f t) = L\frac{\mathrm{d}^2Q}{\mathrm{d}t^2} + R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C}$$

References

Useful references for teachers include:

FEYNMAN, *et al.*, The Feynman lectures on physics *Volume 1*, Chapters 23–5.

SHIVE and WEBER, *Similarities in physics.*

Students may be referred to the article 'Electromechanical similarities' in the Reader *Physics in engineering and technology.*

Dynamic modelling system

The equations which represent the behaviour of *LC*, *LR*, and *LRC* circuits can be written in the form of BASIC statements. For example, experiment H26, Oscillations in a parallel *LC* circuit, is modelled by the following set of equations:

```
VC = Q/C        (p.d. across capacitor)
VR = I*R        (p.d. across resistor)
VL = -VC-VR     (p.d. across inductor)
DI = VL/L*DT    (change in current, from V = L dI/dt)
I = I+DI        (new current)
DQ = I*DT       (change in charge on capacitor)
Q = Q+DQ        (new charge)
T = T+DT        (one time increment)
```

Useful values for the components are:

C = 0.0001; R = 10; L = 10 (where the units are F, Ω, and H respectively);

Q = 0.001 (coulomb, corresponding to an initial p.d. of 10 V across the capacitor);

For initial values:

I = 0 and T = 0 (no current when the switch is closed at T = 0).

DT = 0.01 (second).

These values will give slowly decaying oscillations (plot *I* against *T*) with a frequency of about 5 Hz. The dynamic modelling system makes it easy to explore the effect of varying the values of *L*, *R*, and *C*.

Similar sets of equations can be written to model, for example, the behaviour of an *LRC* circuit to which a sinusoidally varying voltage is applied.

Unit I
LINEAR ELECTRONICS, FEEDBACK AND CONTROL

Wilf Mace
King Edward VII School, Sheffield

Suggested time allocation: three weeks

PLAN OF THE UNIT

Section I1
Basic operational amplifier circuits

Unit C 'Digital electronic systems'	►	**Introduction: linear electronics and control**
current, p.d., resistance, (Unit B)	►	**Simple inverting amplifier circuit behaviour: explanation in terms of characteristics of operational amplifier, negative feedback**
		Summing amplifier, subtraction with inverter
charge, capacitance, (Unit B)	►	**Capacitive feedback, integration**
		Non-inverting amplifier circuit; unity-gain follower, input and output impedance, buffer
		Differential amplifier, comparator

Section I2
More feedback; control

exponential decay (Units B and F)

▶ **Solving $dV/dt = \pm kV$; analogue computing**

oscillation (Unit D)

▶ **Solving $d^2V/dt^2 = -\omega^2 V$: quadrature oscillator**

Oscillation in feedback loops in general: delayed feedback and phase shift

Control systems: open loop, closed loop; on–off and continuous control

Oscillation in continuous control systems, roles of inertia and damping

Feedback and control in biology, economics, etc.

Section I3
Putting electronics to use

Units B, 'Currents, circuits, and charge', F, 'Radioactivity and the nuclear atom', and H, 'Magnetic fields and a.c.'

▶ **Student practical construction tasks on instrumentation, communications, analogue computing, warning and automatic control devices**

PLAN OF THE UNIT

Section I1
Basic operational amplifier circuits

Prerequisites		Topic
Unit C 'Digital electronic systems'	►	**Introduction: linear electronics and control**
current, p.d., resistance, (Unit B)	►	**Simple inverting amplifier circuit behaviour: explanation in terms of characteristics of operational amplifier, negative feedback**
		Summing amplifier, subtraction with inverter
charge, capacitance, (Unit B)	►	**Capacitive feedback, integration**
		Non-inverting amplifier circuit; unity-gain follower, input and output impedance, buffer
		Differential amplifier, comparator

Section I2
More feedback; control

exponential decay (Units B and F)	►	**Solving $dV/dt = \pm kV$; analogue computing**
oscillation (Unit D)	►	**Solving $d^2V/dt^2 = -\omega^2 V$: quadrature oscillator**

Oscillation in feedback loops in general: delayed feedback and phase shift

Control systems: open loop, closed loop; on–off and continuous control

Oscillation in continuous control systems, roles of inertia and damping

Feedback and control in biology, economics, etc.

Section I3
Putting electronics to use

Units B, 'Currents, circuits, and charge', F, 'Radioactivity and the nuclear atom', and H, 'Magnetic fields and a.c.'	►	**Student practical construction tasks on instrumentation, communications, analogue computing, warning and automatic control devices**

INTRODUCTION

Unit C, 'Digital electronic systems', dealt with systems in which only two states are possible (the output is high or low, 1 or 0). This Unit takes a second look at electronics (in this case linear systems, *i.e.* ones in which the output can vary continuously over a range of values), and introduces some powerful and general ideas concerned with feedback and control. Many linear electronic circuits involve feedback and many control systems involve electronics. Some of the important concepts dealt with are amplification, input and output resistance, positive and negative feedback, limiting values, oscillation, open and closed loop systems, and on–off and continuous control.

An initial quick demonstration or two in which electronics is used to control position or light level can introduce some of the ideas which are developed later in the Unit.

Much of the work of the Unit uses an operational amplifier in different circuits. A series of quick experiments establishes the use of the inverting input in a feedback amplifier circuit to invert and amplify, and shows the existence of limiting values; it will also give some confidence in this kind of work. These early results can be consolidated by demonstration and by an experiment to plot the input–output voltage characteristic of the circuit.

Part of the manufacturer's specification can be used to start a discussion in which the essentials of the behaviour of the operational amplifier itself are established. From these (simplified) essentials (infinite input impedance, zero differential input voltage, very high amplification) it follows that the behaviour of a circuit depends on the values of the components (resistors, capacitors) connected around the device, not on the characteristics of the device itself.

Another series of experiments shows that voltage amplification depends on the ratio of feedback to input resistance ($V_{\text{out}}/V_{\text{in}} = -R_{\text{f}}/R_{\text{in}}$); that voltages can be summed (leading to the possibility of analogue-to-digital conversion), and subtracted; that if feedback is via a capacitor then the output is the integral of the input. This work provides many opportunities to use elementary circuit ideas involving resistors, potential dividers, capacitors, and so on.

The non-inverting input is used for the first time, leading to voltage follower (buffer), differential amplifier, and comparator circuits, useful in measurement and control systems.

In Section I2 the emphasis is on feedback and control. A feedback wire is used to make the input of an integrator equal to its output so that $\mathrm{d}V/\mathrm{d}t = -kV$, the familiar exponential decay; two integrators and one inverter can be coupled to produce oscillation. Again, many ideas met earlier in a different context could be revised. The idea that negative feedback leads to stability, and positive (delayed) feedback to oscillation is now generalized to a variety of other systems including mechanical and thermal. Discussion and demonstrations (including some seen at the start of the Unit) lead to the ideas of open- and closed-loop systems, on–off and continuous control, the effect of inertia and delay, and the possibility of oscillation. The wide scope of control theory can be brought out by illustrations from other fields such as biology and economics.

In the final part of the Unit, 'Putting electronics to use', students use the principles and techniques discussed earlier to propose and test solutions to a wide variety of problems.

Because in this Unit more than most it is likely that students will progress at widely differing rates, a number of extension experiments and optional questions have been included. Eighteen learning questions are provided, and these cover virtually all of the theory. References to these questions are given at the beginning of each relevant section.

Apparatus listing

With the exception of demonstration 1a, every experiment and demonstration in this Unit requires at least one operational amplifier unit and its power supply. The operational amplifier units which schools are likely to be using include a number of built-in resistors, at least one capacitor, and perhaps two potentiometers, and means of connecting them easily in the most commonly used configurations. However, the choice of components provided by different manufacturers will not necessarily be identical, and it is also possible that some schools would prefer to use kits which they have made up themselves.

In writing the apparatus lists for each of the experiments and demonstrations in this Unit we have endeavoured to include *all* resistors, capacitors, and potentiometers needed – including any that may be on the operational amplifier unit itself. Nevertheless, access to a selection of resistors (usually in the range 1 kΩ to 1 MΩ) and capacitors (0.001 μF to 1 μF) and some means of connecting them to the circuit (*e.g.* clip component holders and leads) will be useful throughout the Unit. The resistance of the potentiometer(s) used to control the input(s) is not critical. We have listed 'potentiometer (*e.g.* 1 kΩ)' because such potentiometers are used elsewhere in the course. Manufacturers' units may incorporate 10 kΩ potentiometers, which have the advantage of lower current consumption.

Note One extra item of equipment, not included in the *Apparatus guide* lists, will be required for the penultimate demonstration I14 (page 161); a bead thermistor, *e.g.* GL23 (2 kΩ–115 Ω). The thermistors normally provided under Item number 1151 are *not* suitable for this demonstration.

Level of treatment; style of presentation

The range of teachers' (and students') knowledge and experience is probably greater in electronics than in any other part of this course. For some, the ideas presented in this Unit will be very familiar; others will be meeting them for the first time. This *Teachers' guide* and the corresponding *Students' guide* have been written with the latter group in mind. We hope it will be possible for beginners to make sense of what is written here with a minimum need to seek clarification elsewhere. (Nevertheless, a teacher who has never met operational amplifiers before would do well to consult one of the introductory books listed in the references on page 415. Those by Foxcroft and by Plant can be particularly recommended for clear expositions from the start.)

We have tried to resist the temptation to let this Unit grow, and have indicated in this *Guide* what we believe is the minimum for all students. To try to cater for those with prior knowledge and experience of these matters we have suggested a number of extension and buffer experiments. Plenty of more advanced ideas can be found in books such as those by Foxcroft or Plant. Fast students should also be encouraged to tackle the more difficult questions in the *Students' guide*, and there should be considerable opportunity for slower students to be helped by more confident ones.

Timing

Experience has shown that in teaching this kind of material it is important to keep the pace brisk and not to explore all the possibilities which arise in each demonstration and experiment. The time allocation for this Unit has to be rather short, and it is important to allow a sufficient number of periods (at least six) for Section I3, 'Putting electronics to use'. Several experiments are shown as optional, and we suggest that these be omitted until experience has shown that time for them really will be available. Teachers might also consider how far they can use the various learning questions to minimize class time spent on exposition.

THE PLACE OF THE UNIT IN THE COURSE

This Unit relies heavily on basic ideas of current electricity from Unit B, 'Currents, circuits, and charge', and on the notion of a module having input and output which was developed in Unit C, 'Digital electronic systems'. Understanding of capacitors, from Unit B, is needed for integrators and differentiators. Oscillations, first met in Unit D, 'Oscillations and waves', are produced electronically and shown to be a solution of the equation $d^2V/dt^2 = -\omega^2 V$.

We believe that the ideas in this Unit are more sophisticated than those in Unit C, which argues for teaching it fairly late in the course. With such placing it can also be exploited to revise basic ideas of current electricity.

It seems sensible to defer it until after Unit H, 'Magnetic fields and a.c.', since there are a few places where students will need to use ideas about a.c., and some useful applications and illustrations depend on a knowledge of flux and induced e.m.f.

Finally, there would seem to be advantages in teaching this Unit before students embark on their final investigation: not to encourage them to do specifically 'electronics' investigations, but because some of the techniques and principles learned may be useful in planning and instrumenting a wide variety of investigations.

LIST OF SUGGESTED EXPERIMENTS AND DEMONSTRATIONS

Section I1

BASIC OPERATIONAL AMPLIFIER CIRCUITS

Getting started

As an introduction we suggest that not more than one period is used, in which students are told in general what the Unit is about, and concluding with two or three very quick demonstrations. The demonstrations at this point are primarily to arouse interest: looking in detail at how they work must wait until later. The following suggests one way in which this introduction might proceed.

Introduction to the Unit

This Unit is about another aspect of electronics, and also about the uses of electronics, particularly in control systems. By 'control' we mean automatic control: something needs to be done, but we don't want to have to do it all ourselves.

One aspect of control is the predetermined sort. A time switch is programmed to turn radiators on and off at fixed times of day. More elaborate is the control of a washing machine or of a milling machine, making each perform a series of complex operations in the right sequence.

A more demanding aspect is that of automatically regulating something according to unpredictable variations. For example, greenhouse windows can be made to open automatically if the weather becomes hot.

At this point students might offer other examples, such as:

Thermostatically controlled systems: central heating, irons, etc.
Traffic bollard lights controlled by photocells.
Venetian blinds operated by heat or light.
Stabilizers for ships.
Governors for steam engines.

Mechanical, electrical, and chemical processes in industry abound with control systems. In a chemical plant, the temperature, pressure, and rate of flow of reactants may have to be kept to precise values at every instant. For this there are automatic mechanisms, not a person watching dials and adjusting knobs; slight changes in value are sensed automatically and the information fed back to machinery (or circuitry) which applies a compensating adjustment.

So we arrive at the concept of feedback – negative feedback because it is designed to compensate. 'Too fast: make it slower' typifies the function. The idea is not of course confined to mechanisms: we ourselves rely on it, for example in catching a ball, where we use feedback from our eyes to guide our hands, or at the subconscious level in the process of keeping our balance.*

In conclusion, students should see either two or three of demonstrations I1. These need to be prepared in advance by the teacher, and should be shown quite quickly. I1a shows illumination being controlled purely manually, and brings out some of the shortcomings of the process; I1b shows the same situation but governed by automatic control. I1c is a different automatic control situation, but has perhaps more 'spectator appeal'. (It is also of interest as illustrating the principle whereby communications satellites keep themselves accurately facing the Earth, using infra-red radiation instead of visible light.)

Students can see exactly what is happening in I1a, but b and c will at this point be beyond their understanding: they are exhibits, not subjects for analysis. Later in the Unit there will be an opportunity for students to set these up, when they are in a position to understand how they work. They appear in the students' laboratory notes as demonstrations I13 and I15.

DEMONSTRATION
I1 Introduction to linear electronics and control

I1a Manual control of illumination

ITEM NO.	ITEM
72	2 lamps, 12 V, 24 W
27	12 V supply
541/1	rheostat, 10–15 Ω, 5 A
1147	light-dependent resistor
	either
1005	ohmmeter
	or
1033	cell holder with one cell
1507	milliammeter
503–6	retort stand base, rod, boss, and clamp
1000	leads

**Feedback.* The term feedback is deliberately used in a rather loose sense in this discussion. The intention is to establish the general idea that in many kinds of system information about the output is fed back to the input, which in turn affects the output. In electronic systems one can distinguish between current and voltage feedback, series and parallel feedback, but it is not our intention to do so in this course.

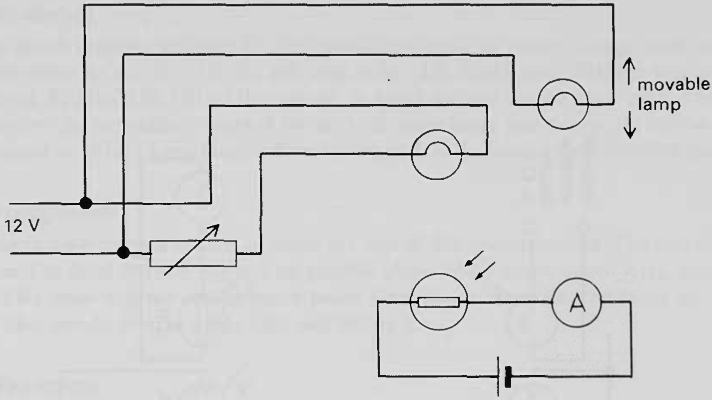


Figure I1
Manual control of illumination.

The light-dependent resistor (L.D.R.) lies on the bench facing upwards, with the variable lamp clamped about 50 cm above it. The other lamp is held by an assistant so that it also sheds light on the L.D.R. The variable lamp is set to about half brightness, and the meter reading noted. The assistant now moves his lamp to represent changes in ambient illumination, and it is the experimenter's task to keep the meter reading at its original value by adjusting the rheostat. In this process feedback via vision from the meter is used to control the variable lamp.

Points to note are: the degree of precision attainable, the time lag in response, and possibly 'overshooting' or even excessive 'oscillation' if the ambient light is increased and decreased at a suitable frequency.

I1b Automatic control of illumination

ITEM NO.	ITEM
1519	operational amplifier with power supply and resistors, 1 kΩ, 1 kΩ, 10 kΩ, 10 kΩ, 100 kΩ, 100 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
72	2 lamps, 12 V, 24 W
1147	2 light-dependent resistors
	either
1005	ohmmeter
	or
1033	cell holder with one cell
1507	milliammeter
1520	power amplifier
59	2 l.t. variable voltage supplies
503–506	retort stand base, rod, 2 bosses, 2 clamps
1000	leads

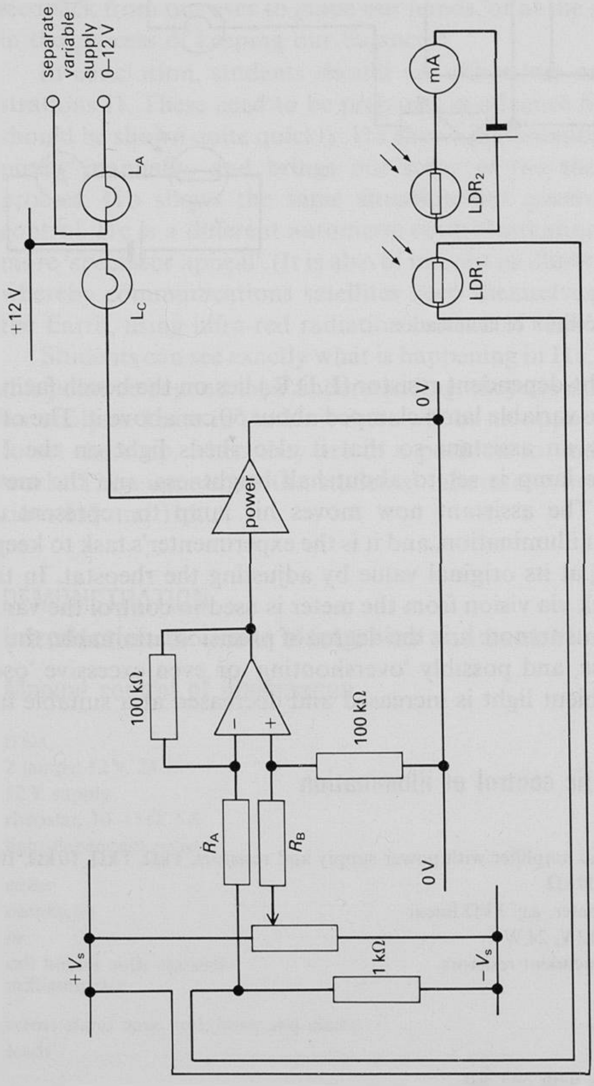


Figure I2
Automatic control of illumination.

The circuit

The circuit is shown in figure I2. The operational amplifier supply voltage could be any value down to +3, 0, −3 V. R_A will need to be 1 kΩ, 10 kΩ, and 100 kΩ in turn (see below). R_B should be 100 kΩ throughout, to avoid making the 'set level' control too sensitive. (In the unlikely event of the set level range being insufficient, R_B can be changed to 10 kΩ.) Long leads will be needed to L.D.R.$_1$ and to the controlled lamp L_C.

Arrangement

A fairly dark room is needed to make the best of this demonstration. The two lamps should be fixed as close together as possible about 50 cm above bench level, with the L.D.R.s close together on the bench below them, facing upwards. The meter for L.D.R.$_2$ should also be where light will fall on it.

Adjustment

Set R_A at 1 kΩ. Turn up the lamp L_A to full brightness. Turn up the voltage of the power amplifier supply to 12 V. Turn the potentiometer to maximum positive potential, so that L_C is bright. Further raise the power amplifier supply voltage as necessary until L_C is as bright as L_A. Finally, adjust the potentiometer again until L_C is only just visibly glowing dull red.

Demonstration

Gradually turn down the brightness of L_A. L_C should brighten, and the meter reading should change very little if at all (see Note *ii* below). Show by covering L.D.R.$_1$ that it is controlling the brightness of L_C. Show by making rapid changes in the light from L_A (including interposing the hand or even switching off and on) that the system corrects instantly and precisely – much better than a human operator.

(If this demonstration is used as part of Section I2, it might be worth changing R_A to 10 kΩ then 100 kΩ, to see how a reduction in gain produces less than total compensation for changes in ambient illumination.)

Explanation (*for class discussion later, when used again as demonstration I13. This might be the task of a student, as a result of the questions posed in the students' laboratory notes*)

L_A provides the 'ambient illumination', *i.e.* daylight, etc., which is subject to unpredictable variation (disturbance). L_C is the 'correcting' lamp which maintains constant illumination as the ambient illumination varies. L.D.R.$_1$ is the *transducer* (note the term) which provides feedback to the control circuit. L.D.R.$_2$ is there for demonstration purposes only, to measure how effective the control system is.

Suppose L_A is reduced in brightness. Less light falls on L.D.R.$_1$, and its resistance increases. Because it is part of a potential-dividing network, the potential applied to R_A becomes more negative. This raises the output potential of the operational amplifier and hence increases the current through the power amplifier and the lamp L_C. The illumination is therefore increased until (see Note *i* below) the resistance of L.D.R.$_1$ is restored to its former value (see Note *ii*), *i.e.* the illumination is restored.

Notes for teachers

i We have here a 'cause-and-effect' loop: light change causes resistance change, resistance change causes light change. 'Until' implies a time sequence; any time delays (in the circuitry, in the transit of light) are of course here negligibly small: the processes are effectively simultaneous.

ii If L_C has to become brighter, the amplifier output must be higher, so its input must be more negative, so the resistance of L.D.R.$_1$ must be somewhat greater. This means that the new equilibrium illumination must be somewhat less than before. If the gain of the circuit is very high the difference will be negligible, but it shows up if the gain is reduced, as suggested above.

I1c Light follower

ITEM NO.	ITEM
	For first circuit (figure I3)
1520	power amplifier
59	2 l.t. variable voltage power supplies
1510	potentiometer, *e.g.* 1 kΩ linear
1147	2 light-dependent resistors
	either
94A	lamp, holder, and stand
	or
	other source of illumination (*e.g.* reading lamp)
	either
	turntable (for 3 cm X-ray diffraction analogue) (if available)
	or
154/1	turntable
9B	small motor/generator unit
1153	insulating tape
1000	leads
	Additional for second circuit (figure I4)
1519	operational amplifier with power supply and resistors, 10 kΩ, 10 kΩ, 100 kΩ, 100 kΩ, 100 kΩ

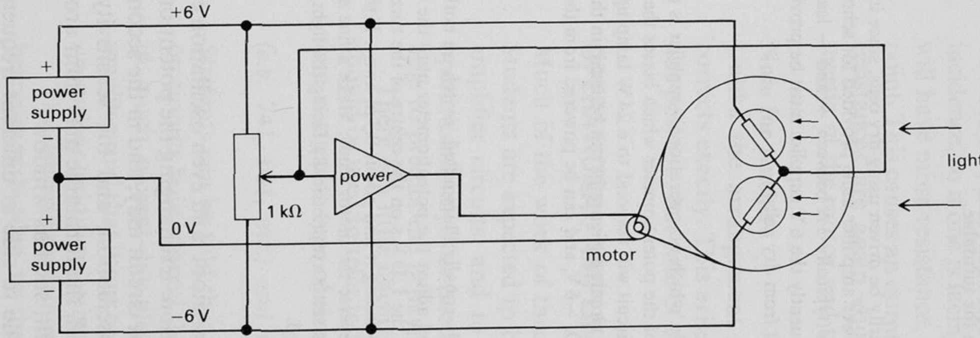


Figure I3
Light follower: first circuit.

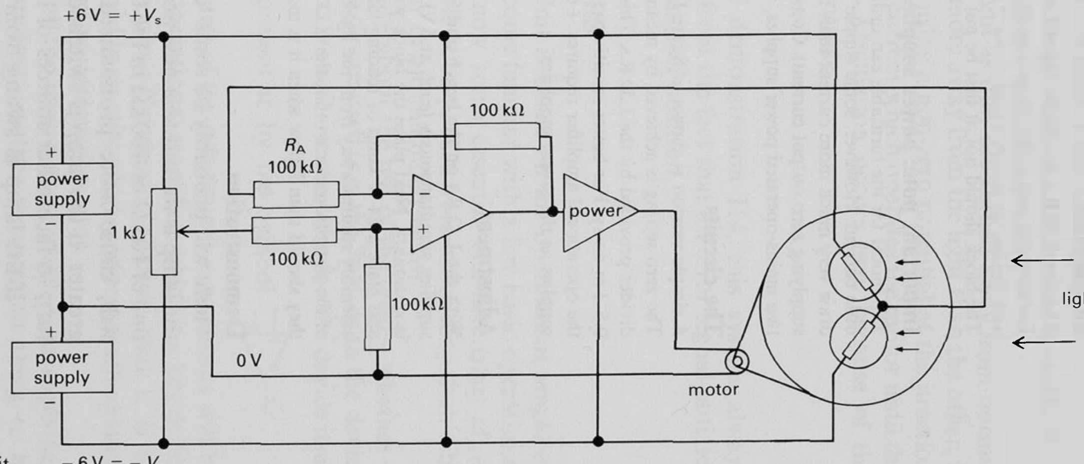


Figure I4
Light follower: second circuit.

Preparation

The two L.D.R.s should be taped firmly to a block of wood so that they face horizontally and at 90° to one another. The three leads from them should be as light and flexible as possible and about 2 m long: 0.45 mm insulated copper wire is suitable. The block should not at first be put on the turntable.

Important note: power supplies

The motor for the turntable can ordinarily be driven using dry cells, since it needs little current. However, some simple power amplifier units provided for schools may draw very much more current than this (typically 3 A between + and − leads when supplying zero output current). Consequently the 6 V supplies must be provided from two mains-operated power supplies, not from dry cells.

The circuits

A simple version is shown in figure I3, in which no operational amplifier is required. The zero setting is achieved by means of the potentiometer which biases the potential divider provided by the L.D.R.s. This circuit will respond to a 24 W lamp up to 0.5–1 m away. For better sensitivity the circuit of figure I4 can be used; in this circuit the operational amplifier requires +6, 0, −6 V, and can be powered from the same supplies as the power amplifier.

Adjustment

With the L.D.R.s on the bench and fairly equally illuminated, switch on both power supplies simultaneously (each at 6 V) and adjust the potentiometer until the turntable is stationary. Next place the block with the L.D.R.s on the centre of the turntable, and place a 24 W lamp or reading lamp facing it and about 0.5 m away. If the assembly rotates away from the light, reverse the motor leads. Make a fine adjustment of the potentiometer so that the L.D.R.s settle symmetrically facing the light, which they should then follow when it is moved.

Demonstration

There will probably be some 'overshoot' and even oscillation each time the lamp is moved to a new position. Exchanging the potentiometer for one of 10 kΩ or 100 kΩ in the first circuit may, and in the second circuit will, cause more pronounced oscillation, and the sensitivity will be greater. If the lamp is switched off, the turntable may hunt around and turn to face other sources of light, such as windows.

If the lamp is swung from side to side at different frequencies, the unit will attempt to follow, but at its resonant frequency it will be very obviously out of phase with the lamp movements.

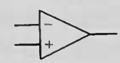
If in the second circuit the gain is increased (by reducing the value of R_A), it is likely that the oscillation will start spontaneously and increase uncontrollably.

Explanation (*for class discussion later, when used again as demonstration I15; this might be the task of a student, as a result of the questions posed in the students' laboratory notes*)
The L.D.R.s intercept less light as they are turned away from normal incidence, so if one is facing more away from the light than the other, it will have more resistance, causing a non-zero potential at the junction point. This causes an output current which operates the motor until the resistances are again equal. Oscillation occurs here because of the inertia of the motor and turntable.

Note for teachers
Unlike the arrangement of demonstration I14, this system always corrects exactly. This arises from the fact that an error signal initiates not position, but rate of change of position.

OPERATIONAL AMPLIFIERS

Much of the work of this Unit is based on the operational amplifier. Students are expected to become familiar with a few basic operational amplifier circuits, and to know some essential characteristics of an operational amplifier (see table I1 on p. 122), from which they should be able to predict the behaviour of an unfamiliar circuit.

There is a little uncertainty about the use of the term – whether it should be used to refer to the device itself (*e.g.* the 741), or to the device in a particular type of circuit. We will use the term for the device itself (*e.g.* 741, 081, etc.) and represent it by the symbol  .

Students can be told that many of the subsequent experiments will be on circuits built around the operational amplifier, which is a key device in much present-day electronics. The name indicates that it is an *amplifier* and that it can be used in circuits which perform mathematical *operations* (addition, subtraction, integration, etc.).

This is worth explaining: all too often a name which is not explained becomes a burden to the student, another meaningless thing to be learned. But in many cases the name does mean something. Understanding the name removes the burden and eases learning.

As a matter of background information, students should of course see an actual operational amplifier: in manufacturers' units the amplifier itself is generally clearly visible. The diagram of figure I5, showing the internal circuitry, is also reproduced in the *Students' guide.*

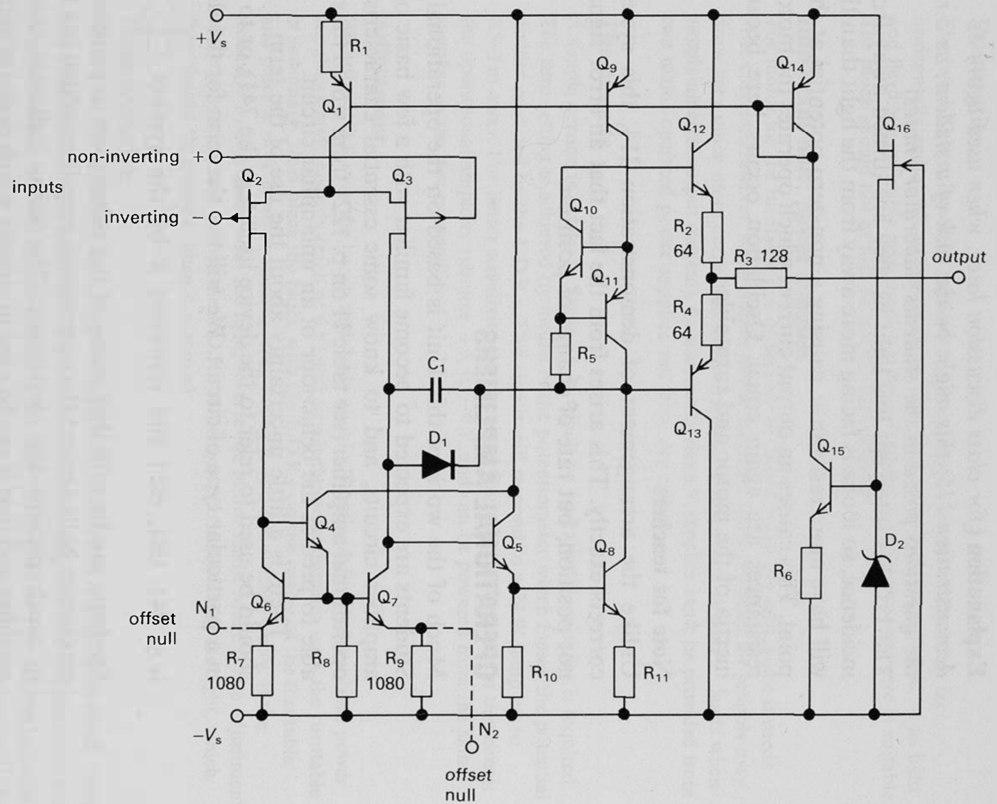


Figure I5
Schematic diagram of the TL081 operational amplifier.
From The BIFET Design manual, *Texas Instruments Ltd.*

Previous knowledge required

Knowledge of the basic current electricity from Unit B, 'Currents, circuits, and charge', will be needed: students could be advised to make sure now that they are fully familiar with $V = IR$; the potential divider; $Q = CV$; $I = \mathrm{d}Q/\mathrm{d}t$.

Questions

Questions 1 to 4 are introductory questions about p.d., current, and resistance; questions 5 to 7 deal with capacitors.

A group of experiments to introduce the operational amplifier

The aims of this first group of short experiments are quite modest: to give students a first experience and some confidence in working with the device; to introduce the use of the inverting input with resistive feedback to give amplification; and to show that the output is limited. The initial approach can be to regard the circuit as a black box and to ask students simply to report what output it gives for certain inputs. It is worth pointing out to students that in this Unit we are concerned with how the operational amplifier circuit behaves between its limiting positions, whereas in Unit C, 'Digital electronic systems', we used gates at their limiting positions and not in the intermediate stages.

The operational amplifier can be presented as a rather more elaborate kind of 'four-terminal box': it needs a power supply, and has two inputs and one output. Explanations as to why it behaves as it does should come later, after experiment I4.

Students will need to be told how to connect the power supply (+, 0, and −) to the operational amplifier unit*, and that inputs are connected between the appropriate terminal and the 0 V line. Similarly, outputs are measured between the output terminal and 0 V (not the negative power rail). Only the inverting input (marked −) will be used at this stage. The non-inverting input will be connected to 0 V.

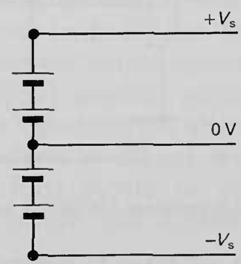


Figure I6
Use of dry cells as power supply for operational amplifier.

*Dry cells can be used in place of a power supply unit with +, 0, and − terminals. Some operational amplifiers work off as little as $\pm$ 3 V, and for these four $1\frac{1}{2}$ V cells will suffice (figure I6), though this will of course limit output voltages to less than $\pm$ 3 V. In circuit diagrams we use the symbols $+V_s$, 0 V, $-V_s$ to denote the power rails.

For clarity the power supply is omitted from all circuit diagrams.

We suggest that each pair or group of students should do experiment I2a, followed at once by one other allocated from I2b, c, or d. The results of the four experiments can then be summarized with the class as a whole. A fast group could also be given the task of setting up demonstration I3, which will help to consolidate the observations so far.

As a fill-in we have included one more optional experiment I2e, but this is not an essential part of this first exploratory exercise.

EXPERIMENT

I2 Introduction to operational amplifiers

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistors, 10 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
1000	leads

I2a Input and output voltages

Apparatus as for experiment I2 plus:

ITEM NO.	ITEM
	either
1507	2 voltmeters, 1 V, 10 V d.c.
	or
1511	oscilloscope
1507	voltmeter

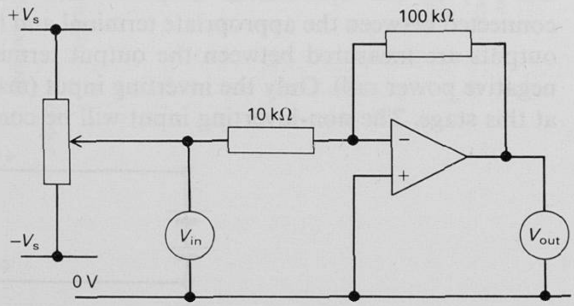


Figure I7
Circuit for experiment I2a (power supply connections are not shown).

Students are asked to observe (simply, not by detailed graph plotting) how the output voltage V_{out} depends on the input voltage V_{in}.

Inputs both positive and negative with respect to the 0 V line should be tried. Students should note that positive input voltages give negative output voltages and vice versa (which is why the input is called the inverting input); that the numerical value of output voltage is greater than that of the input voltage (the device amplifies); that the output voltage never goes above the positive or below the negative supply voltage: indeed, the limiting value of V_{out} is a little less than the power supply voltage.

I2b Response to a.c.

Apparatus as for experiment I2 plus:

ITEM NO.	ITEM
	either
1109	signal generator
	or
27	transformer
1511	oscilloscope

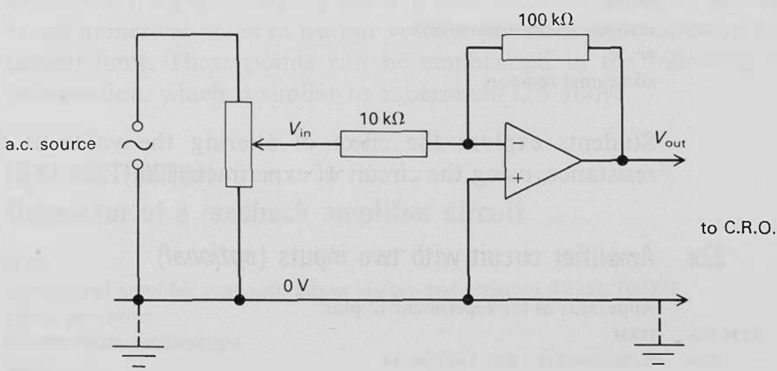


Figure I8
Operational amplifier with a.c. input.

If the signal generator and oscilloscope both have an earthed terminal, these should be the ones connected to the 0 V line. Students should use the oscilloscope to compare input and output signals. For small inputs the output should be a very faithfully amplified version of the input; but if the peak voltage of the input signal exceeds about one-tenth of the supply voltage the output signal will be clipped, becoming eventually virtually a square wave. Students should be able to relate this observation to the maximum positive and negative output voltages observed in experiment I2a.

I2c Effect of changing input resistance

Apparatus as for experiment I2 plus:

ITEM NO.	ITEM
	either
1017	resistance substitution box
	or
1151	additional resistors

Students explore the effect of altering the value of the input resistance, using the circuit of experiment I2a.

I2d Effect of changing feedback resistance

Apparatus as for experiment I2 plus:

ITEM NO.	ITEM
	either
1017	resistance substitution box
	or
1151	additional resistors

Students explore the effect of altering the value of the feedback resistance, using the circuit of experiment I2a.

I2e Amplifier circuit with two inputs (*optional*)

Apparatus as for experiment I2 plus:

ITEM NO.	ITEM
1510	potentiometer, *e.g.* 1 kΩ linear
	either
1151	resistor, 10 kΩ
	or
1017	resistance substitution box
	either
1507	voltmeter
	or
1511	oscilloscope

Students use the circuit of figure I9 to explore the relation between input and output voltages. The outcome will not, however, be taken up in the following consolidation demonstration I3.

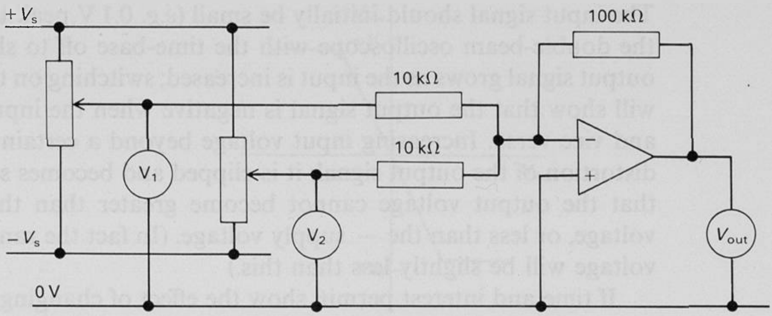


Figure I9
Operational amplifier circuit with two inputs.

Consolidation
The points which should have emerged from the introductory experiments are: if input voltage is positive then output is negative, and vice versa; numerical value of output voltage increases with input up to a certain limit. These points can be emphasized in the following demonstration, which is similar to experiment I2b above.

DEMONSTRATION
I3 Behaviour of a feedback amplifier circuit

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistors, 10 kΩ, 100 kΩ
1109	signal generator
1511	double-beam oscilloscope
1000	leads

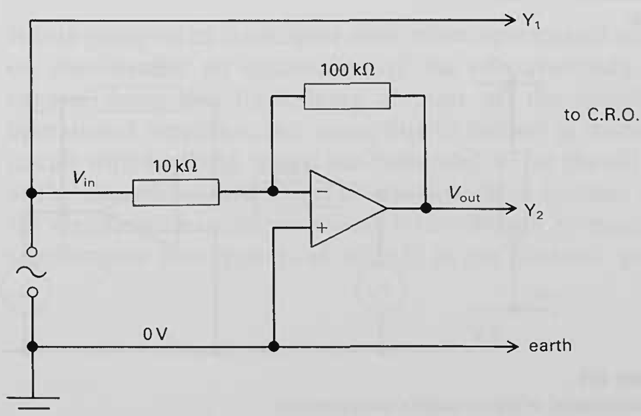


Figure I10
Measurement of input and output voltages on double-beam oscilloscope.

The input signal should initially be small (*e.g.* 0.1 V peak to peak). Use the double-beam oscilloscope with the time-base off to show that the output signal grows as the input is increased; switching on the time-base will show that the output signal is negative when the input is positive and vice versa. Increasing input voltage beyond a certain limit causes distortion of the output signal: it is clipped and becomes square. Show that the output voltage cannot become greater than the + supply voltage, or less than the − supply voltage. (In fact the range of output voltage will be slightly less than this.)

If time and interest permit, show the effect of changing the value of either input or feedback resistor, *i.e.* that amplification depends on the ratio R_f/R_{in}.

More careful experimentation

Before discussing why the circuit behaves as it does, students can do a more careful experiment, plotting a graph of output against input voltage. To economize in time, it would be well to ensure that they set up the graph first, then plot their readings directly on it; a table of results is not needed.

EXPERIMENT
I4 Input–output characteristic of a feedback amplifier circuit

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistors, 10 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
1507	voltmeter, 1 V d.c.
1507	voltmeter, 10 V d.c.
1000	leads

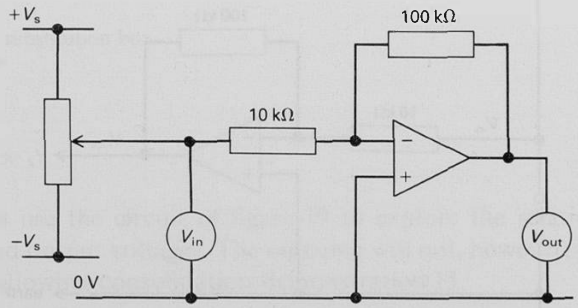


Figure I11
Measurement of input–output characteristic.

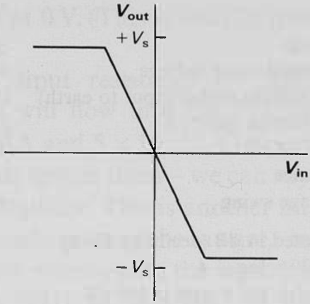


Figure I12
Input–output characteristic.

The potentiometer is connected across the + and − terminals of the power supply of the operational amplifier. The smaller range voltmeter is used to measure input voltages. The power supply voltage relative to 0 V should not exceed the range of the meter used to measure output voltage (*i.e.* 10 V d.c.). Students measure V_{out} for various values (positive and negative) of V_{in}, and plot a graph (figure I12). They should obtain a straight line of negative slope, which turns and becomes horizontal when (or slightly before) the output voltage reaches the supply voltage ($+V_s$ or $-V_s$). The slope of the line should be -10.

Questions

Questions 8 to 11 could be answered at this stage.

THEORY OF THE OPERATIONAL AMPLIFIER

Before going on to investigate some other operational amplifier circuits we should offer an explanation of the effects already observed. We suggest doing this by looking at (part of) the specification for an operational amplifier, and using this to deduce in discussion how the circuit with feedback which has been used so far should behave. Some of the essential points about the device which become apparent from the specification could, if desired, be confirmed by measurement. (The specification also appears as table I1 in the *Students' guide*.)

Type	741	081
Supply voltage	± 3 to ± 18 V	± 3 to ± 18 V
Max. differential input voltage	30 V	± 30 V
Max. input voltage (either input to earth)	15 V	$\pm V_s$ (supply voltage)
Open-loop voltage gain*, A	2×10^5 (106 dB)	2×10^5 (106 dB)
Input resistance	2 MΩ	$10^{12}\,\Omega$
Output voltage swing	± 13 V	± 13.5 V

*Gain is quoted in dB (decibels). Using
gain (dB) $= 20 \times \log_{10} A$
we find $A = 2 \times 10^5$ if gain is 106 dB.
Students should be given the value 2×10^5 which is the ratio V_{out}/V_{in}, not 106 dB.

Table I1
Partial specification for an operational amplifier

Data are easily obtained, for example from the RS Components Catalogue, and it may be worth asking students to extract the data for a particular device for themselves. (The Catalogue is available from RS Components Ltd, P.O. Box 99, Corby, Northamptonshire, NN17 9RS.) The six pieces of information listed here are sufficient for our purpose.

So far we have used only the *inverting input* of the amplifier; the non-inverting input has always been connected to 0 V. Later we shall use the non-inverting input. The data specifies a maximum voltage between either input and 0 V (earth). It also gives a *maximum differential input voltage*. This is the maximum potential difference which can be applied between the two inputs (and its value is equal to or slightly less than the difference between + and − supply voltages). The operational amplifier in fact amplifies the potential difference between its two input terminals by the factor A:

$$V_{out} = A(V_+ - V_-)$$

where V_+ and V_- are the voltages at the non-inverting and inverting terminals respectively.

The open-loop voltage gain A, *i.e.* the gain when there is no feedback from output to input, is enormous: 2×10^5 compared with about 10 in the experiment *with* feedback. 'What p.d. between the two inputs will be needed to produce an output voltage of 10 V (open loop)?' ($10/(2 \times 10^5)\,\text{V} = 5 \times 10^{-5}$ V, or 50 μV.)

This is so small in comparison with voltages measured elsewhere in operational amplifier circuits that we can ignore it: we can assume that there is *no* potential difference between the two input terminals. This is an important characteristic to bear in mind when analysing operational amplifier circuits. It follows that if, as is often (but not always) the case, the non-inverting input is connected to 0 V, we can assume that the inverting input is also at 0 V. (This is usually referred to as the 'virtual earth'.)

For the 741 the input resistance is 2 MΩ, for the 081 it is 10^{12} Ω. 'What current will flow into the 741, or the 081, at 50 μV?' ($I = V/R = 2.5 \times 10^{-11}$ A and 5×10^{-17} A respectively.) These currents are so small that we can ignore them – we can say that no current flows into the operational amplifier. This is another important characteristic of the device. (As teachers may be aware, in the very simplified calculations here we are overlooking the input bias current, which may be about 80×10^{-9} A (741), or 30×10^{-12} A (081). See, for example, CLAYTON, *Operational amplifiers*, 2nd edn, pages 63–4.)

The open-loop voltage gain – nominally* 2×10^5 – is enormous but not particularly useful. We rarely want to amplify a voltage by such a huge factor. Moreover, the open-loop performance is not reliable: it tends to be sensitive to temperature, and the gain falls off significantly at high frequencies. (Any attempt to use the amplifier in an open-loop circuit will probably show the output fluctuating wildly, however small the input.) The amplifier can be made stable (and its temperature and frequency range increased) by negative feedback. This very much reduces the gain, and is what has been used in the circuits we have used so far.

An important distinction

The above discussion of the characteristics of the operational amplifier has included mention of various parameters including (741 values) input resistance (≈2 MΩ), input current (25×10^{-12} A), and potential difference between the two inputs (50 μV). It is important that students understand that these apply to the *device* itself. In what follows we are concerned with a *circuit*: then the input current I_{in} is the current in the input resistor R_{in}. R_{in} is typically 10 kΩ, and I_{in} is 0.1 mA. The input voltage to the *circuit* (measured in experiments I2, I4) may be several volts; it is the potential difference between the end of the input resistor furthest from the operational amplifier and 0 V. To avoid confusion we shall use V_- and V_+ for the potential at the inputs to the operational amplifier itself; V_{in} and V_{out} for the input and output voltage to the circuit (see figure I13).

Input voltage (V_{in}) and voltage at the inverting input (V_-) are *not* the same. (In figure I13 $V_+ = 0$ and V_- is virtually 0.)

Questions

Questions 12 and 14 give practice with these ideas. Question 13 is an optional learning question about dB.

*The gain of individual units can in fact vary by up to a factor of 2 either way.

AMPLIFIER WITH NEGATIVE FEEDBACK: THE THEORY

If students have grasped the two salient points about the operational amplifier itself – negligible input current, and negligible p.d. between the two inputs – they should be able to see why the feedback amplifier circuit (figure I13) behaves as it does.

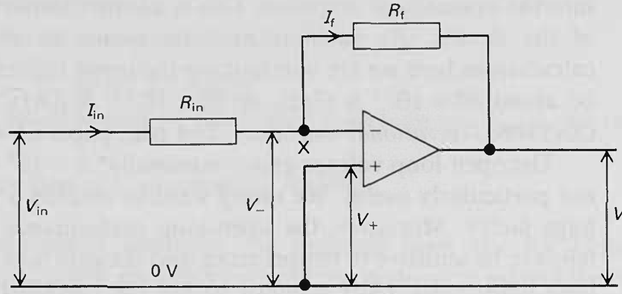


Figure I13
Amplifier with negative feedback.

Questions

Questions 15 to 17 are learning questions which deal with this. Question 18 is optional: it discusses exactly how feedback stabilizes gain. Question 44 is an optional question about feedback and gain in more 'control technology' terms.

If students have tackled questions 15 to 17 by themselves it may not be necessary to repeat in class the argument summarized below.

The non-inverting input is connected to 0 V; the inverting input is therefore also at 0 V. The current in the input resistor, R_{in}, is therefore

$I_{in} = (V_{in} - 0)/R_{in} = V_{in}/R_{in}$

Similarly, the current in the feedback resistor R_f (in the direction marked) will be

$I_f = (0 - V_{out})/R_f = -V_{out}/R_f$

But since no current goes in to or out of the amplifier itself these two currents are the same. So

$V_{in}/R_{in} = -V_{out}/R_f$, from which

$V_{out}/V_{in} = -R_f/R_{in}$

We now see why our earlier experiments with $R_f = 100\,\mathrm{k\Omega}$ and $R_{in} = 10\,\mathrm{k\Omega}$ give a voltage gain of -10. And we see how the operational amplifier can be used to perform the mathematical operation of multiplying by a constant factor. (Of course for division, *i.e.* multiplying by a factor less than 1, $R_{in} > R_f$.)

Here are some points to emphasize about the relationship $V_{out}/V_{in} = -R_f/R_{in}$:

i The 'closed loop' gain (*i.e.* gain with feedback) does not depend on A, the open-loop gain of the operational amplifier itself, but only on the values of the input and feedback resistors.
ii Because closed-loop gain is independent of A it will not be affected if A changes at high frequency, or due to temperature, etc.
iii All this applies as long as the operational amplifier is not saturated.

The essential points – negligible input current and negligible potential difference between the two inputs – can be demonstrated fairly easily, and although the following demonstration may help to emphasize these points, for many classes it may be more profitable to go on to other uses of the operational amplifier.

OPTIONAL DEMONSTRATION

I5 Investigation of currents and voltages in an operational amplifier

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistors, 10 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
1000	leads

Connect the circuit as in experiment I4 (page 120) with input resistance of 10 kΩ and feedback of 100 kΩ.

I5a Currents

Apparatus as for experiment I5 plus:

ITEM NO.	ITEM
1507	2 microammeters, 100 μA

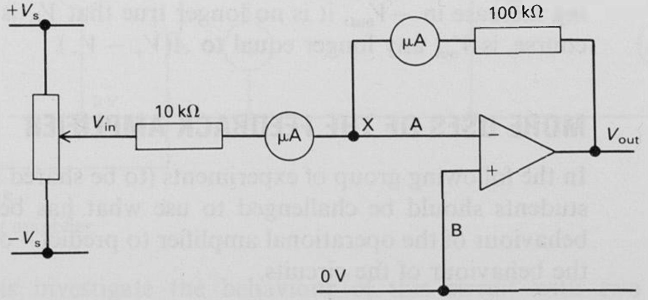


Figure I14
Measuring input and feedback currents.

To show that the current into the operational amplifier itself is negligible, put microammeters into the circuit as shown in figure I14. The difference between their readings should be negligibly small for a range of settings of the potentiometer. On some units it may be possible to insert a microammeter between X and the input (*i.e.* at point A in figure I14), though it might be found that oscillations occur, due to the meter's stray capacitance to earth. This meter should show negligible current. You may also want to show that there is no measurable current between the non-inverting input and earth.

I5b Voltages

Apparatus as for experiment I5 plus:

ITEM NO.	ITEM
1507	2 voltmeters

Since the resistance inside the operational amplifier between the two inputs is very high (2 MΩ for the 741, $10^{12}\,\Omega$ for the 081), it would be meaningless to attempt to measure the p.d. between them with a voltmeter which could well have a resistance of only 10 kΩ. The only way to verify that the inverting input is virtually at zero potential is to calculate the value from the input and output voltages and the values of the resistors. Since the addition of a voltmeter across the input potentiometer may slightly lower the input potential, two voltmeters should be used simultaneously, rather than one used to read the two potentials in turn.

Assumptions and limits

If the output is taken to its limiting value, the situation changes. The 2 MΩ input resistance still allows almost no current to enter or leave the inverting input, so the currents in the two resistors must remain equal. Since V_{in} has finally been increased in magnitude without a corresponding increase in $-V_{out}$, it is no longer true that V_- is near zero. Nor, of course, is V_{out} any longer equal to $A(V_+ - V_-)$.

MORE USES OF THE FEEDBACK AMPLIFIER

In the following group of experiments (to be shared amongst the class), students should be challenged to use what has been said about the behaviour of the operational amplifier to predict – or at least explain – the behaviour of the circuits.

EXPERIMENT
I6 More uses of the feedback amplifier

The important experiments in this group are I6a, Summing amplifier, and I6c, Integration. If students have time to do only one of these they should at least see the other demonstrated. I6b, Subtractor, and I6d, Differentiation, are extensions for fast students.

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistors, 10 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
1507	2 voltmeters
1000	leads

I6a Summing amplifier

Apparatus as for experiment I6 plus:

ITEM NO.	ITEM
1510	potentiometer, *e.g.* 1 kΩ linear
	either
1017	2 resistance substitution boxes
	or
1151	additional resistors
	either
1507	voltmeter
	or
1511	oscilloscope

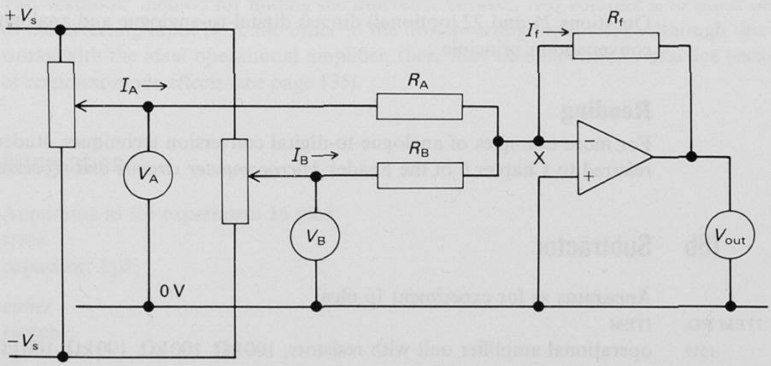


Figure I15
Summing amplifier.

Students investigate the behaviour of this circuit with two input resistors, at first with $R_A = R_B$.

Questions

Questions 19 and 20 are learning questions which deal with the theory below.

The currents in R_A and R_B are $I_A = V_A/R_A$ and $I_B = V_B/R_B$ (X is at 0 V). No current goes into the inverting input so the current I_f in R_f must be $I_A + I_B$. But $I_f = -V_{out}/R_f$.

So $-V_{out}/R_f = V_A/R_A + V_B/R_B$

$$\Rightarrow V_{out} = -(V_A \times R_f/R_A + V_B \times R_f/R_B)$$

The circuit multiplies the two input voltages by different factors, and then adds them: it is a summing amplifier with gains which depend on the values of the resistances.

If $R_A = R_B = R_{in}$, then

$$V_{out} = -(V_A + V_B) \times R_f/R_{in}$$

i.e. each input is multiplied by the same factor. And if all the resistors have the same value ($R_A = R_B = R_{in} = R_f$) then $V_{out} = -(V_A + V_B)$. The output voltage is simply the (negative of the) sum of the input voltages. The operational amplifier is performing another mathematical operation.

An important application of this circuit is in one form of digital-to-analogue conversion. Setting up such a circuit could be used as a buffer experiment for fast students.

Questions

Questions 21 and 22 (optional) discuss digital-to-analogue and analogue-to-digital conversion in principle.

Reading

For more examples of analogue-to-digital conversion techniques, students can be referred to Chapter 4 of the Reader *Microcomputer circuits and processes.*

I6b Subtractor

Apparatus as for experiment I6 plus:

ITEM NO.	ITEM
1519	operational amplifier unit with resistors, 100 kΩ, 100 kΩ, 100 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
	either
1507	voltmeter
	or
1511	oscilloscope

Students now know that a feedback amplifier can be used to multiply by a constant factor; and that it can be used to add two voltages. The task here is to use *two* operational amplifiers to *subtract* voltages, *i.e.* to produce an output $V_{out} = V_1 - V_2$. Figure I16 shows the solution.

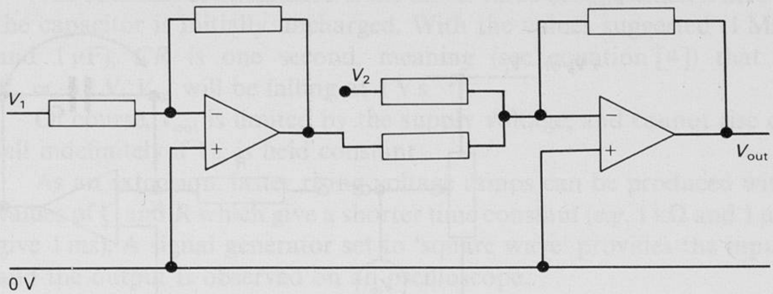


Figure I16
Use of two operational amplifiers to subtract.

All resistors have the same value (*e.g.* 100 kΩ). The output of the first unit is $-V_1$. The second unit sums the two inputs and inverts the sum, so $V_{out} = -(V_2 - V_1) = V_1 - V_2$.

Question

Question 23 sets subtracting as a problem.

Differential amplifier

The 'textbook' method for finding the difference between two voltages is to input one to the inverting input (V_-), the other to the non-inverting input (V_+). Although this works with the ideal operational amplifier, there may be difficulties in practice because of common-mode effects (see page 135).

I6c Integration

Apparatus as for experiment I6 plus:

ITEM NO.	ITEM
1151	capacitor, 1 μF
	either
	switch
	or
52L	mounted bell push
507	stopwatch

The capacitor is initially discharged. Students are asked to observe what happens when the input is switched from 0 V to a few volts negative. The output voltage will rise steadily. They should be able to explain this using what they have learned about the behaviour of the operational amplifier, together with the equations $V = IR$, $Q = CV$, and $I = dQ/dt$. With 1 MΩ and 1 μF, setting V_{in}, for example, to 0.5 V should cause V_{out} to fall at 0.5 V s^{-1}. This should be checked with a stopwatch.

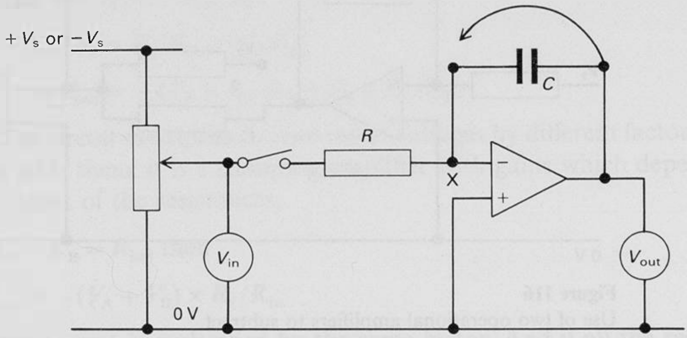


Figure I17
Operational amplifier circuit for integration.

Questions

Questions 24 and 25 are learning questions which deal with the theory.

The point X (figure I17) is at 0 V. The current in R is

$I = V_{in}/R$ [1]

Since no current goes into the operational amplifier, this same current is charging C. It is therefore

$I = dQ/dt$ [2]

The potential difference across the capacitor in the direction of the current is $V = 0 - V_{out}$, so the charge (on the lefthand plate) is

$Q = -CV_{out}$ whence

$dQ/dt = -C\,dV_{out}/dt$

So from [2]

$I = -C\,dV_{out}/dt$ [3]

Equating the two expression for I from [1] and [3] gives

$V_{in}/R = -C\,dV_{out}/dt$ or

$dV_{out}/dt = -(1/CR)\,V_{in}$ [4]

In words, the rate of change of V_{out} is proportional to $-V_{in}$.

The result [4] can also be written as

$$V_{\text{out}} = -(1/CR) \int V_{\text{in}}\, \mathrm{d}t + \text{constant}$$

The constant of integration is the initial value of V_{out}, which is zero if the capacitor is initially uncharged. With the values suggested (1 MΩ and 1 μF), CR is one second, meaning (see equation [4]) that if $V_{\text{in}} = +1$ V, V_{out} will be falling at $1\ \text{V s}^{-1}$.

Of course, V_{out} is limited by the supply voltage, and cannot rise or fall indefinitely if V_{in} is held constant.

As an extension, faster rising voltage ramps can be produced with values of C and R which give a shorter time constant (*e.g.* 1 kΩ and 1 μF give 1 ms). A signal generator set to 'square wave' provides the input, and the output is observed on an oscilloscope.

It is worth mentioning that the time-base of an oscilloscope (or television) uses such a ramp to move the spot across the screen at a steady rate from left to right. In addition, some arrangement is needed to reduce the intensity of the electron beam to zero while it is rapidly returned to its starting point.

Question

Question 26 is about the integrator.

I6d Differentiation

Apparatus as for experiment I6 plus:

ITEM NO.	ITEM
1018	capacitance substitution box
1511	oscilloscope
1109	signal generator

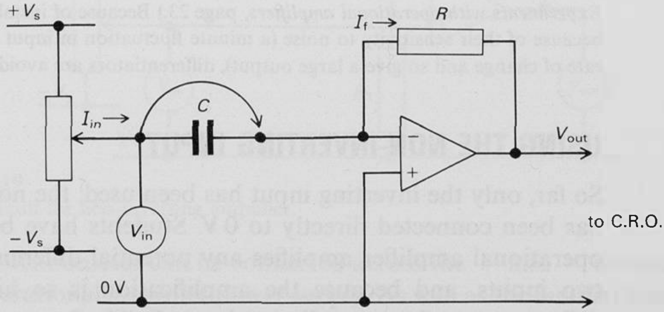


Figure I18
Operational amplifier circuit for differentiation.

If the capacitor is initially discharged, a negative voltage spike should appear at the output when V_{in} is made a few volts positive, and a positive spike when V_{in} is returned to zero or made negative. A signal generator set to square wave or saw-tooth and a range of capacitors will enable students to investigate the response of circuits with different time constants at various frequencies. $R = 100\,\mathrm{k\Omega}$, $C = 0.0001\,\mu\mathrm{F}$, and $f = 2\,\mathrm{kHz}$ are useful values.

Question

Question 27 is a learning question which deals with the theory of the circuit.

Here again, students should be able to explain how the circuit acts to differentiate the input.

$$I_{\mathrm{in}} = \frac{\mathrm{d}Q}{\mathrm{d}t} = C\,\frac{\mathrm{d}V_{\mathrm{in}}}{\mathrm{d}t}$$

$$I_{\mathrm{f}} = (0 - V_{\mathrm{out}})/R$$

$$I_{\mathrm{in}} = I_{\mathrm{f}} \qquad \text{so}$$

$$-\frac{V_{\mathrm{out}}}{R} = C\,\frac{\mathrm{d}V_{\mathrm{in}}}{\mathrm{d}t} \qquad \text{or}$$

$$V_{\mathrm{out}} = -CR\,\frac{\mathrm{d}V_{\mathrm{in}}}{\mathrm{d}t}$$

Practical limitations

The circuit may not in practice give the sharp spikes that are the result of differentiating a 'square wave' input. The performance of the circuit can be improved by adding a low-value resistor (*e.g.* $100\,\Omega$) in series with the input capacitor and a low-value capacitor (*e.g.* $0.01\,\mu\mathrm{F}$) in parallel with the feedback resistor. (The explanation of this is beyond the scope of this Unit, but see, for example, CLAYTON, *Experiments with operational amplifiers*, page 23.) Because of instability problems, and because of their sensitivity to noise (a minute fluctuation in input can have a large rate of change and so give a large output), differentiators are avoided wherever possible.

USING THE NON-INVERTING INPUT

So far, only the inverting input has been used; the non-inverting input has been connected directly to 0 V. Students have been told that the operational amplifier amplifies any potential difference between these two inputs, and because the amplification is so high the potential difference must be so small as to be negligible. So we could assume that the potential at the inverting input was effectively zero (virtual earth).

We now go on to use the non-inverting input (+). Feedback to the inverting input is still necessary. Neither input will be at 0 V, but it is still the case that the potential difference between the two is negligibly small.

In the following series of experiments fast students will naturally get further. All should complete the first two, and a discussion afterwards will emphasize the important points.

EXPERIMENT
I7 Using the non-inverting input

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistors, 100 kΩ, 100 kΩ
1507	voltmeter, 10 V d.c.
1000	leads

I7a Follower circuit with variable gain

Apparatus as for experiment I7 plus:

ITEM NO.	ITEM
1510	potentiometer, *e.g.* 1 kΩ linear
1507	voltmeter
	either
1017	2 resistance substitution boxes
	or
1151	additional resistors

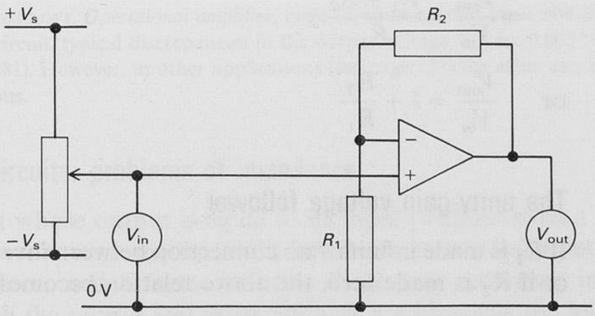


Figure I19
Basic circuit for non-inverting amplifier.

The potentiometer can be connected across the + and − terminals of the operational amplifier power supply. As well as the familiar feedback resistor (R_2), a resistor, R_1, is connected between the inverting input and 0 V (see figure I19). With $R_1 = R_2$ (= 100 kΩ, say), students investigate how V_{out} varies with V_{in}. They will find there is no change of sign

(V_{out} is positive if V_{in} is positive); hence the term non-inverting input. For $R_1 = R_2$ they should find $V_{out} = 2V_{in}$. Let them use what they know about the operational amplifier (no current in or out of either input; negligible p.d. between the two inputs – but in this circuit neither input is at 0 V: the two *are* at the same potential but it is not 0 V) to explain this observation and to predict what V_{out}/V_{in} should be for other values of R_1 and R_2. They can then test their predictions, including the case where $R_2 = 0$ and R_1 is infinite.

Questions

Questions 28 to 31 deal with the theory of this circuit.

The current in R_1 and R_2 is the same (none enters or leaves the negative input). The potential at the negative input is

$$V_- = (V_{out} - 0)\frac{R_1}{R_1 + R_2}$$

$$= V_{out}\frac{R_1}{R_1 + R_2}$$

But since there is no potential difference between the two inputs, $V_- = V_+$. In this case, V_+ is the actual input voltage to the circuit, V_{in}.

So $$V_{in} = V_{out}\frac{R_1}{R_1 + R_2}$$

$$\Rightarrow \frac{V_{out}}{V_{in}} = \frac{R_1 + R_2}{R_1}$$

or $$\frac{V_{out}}{V_{in}} = 1 + \frac{R_2}{R_1}$$

The unity-gain voltage follower

If R_1 is made infinite – no connection between inverting input and 0 V – or if R_2 is made zero, the above relation becomes

$V_{out}/V_{in} = 1$

and the ratio no longer depends on the values of resistances; in fact R_2 can be replaced by a piece of wire (figure I20).

Students should set up this circuit and confirm that $V_{out} = V_{in}$: the output follows the input.

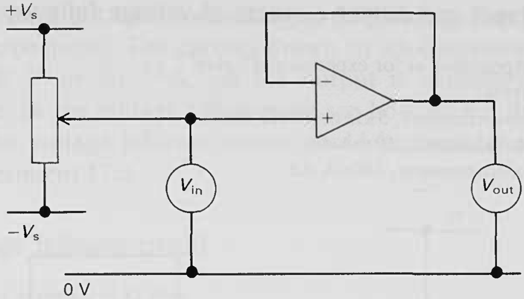


Figure I20
Follower circuit with unity gain.

Questions

Questions 32 to 34 deal with the follower.

Note to teachers: common-mode voltage errors

For the 'ideal' operational amplifier, the output voltage is proportional to the potential difference between the two input terminals. In practice, with real devices the output is also affected slightly by the actual values of V_+ and V_-, not just their difference. This is called the common-mode voltage error: even if V_+ and V_- are equal there will be an output voltage if they are not zero.

A better expression for V_{out} is $V_{out} = A(V_+ - V_-) + A'(V_+ + V_-)/2$. When the non-inverting input is connected to 0 V, both V_+ and V_- are effectively zero, and the second term can be forgotten. But when there is an input to V_+, as in the circuits of figures I19 and I20, this is no longer true: the further the input voltages depart from zero, the greater the error. For the 741 and 081, A is nominally 2×10^5. For A' approximate values are 6 (741) and 32 (081). (Manufacturers refer to the 'common-mode rejection ratio', A/A', expressed in dB.) An analysis of the effects of the error is given in FOXCROFT, *Operational amplifier*, page 72, showing that for a non-inverting amplifier circuit, typical discrepancies in the output voltage V_{out} are 0.003 % (741) and 0.016 % (081). However, in other applications (see page 139) the effect can be much more serious.

Buffer circuits; problems of impedance

A circuit whose output is equal to its input (without even a change of sign!) doesn't seem very useful. In fact, its use is as a *buffer* between, say, a transducer and a meter or other display or recording instrument. Although the output and input voltages are identical, the follower can supply a much bigger current than it takes from the voltage source. For example, a photodiode generates a voltage which depends on illumination, but it may not be able to supply enough current to drive, say, a display meter or chart recorder. A voltage follower acting as a buffer between the two would solve the problem.

The fact that the output current of the follower is much greater than the input current can be seen in an extension to experiment I7a.

I7b Input and output currents of voltage follower

Apparatus as for experiment I7 plus:

ITEM NO.	ITEM
1510	potentiometer, 5 kΩ or 10 kΩ linear
1507	milliammeter, 10 mA d.c.
1507	microammeter, 100 μA d.c.

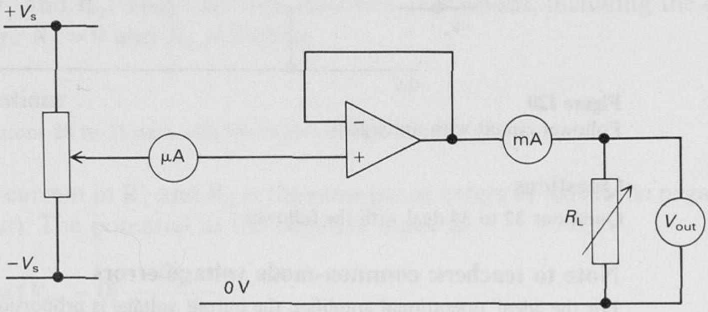


Figure I21
Input and output currents of voltage follower.

The microammeter measures the input current to the circuit. The variable resistor R_L is connected as a load in parallel with the voltmeter which measures output voltage. The milliammeter measures output current, which is seen to be much larger than the input current, and of course depends on the resistance of R_L. Students may wonder where the extra current comes from (the power supply to the operational amplifier).

Teachers might like to be aware of two points here. Firstly, the input current: on our simplified explanation of the operational amplifier the p.d. between the two inputs is V_{out}/open-loop gain (say 10^{-5} V for 1 V output), so the input current we measure might be expected to be that which will give a p.d. of 10^{-5} V across the input resistance of 2 MΩ (741). This amounts to 5×10^{-12} A! However, there will in fact be an 'input offset current' superimposed on this, and for the 741 this could well be of the order of 10^{-8} A. But this still means that the effective input resistance of the circuit is very high – much higher than that of the 2 MΩ value of the 741 operational amplifier's own input resistance.* Secondly, the output voltage is 'ideally' kept equal to the input voltage by the feedback, but of course there is a limit to the total current which the operational amplifier can supply. Nevertheless, changes in output (load) current can, up to a point, cause only small changes in V_{out}, so the effective output resistance ($-\Delta V_{out}/\Delta I_{out}$) can be very low indeed.

An electrometer/d.c. amplifier or a digital voltmeter will have been used often in the course to measure voltages of sources which cannot supply

*The input resistance of the 081 is much higher (10^{12} Ω) and its input offset current is much less.

an appreciable current, or whose voltage would change if a current were drawn (*e.g.* capacitors). The current drawn by an electrometer may be as low as 10^{-11} or 10^{-12} A, yet the output is enough to deflect a milliammeter. In the voltage follower we see how such a device might work, and the voltage follower circuit could be used to demonstrate this (see experiment I7c).

I7c Use of voltage follower circuit

Apparatus as for experiment I7 plus:

ITEM NO	ITEM
1151	capacitor, 50 μF
1033	cell holder with one cell

Charge the capacitor from the 1.5 V cell and then try to measure the p.d. across it with a moving-coil voltmeter (figure I22).

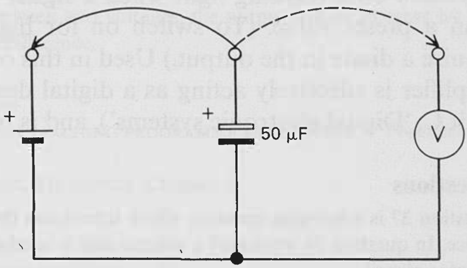


Figure I22
Attempt to measure p.d. across capacitor with voltmeter.

The reading on the voltmeter rapidly drops as the capacitor discharges through it.

Then use the voltage follower as a buffer between the charged capacitor and the voltmeter (figure I23). Now the only current drawn from the capacitor is the tiny input current to the operational amplifier, and the voltmeter reading is sensibly constant.

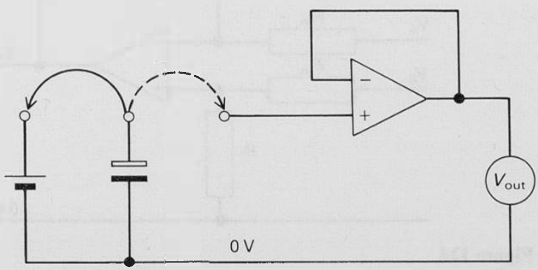


Figure I23
Measuring voltage across capacitor using voltage follower.

Note that a 50 μF capacitor is recommended. There is a complication arising from the input offset current referred to on page 136. This is a current which unavoidably flows between the two input terminals. With the polarity shown in figure I23, its effect is to accelerate the discharge of the capacitor (and if the capacitor is initially discharged it will charge it up steadily in the opposite direction!). With capacitance values of the order of 0.01 μF or less, the result with the 081 is an unacceptable drift in the output voltage. For the 741 the current is much greater, and 1 μF is the practicable lower limit.

The operational amplifier as a differential amplifier

Since the operational amplifier responds to the difference in potential between its two inputs, it can be used to detect or to measure the difference between two input voltages. From $V_{out} = -A(V_- - V_+)$ it follows that if $V_+ > V_-$ the output voltage will be positive, and if $V_+ < V_-$ the output voltage will be negative. So it could be used, for example, to switch on a warning light when a signal voltage is higher or lower than a preset value. (To switch on for high but not for low would require a diode in the output.) Used in this on–off way, the operational amplifier is effectively acting as a digital device (like the gates used in Unit C, 'Digital electronic systems'), and is referred to as a *comparator*.

Questions

Question 37 is a learning question which introduces the comparator as a warning device. In question 38 (optional) a comparator is used to build an astable multivibrator circuit.

However, the differential amplifier can also be used in a linear way, using the circuit of figure I24. In this circuit the output voltage is $V_{out} = (V_B - V_A)R_f/R_{in}$; sensitivity is increased if $R_f > R_{in}$. If $R_f = R_{in}$ then $V_{out} = V_B - V_A$.

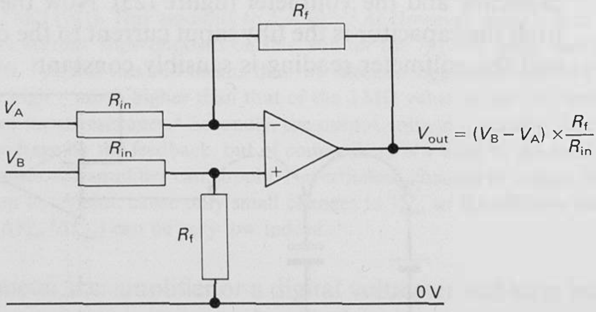


Figure I24
Differential amplifier.

This arrangement could be used, for example, with a high gain as a sensitive bridge balance detector to measure small resistance changes as a deflection from zero, or as in sections I2 and I3 for a variety of control purposes.

Questions

Questions 35 and 36 deal with the theory and use of the differential amplifier.

From the above it appears that we have a device whose output is proportional to the difference between two voltages, and could therefore be used as a single unit subtractor – in contrast to the subtractor suggested in experiment I6b, which needed two units (one to multiply by -1, and one to add). However, common-mode voltage effects (see above, page 135) complicate the story: V_{out} also depends on the absolute values of V_A and V_B, and the larger they are, the greater the error this circuit will give as a subtractor. Analysis (see FOXCROFT, *Operational amplifier*, page 73) shows that the percentage discrepancy is proportional to $V_B/(V_A - V_B)$, so it is large when one is measuring small differences between large voltages. Typically, if one is measuring a 0.1 % difference between two voltages, the output will be in error by 3 % (741) or 17 % (081) of that difference.

Reading for students

ELECTRONIC SYSTEMS TEACHING PROGRAMME ESP700 Book 4 *Feedback systems*, Sections 7, 8, 9.
CLOSE and YARWOOD, *Electronics*. Chapter 5.

Reading for teachers

CLAYTON, *Operational amplifiers*.
FOXCROFT, *Operational amplifier*.
GOUGH *et al.*, *Notes for guidance on the electronics option*.
HOROWITZ and HILL, *The art of electronics*. Chapter 3.
PLANT, *Operational amplifier applications*.

SECTION I2

MORE FEEDBACK; CONTROL

In this Section the emphasis is on feedback: first some further uses in electronic circuits, then more generally in a variety of systems leading on to a consideration of some of the essential features of control systems.

Integrator with feedback

Question

Question 39 is a learning question which deals with the theory below.

Students have seen in experiment I6c that for the circuit of figure I25

$$V_{out} = -\frac{1}{CR}\int V_{in}\,dt + \text{constant}$$

or

$$\frac{dV_{out}}{dt} = -\frac{1}{CR}V_{in}$$

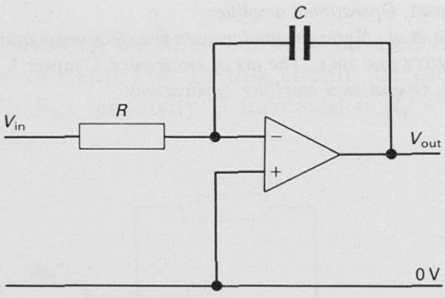


Figure I25
Integrator circuit.

Suppose a wire (zero resistance) connects the output to the input (figure I26). The input and output voltages must now be equal: $V_{out} = V_{in} = V$, and so $dV/dt = -V/CR$. Students should recognize this as the differential equation for exponential decay, and that it implies $V = V'\,e^{-t/CR}$. Here the constant V' is the initial p.d. across the capacitor. This can be tested experimentally.

EXPERIMENT
I8 Integrator with feedback

ITEM NO.	ITEM
1519	operational amplifier unit with power supply and resistor, 1 MΩ, capacitor, 1 μF
1511	oscilloscope
1000	leads

I8a Negative feedback: solving $dV/dt = -V/RC$

Apparatus as for experiment I8 plus:

ITEM NO.	ITEM
1033	cell holder with four cells

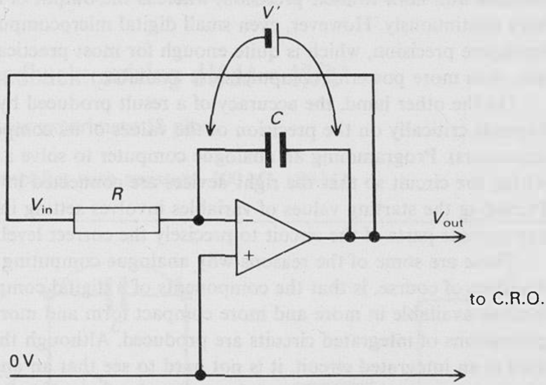


Figure I26
Integrator with feedback.

The output voltage is fed back to the input by a wire; there is no external input voltage to the circuit. There must, however, be a p.d. (V') across the capacitor initially, and this can be any value up to the limiting value of V_{out}. Students observe the decay of the voltage at the output terminal on an oscilloscope set to a slow time-base speed (about $1\,\mathrm{cm\,s^{-1}}$). The familiar exponential decay should be seen; the time constant depends on the values of R and C.

Students might be confused by the fact that there appear to be two kinds of feedback going on in figure I26: from the output to the inverting input via the capacitor, and from the output to the input resistor by a wire. It is the former which makes $dV_{out}/dt = -V_{in}/CR$; and it is the latter which makes the output voltage equal to the input voltage to the circuit, hence making V_{out} and V_{in} equal and so giving $dV/dt = -V/CR$.

Students who finish experiment I8a early can go on to I8b.

Opportunities for revision

This Unit provides many opportunities for revision, and at this stage teachers may want to spend a little while revising some basic ideas about exponential decay, time constant, and so on, in a variety of contexts including radioactivity.

Analogue computing

The last experiment is another example of how operational amplifier circuits can be used to compute (addition, multiplication by a constant, etc., have already been seen). These circuits are the basis for analogue computers, so called because the voltages produced are analogous to variables being computed, as in the example of radioactive decay. Digital computers, whose basic gates and memory devices were the subject of Unit C, 'Digital electronic systems', operate on binary digits. Thus they have a fundamental limit to their precision, whereas the output of an analogue computer can vary continuously. However, even small digital microcomputers work to six- or even ten-figure precision, which is quite enough for most practical purposes. (And if it is not, then more powerful computers are available.)

On the other hand, the accuracy of a result produced by an analogue computer depends critically on the precision of the values of its components (resistors, capacitors). Programming an analogue computer to solve a particular problem means wiring the circuit so that the right devices are connected in the correct configuration. Providing the starting values of variables involves setting initial values of voltage at appropriate parts of the circuit to precisely the correct level.

These are some of the reasons why analogue computing is now little used. Another, of course, is that the components of a digital computer (memory, etc.) become available in more and more compact form and more cheaply as new generations of integrated circuits are produced. Although the operational amplifier itself is an integrated circuit, it is not hard to see that an analogue computer capable of solving complex problems might need to be a pretty bulky affair.

Nevertheless, a little analogue computing may have educational value, especially as it is relatively easy to set up circuits which produce voltages analogous to many physical situations. FOXCROFT, *Operational amplifier*, gives detailed solutions to several problems including free fall, bouncing ball, and damped oscillations.

Reading

An example in which a circuit containing four operational amplifiers can be used to solve a problem in mechanics is given in the article 'Electromechanical similarities' in the Reader *Physics in engineering and technology*.

An example of positive feedback

All the examples seen so far have been of *negative* feedback: the signal fed back from the output has been of opposite sign to the input. Students may not have realized this, because when the inverting input is used the output is automatically negative if the input is positive, and vice versa. But the point must be made clear now before going on to positive feedback. We have seen that negative feedback has the effect of reducing the gain of the circuit, and making it more stable. Suppose the input signal suddenly increases (becomes more positive); the output voltage will become more negative. So the feedback voltage becomes more negative and the combined input (*i.e.* original input plus feedback) changes less than it would in the absence of feedback. And hence the gain of the system is less.

Positive feedback may be easier to explain. If the feedback signal *increases* in response to an increase in the input signal then we seem to have the makings of a runaway situation. But before testing this speculation we need to know how to provide the positive feedback. If students are given the hint that *two* operational amplifiers are needed, they should be able to suggest the solution: one is used as an integrator, the second as an inverter (*i.e.* to multiply by -1). Students who finish experiment I8a early can do the following extension experiment.

I8b Positive feedback: solving $\mathrm{d}V/\mathrm{d}t = V/RC$

Apparatus as for experiment I8 plus:

ITEM NO.	ITEM
1519	operational amplifier with resistors, 100 kΩ, 100 kΩ

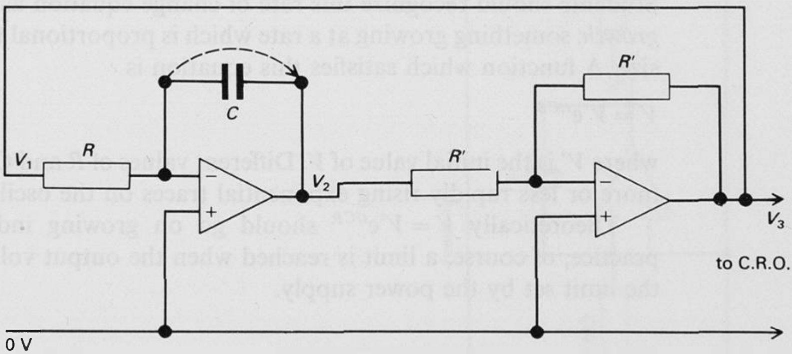


Figure I27
Positive feedback.

Students observe the output of the second unit on the oscilloscope. Short-circuiting C will bring the output V_3 to zero. When this connection is broken the output will rise exponentially until it reaches its limiting value. The time constant (time needed to rise by a factor of e [$\approx$2.72]) is CR, and component values could be changed to illustrate this.

Question

Question 40 deals with this circuit.

The input and output voltages of the two circuits are V_1, V_2, V_3 (figure I27). For the first unit,

$$dV_2/dt = -V_1/CR$$

For the second unit,

$V_3 = -V_2$ (if input and feedback resistors are equal)

So

$$-dV_3/dt = -V_1/CR$$

When the feedback wire is connected, we make

$$V_1 = V_3\,(=V)$$

And so

$$dV/dt = V/CR.$$

Students should recognize this rate of change equation as exponential *growth*: something growing at a rate which is proportional to its present size. A function which satisfies this equation is

$$V = V'e^{t/CR}$$

where V' is the initial value of V. Different values of R and C should give more or less rapidly rising exponential traces on the oscilloscope.

Theoretically $V = V'e^{t/CR}$ should go on growing indefinitely. In practice, of course, a limit is reached when the output voltage reaches the limit set by the power supply.

A circuit to produce oscillation

Oscillations are a familiar theme of the course: mechanical oscillations in Unit D, 'Oscillations and waves'; electrical oscillations in Unit H, 'Magnetic fields and a.c.'. And the signal generator which produces sinusoidal or 'square-wave' output at the turn of a switch (and may also serve as an amplifier) has been much used throughout the course. The next demonstration shows one way in which sinusoidal oscillations can be produced electronically, and for some students it offers a nice link between physics and mathematics.

DEMONSTRATION

I9 A circuit to produce oscillation

ITEM NO.	ITEM
1519	3 operational amplifier units with power supply and resistors, 1 MΩ, 1 MΩ, 1 MΩ, 1 MΩ, and capacitors, 1 µF, 1 µF
1033	cell holder with one cell
1511	double-beam oscilloscope
1017	2 resistance substitution boxes
1018	2 capacitance substitution boxes
1000	leads

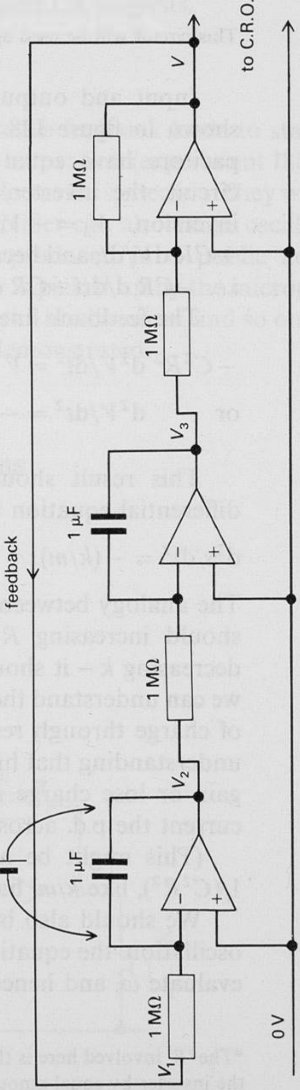


Figure I28
Circuit to produce oscillation.

The strategy here could be either to analyse the circuit first to predict its behaviour, or to see it working first and analyse it afterwards (the analysis is given below).

To start the oscillation, the cell providing V' is momentarily connected as shown in figure I28. Whilst oscillation is taking place, its frequency should be measured. The circuit needs to be left set up until after discussion of it, so that the effect of changing resistance and capacitance values can be checked against experiment.

This circuit will be used again in demonstration I11: it may be worth leaving it set up.

Input and output voltages of the three stages of the circuit are shown in figure I28. Assume that all the resistors and both the capacitors have equal values. Starting at the righthand side of the circuit, the inverter, and calling its output V, then because of the inversion, $V_3 = -V$. Because of the middle unit, V_2 must be $+CR\,\mathrm{d}V/\mathrm{d}t$, and because of the lefthand unit, V_1 must be $-CR\,\mathrm{d}V_2/\mathrm{d}t$, *i.e.* $-CR\,\mathrm{d}/\mathrm{d}t(+CR\,\mathrm{d}V/\mathrm{d}t)$, which is $-C^2R^2\,\mathrm{d}^2V/\mathrm{d}t^2$.

The feedback line makes V_1 equal to V, giving

$$-C^2R^2\,\mathrm{d}^2V/\mathrm{d}t^2 = V$$

or $$\mathrm{d}^2V/\mathrm{d}t^2 = -\frac{1}{C^2R^2}V$$

This result should be recognized by students as the form of differential equation they met in Unit D, 'Oscillations and waves':

$$\mathrm{d}^2s/\mathrm{d}t^2 = -(k/m)s$$

The analogy between the two equations can be explored: 'What effect should increasing R* or C have?' (The same as increasing m or decreasing k – it should slow down the oscillation.) At the same time, we can understand the physics of this system by thinking about the flow of charge through resistors on to capacitors. We would hope for the understanding that higher R means smaller currents, so capacitors will gain or lose charge more slowly. Lower C means that for the same current the p.d. across the capacitor will rise and fall more rapidly.

(This might be a moment to recall dimensions, and show that $1/(C^2R^2)$, like k/m, has dimensions T^{-2}.)

We should also be able to calculate the expected frequency of the oscillation: the equations are $\mathrm{d}^2s/\mathrm{d}t^2 = -\omega^2 s$, and $f = \omega/2\pi$, so we can evaluate ω, and hence f, if we know the values of C and R.

*The 'R' involved here is the value of the resistors in the integrators. Changing those in the inverter by equal amounts would of course make no difference.

Question

Question 41 is about frequency calculation.

Students should be able to see how, in principle at least, the oscillation can be converted from sinusoidal to square (limiting values of V_{out}); and it would seem quite sensible of manufacturers to let us use the operational amplifiers within a unit to amplify signals as well as to produce oscillation. But the actual circuitry of a signal generator is likely to be more complicated than figure I28 suggests.

Oscillation and feedback

Students know that in amplifiers, negative feedback leads to stability; and that positive feedback can lead to saturation (experiment I8b, and the bistable circuit in Unit C, 'Digital electronic systems'). They will not know that positive feedback in an amplifier can also cause oscillation.

The annoying squeal sometimes produced in a public address system if a signal from the loudspeaker is picked up by the microphone and fed back to the amplifier, hence to the loudspeaker, and so on, is an example of this, and it can easily be demonstrated.

DEMONSTRATION
I10 Feedback in public address systems

ITEM NO.	ITEM
1519	operational amplifier unit with power supply
	either
1510	potentiometer, 1 MΩ
	or
1017	resistance substitution box
157	microphone
1151	small loudspeaker (not earpiece)
1000	leads

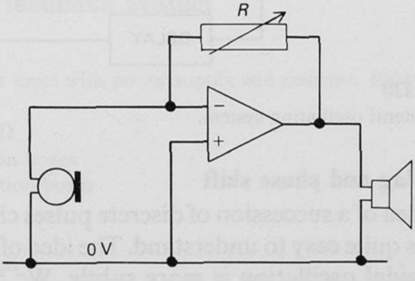
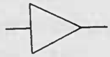


Figure I29
Acoustic feedback.

The value of R should be adjusted so that when the loudspeaker and microphone are about a metre apart, no oscillation occurs. (It may be necessary to cover the microphone with the hand to achieve this.) Scratching the microphone should give an audible sound from the loudspeaker. When the two are brought closer, oscillation should start, and it may be possible to get some idea of how the frequency changes according to the distance apart. If R is suitably reduced, oscillation will not occur, or may be heard dying away at some critical value. It may also be possible to adjust separation and gain so that oscillation only starts if a sharp sound (*e.g.* snapping the fingers) is made close by.

'What is happening here?' (Each pulse from the loudspeaker reaches the microphone causing a second pulse to be sent out, and the process is repeated.) 'What starts the oscillation?' (Any stray sound.) 'What determines whether the oscillation dies away or not?' (The gain of the amplifier, and also what fraction of the loudspeaker's sound is available to the microphone.) 'What limits the loudness of the oscillation?' (Perhaps saturation of the amplifier, perhaps energy dissipation in resistors.) 'What determines the frequency?' (There has to be a time lag between one pulse and the next – this time lag occurs in the travel of sound between the two transducers, and possibly also in the amplifier itself. The periodic time, T, of the oscillation depends on the delay.)

We can represent oscillating systems in general by the diagram of figure I30, in which A is the amplification factor, β (<1) is the feedback factor (feedback signal $= \beta \times$ amplifier output signal), and all of the delay is represented by the box so marked. (The symbol  represents a non-inverting amplifier.)

For oscillation to be maintained, the *loop gain* $A\beta$ must be at least 1.

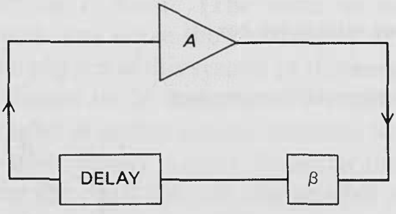


Figure I30
The general oscillating system.

Time lag and phase shift

The idea of a succession of discrete pulses chasing one another round the loop is quite easy to understand. The idea of some sort of steady stream of sinusoidal oscillation is more subtle. We hope that students will have learned from Unit D, 'Oscillations and waves', that such oscillation, though apparently more complicated (continuously varying acceleration, velocity, and displacement), is fundamentally more simple and dynamically more natural. How can we extend the 'delay' idea to cope with this?

Suppose we have a system like that in figure I30, but with a loop gain $A\beta$ less than 1, and suppose we inject an oscillating disturbance into it, as in figure I31.

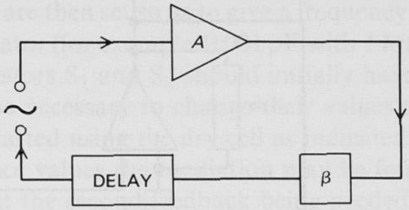


Figure I31
Forced oscillation.

If the injected signal has a periodic time equal to the time lag, how will the feedback and injected signals compare? They will be 2π out of phase – and therefore will reinforce totally. In the ensuing oscillation the energy loss will be balanced by the injected energy. What happens if we now change the frequency? The phase shift is no longer 2π, the injected and feedback signals are no longer in phase, reinforcement is only partial, so less energy is fed to the input and the oscillation will be feebler. If we switch off the injected signal, the oscillation will die away.

At this point we might remind students that they have seen something very like this before – resonance, studied in Unit D.

Questions

Questions 42 and 43 on time lag, phase shift, and superposition might be a useful preliminary here.

OPTIONAL DEMONSTRATION

I11 Oscillation in a feedback system

ITEM NO.	ITEM
1519	3 operational amplifier units with power supply and resistors, 100 kΩ, 100 kΩ or 1 MΩ, 1 MΩ
1510	potentiometer, *e.g.* 1 kΩ
1017	4 resistance substitution boxes
1018	2 capacitance substitution boxes
1109	signal generator
1511	oscilloscope
1033	cell holder with one cell
1000	leads

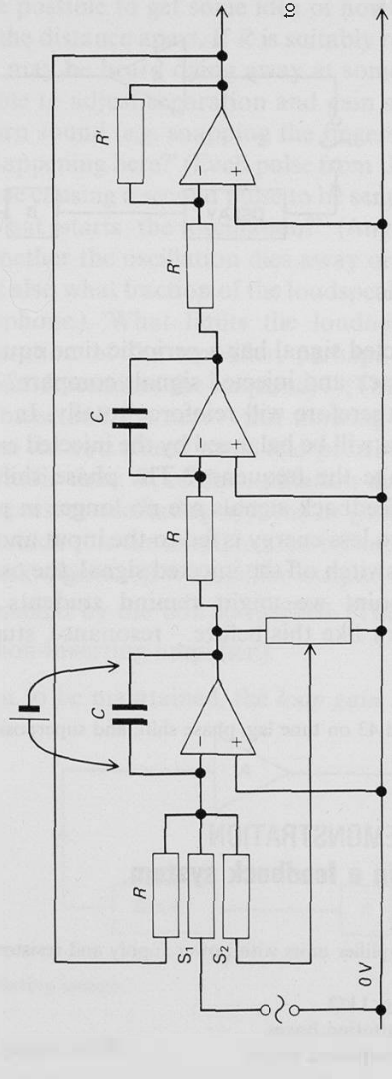


Figure I32
Forced oscillation circuit.

In the circuit of figure I32 both resistors R′ should be either 1 MΩ or 100 kΩ. The circuit is the same as for demonstration I9, except that an additional feedback loop is provided. The total feedback current into the first unit, representing d^2x/dt^2, now has two components, the new one being proportional to $-dx/dt$, representing 'velocity damping'. This should initially be set at zero, and the signal generator turned off. The values of C are then set so as to give a frequency within the range of the signal generator (for example, 0.001 μF with 1 MΩ gives a frequency of 160 Hz). Resistors S_1 and S_2 should initially have the same value as R, but it may be necessary to change their values to give best results. Oscillation is started using the dry cell as indicated in figure I32. With small capacitance values the oscillation may be found to be markedly damped without the second feedback being needed: in any case, fairly rapid damping should be ensured by adjusting this feedback if necessary. When the signal generator output is turned up and the frequency swept, a peaked frequency response should be apparent.

The effect of changing the damping can be explored, and with minimum damping it may be possible to see transient behaviour.

Note: it is advisable *not* to monitor the input from the signal generator using the second trace, since there are phase complications analogous to the phase relations that occur in mechanical systems, linked with the energy transfer between the driving and driven systems.

Finally, what happens if in our general oscillating system of figure I31, we increase A so as to increase the loop gain to 1? We have raised the energy supplied by the amplifier so that the net energy loss is zero – so the injected signal is no longer needed: the system will maintain its own oscillation. And the frequency of the oscillation? It is that for which the phase shift is 2π.

Textbooks call the kind of oscillator we used in demonstration I9 a *quadrature oscillator*, because the output of the second unit is in quadrature ($\pi/2$ out of phase) with the output of the first unit. The third unit is an inverter, so the total phase shift around the circuit in the three units is $\pi/2 + \pi/2 + \pi = 2\pi$; the signal fed back to the input is in phase – the feedback is positive.

In practice frequency generators use different types of basic oscillator circuit; the phase of the output of an amplifier can be changed in a variety of ways, using networks of R and C or of R and L. Oscillators using these techniques are called *phase shift oscillators*.

The above discussion on feedback and oscillation was intended to bring out two ideas: *1* the significance of phase shift as a development of the idea of time lag, and *2* the way in which any feedback system with delay, even if it does not itself generate sustained oscillation, may still respond to disturbances of appropriate frequencies, and may show 'damped oscillation' behaviour when disturbed. These ideas will be met again in the next section.

CONTROL SYSTEMS

Negative feedback leading to stability and delayed feedback leading to oscillation are not restricted to electronics. We now go on to take a brief look at these and some other important notions concerned with control which have general application in a range of systems including biological and economic systems as well as in engineering, manufacture, and everyday household appliances. These ideas are increasingly important to engineers and designers, and, whether we are aware of it or not, such systems are becoming more and more a part of the world we live in. For these reasons we think students should have a chance of a first meeting with the ideas, but the pressures on an A-level physics course mean that no more than an introduction will be possible. It seems right to link this with the work on electronics for two reasons: mechanical control devices (like Watt's classic two-ball steam engine governor) are increasingly being replaced by electronic ones; and, as has been said, many of the same general concepts can usefully be used in both control systems and electronic circuits.

Much of the suggested teaching can be done by question and answer in the classroom. Because the material may be unfamiliar to teachers, from time to time in the subsequent pages we give examples of the kinds of question we think students should be able to answer in classroom discussion (with appropriate answers in parenthesis).

Note on timing

Because we have given a fairly detailed set of teaching suggestions, the remaining part of this Section occupies a large number of pages, and this could give a misleading impression of the amount of time it should take to teach. The ideas presented are not essentially difficult, and what occupies several pages of print may sometimes require only a few minutes of actual class discussion time.

To give a gradual introduction to the new concepts required, we suggest first considering a system with *no* feedback – an open-loop system.

Systems with no feedback – open-loop systems

Examples one might discuss here include a room in which a heater is on all the time and a car (with no brakes) whose accelerator pedal is locked in a fixed position.

'What will happen to the temperature of the room or the speed of the car?' (They will reach steady values: the temperature in the room will rise until it reaches a steady value and the car will accelerate until it reaches a steady speed.)

'Why doesn't the temperature or speed increase any further?' 'What determines the final value?' (Temperature stops changing when rate at which energy is supplied by heater equals rate of energy flow to surroundings and speed levels off when power to overcome wind resistance, etc., is equal to power supplied by engine.)

'Would these final values be the ones we wanted?' (Maybe – but only if heater is of appropriate power and if accelerator setting was well chosen.)

'Suppose the weather changes, or the road begins to go downhill?' (Temperature, speed change.)

To summarize and introduce some useful general terms: systems like this in which the *input* is fixed (power of heater, accelerator setting) cannot produce a fixed *output* (temperature in room, car's speed) if there is a *disturbance* to the system (change in temperature, gradient of road). Such systems are represented by a simple block diagram (figure I33).

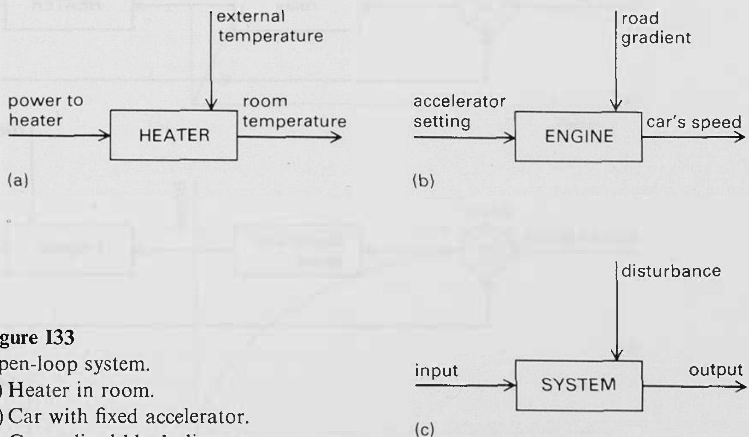


Figure I33
Open-loop system.
(a) Heater in room.
(b) Car with fixed accelerator.
(c) Generalized block diagram.

Systems with feedback – closed-loop systems

Students will be eager to point out how to improve the crude systems just discussed (put a thermostat in the room or a driver in the car).

'What are the essential functions of the thermostat and the driver?' (Thermostat compares temperature in room with preset value and driver looks at speedometer to see if car's speed is appropriate to road conditions.)

'Is that all? What else must be done?' (Thermostat trips microswitch or operates relay so that heater goes off if room is a few degrees above thermostat setting, and is switched on if temperature falls a little below and driver lifts or lowers his foot.)

These ideas can be generalized. A system in which the output is to be *controlled* – held at a predetermined value in spite of disturbance – requires feedback of information. The output signal is *compared* with the *reference* signal (*e.g.* the desired temperature or speed) by an *error detector* or *comparator.*

Any difference between the feedback and reference signals gives rise to an *error signal* which in turn acts perhaps through a *control unit* to alter the power to the *plant.* Sometimes several functions are combined: the driver acts as error detector and controller. Figure I34 shows block diagrams of some closed-loop systems.

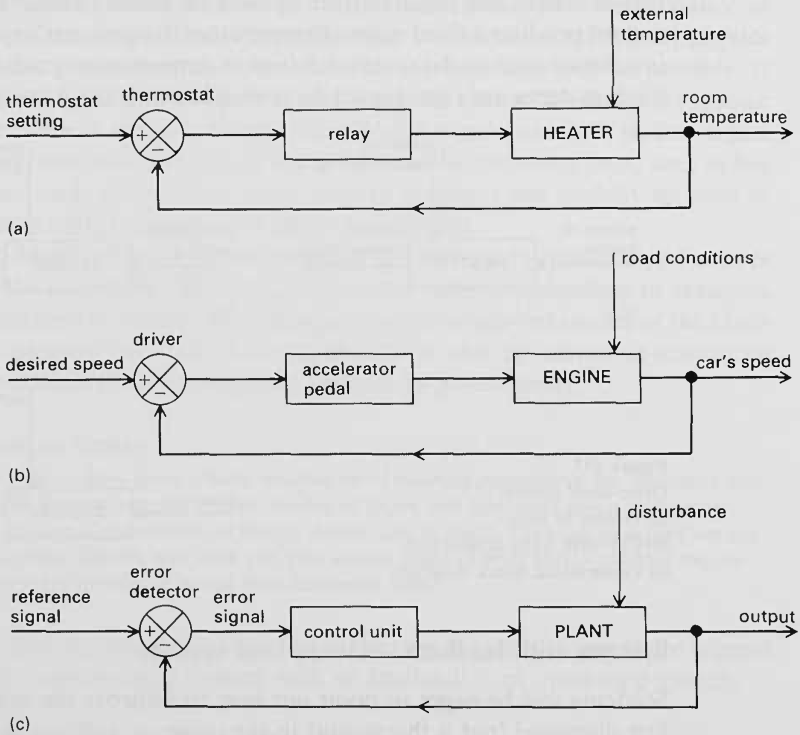


Figure I34
Closed-loop systems.
(a) Heater with thermostat.
(b) Car with driver.
(c) Generalized block diagram.

This way of thinking about systems can be applied very generally, and students can be invited to see how it fits familiar systems (lavatory cistern, electric iron, refrigerator, etc.), though as the example of the driver in the car shows, it is not always easy to separate the different elements.

On–off and continuous control

The same examples could be used to introduce another general idea. The heater in an iron is either on or off, but the power output of a car engine is continuously variable. How does this difference affect graphs of temperature or speed against time? (Temperature oscillates about the reference level, the thermostat setting; but given a good driver, speed will stay constant; figure I35.)

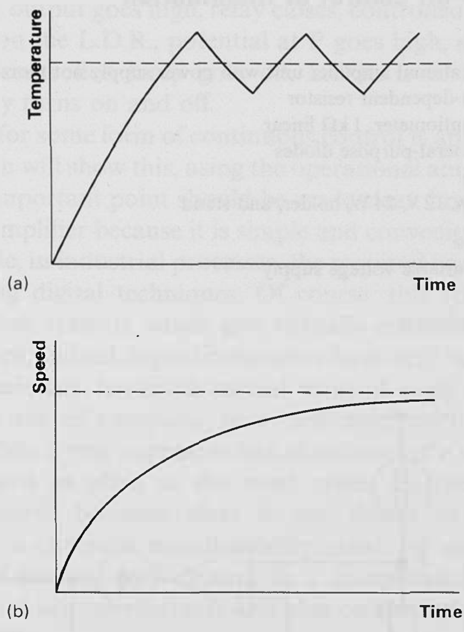


Figure I35
(a) Room temperature against time: on–off system.
(b) Car's speed against time: continuous system.

We might imagine a car controlled by an on–off system: a fixed throttle setting and an on–off switch. The ride would be quite jerky, but it would work, in principle. An absurd example could be used to illustrate that an on–off system cannot always be used. A lighting system is required to maintain a minimum level of illumination in a room, making use of available daylight whenever possible. The lights have been off during the day; during the evening daylight entering the room declines, and illumination falls below the reference level.

'What will happen if the lights are controlled by an on–off system?' (All lights go on.)

'Total illumination (daylight + room lighting) now exceeds reference level. What happens?' (Lights go off.) 'And then?' (Illumination is below reference level, lights go on, ...)

If the room also has blinds which are made to come down automatically when daylight *exceeds* reference level we have an even more chaotic situation!

DEMONSTRATION
I12 On–off control of illumination

ITEM NO.	ITEM
1519	operational amplifier unit with power supply not more than +6, 0, −6 V (see below)
1147	light-dependent resistor
1510	potentiometer, 1 kΩ linear
1151	2 general-purpose diodes
1513	relay
94A	lamp, 12 V, 24 W, holder, and stand
	either
59	l.t. variable voltage supply
	or
27	transformer
1000	leads

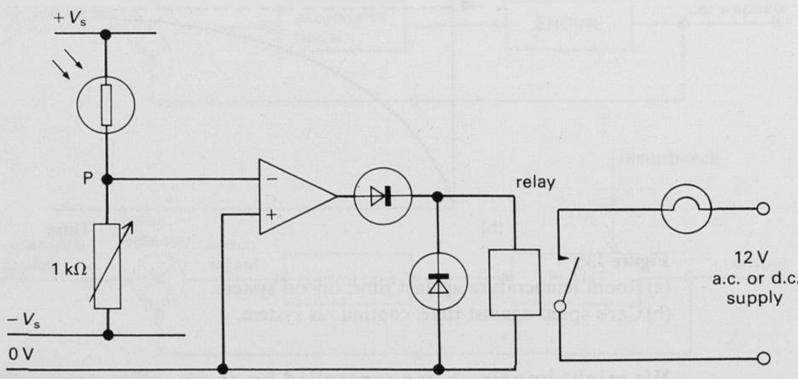


Figure I36
On–off control of illumination.

The variable resistor and light-dependent resistor (L.D.R.) are connected in series across the power supply, which needs to be +6, 0, −6 V because of the relay.

The controlled lamp should be 10–20 cm above the L.D.R.

It is convenient to work in a fairly dark room, using another lamp, *e.g.* a reading lamp, to vary the illumination on the L.D.R. With this lamp (or the room lights) on, set the potentiometer to about its mid-position. Adjust if necessary so that the controlled lamp is off when this light falls on the L.D.R. and comes on when it is in shadow.

In this demonstration we are using the operational amplifier simply as a digital device – a NOT gate.

When no light falls on the L.D.R. (figure I36), its resistance rises, the potential at P falls, the output of the operational amplifier goes high, the relay closes, and the lamp comes on.

Show that while light falls on the L.D.R. the controlled lamp is off, and it comes on when the L.D.R. is covered. What happens when the room lights are switched off (or the reading lamp is removed)? Potential at P goes low, output goes high, relay closes, controlled lamp comes on, its light falls on the L.D.R., potential at P goes high, output goes low, relay opens, controlled lamp goes off, L.D.R. is in darkness ... the system rapidly turns on and off.

The need for some form of continuous control is apparent. The next demonstration will show this, using the operational amplifier as a linear device. One important point should be made clear, however: we use an operational amplifier because it is simple and convenient for us, but on the larger scale, in industrial processes, the required control is normally achieved using digital techniques. Of course, this refers to complex digital electronic systems, which give virtually continuous variation by very small steps. Indeed, digital computers have very largely taken over the tasks previously requiring special 'control units'. A few lines of program in a shared computer, or a 'dedicated' microprocessor, now replace what was a very expensive box of electronics – and the program can be changed as often as the need arises. Further, the need for computer control becomes clear if one thinks of more complex situations. In a chemical manufacturing plant, for example, the optimum rate of stirring may depend in a complicated way upon the temperatures of several reactants and also on the output flow rate.

DEMONSTRATION
I13 Continuous control of illumination

See demonstration I1b on page 107.

Reading

This might be a good moment to draw students' attention to the article 'The scope and relevance of control engineering' in the Reader *Physics in engineering and technology.*

Questions

Questions 45 to 48 are about other control systems.

OSCILLATION IN CONTROL SYSTEMS

'Oscillations of temperature above and below the reference level were acceptable in the heating system – why not in the on–off lighting system?' (Temperature variations above and below reference level were slow and small, scarcely noticed; light level fluctuates rapidly and to extremes.)

'Both are on–off systems: why does one react so much more quickly, and therefore go to extreme values? Why does the heating system respond slowly?' (Because of the thermal capacity of the room – or to be more precise, because of the relationship between the room's thermal capacity and the power output of the heater – a very powerful heater in a very small room would switch on and off rapidly, and if the heater itself had a high thermal capacity the room would continue to heat up even after the power supply had been switched off.)

The last point can be generalized: the *inertia* of a system (mass in a mechanical system; thermal capacity in a heating system; inductance in an electrical system; ...) helps determine how quickly it will react and how much it will oscillate.

Oscillation can also occur in continuously controlled systems. An example, in which a very long time delay causes the feedback to be positive, might make this clear. One might discuss a thermostatically controlled central heating system which is intended to maintain a constant temperature, day and night. There are no preset time switches.

'What would be the effect of a twelve-hour delay in the feedback and control system? After sunset more energy begins to be lost from the building: the temperature drops. When will the heater be turned up?' (Twelve hours later – just as the outside temperature is rising.) 'Soon the temperature in the building rises. Will the heater be turned down?' (No, not until twelve hours later – about sunset time!)

Thus a system which was intended to maintain a constant temperature in fact amplifies oscillations in the disturbance.

Or suppose that due to a crop failure or manufacturing disaster one year, a particular commodity, X, is in short supply. Prices rise. Farmers and manufacturers decide to switch to growing or making X. But crops take time to grow, and factories can't be converted to a new product immediately. The shortage lasts a while – perhaps a year or two – until suddenly the market is flooded with X. Prices plummet. Farmers and factory owners get out of X. Another shortage. Prices rise. ... The oscillations again are due to delay in feedback.

We could represent this kind of situation in a block diagram (figure I37). Suppose some disturbance causes the output to oscillate slightly. At an instant when the output is above the reference level, the error signal should be negative so that the system will react to decrease the error. But suppose there are delays somewhere in the feedback loop such that by the time the compensation is applied to the output signal the delay amounts to half a cycle of the disturbing signal. The error 'compensation' will now cause the system to increase the output still further. Because the error detector *subtracts* the feedback from the reference signal it constitutes a phase change of π – this is negative feedback. If the other delays in the feedback and control elements amount to another π then the total delay will be 2π and the system will oscillate.

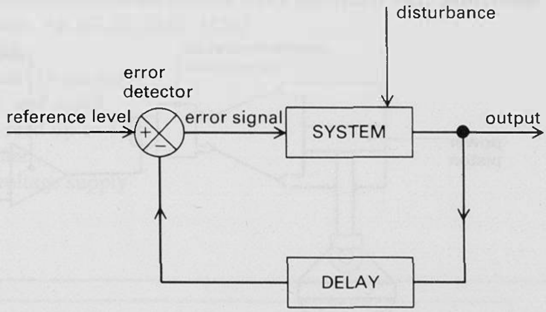


Figure I37
Feedback with delay.

Comparison with a simple mechanical system might help. A child sits on a swing which is at rest in its equilibrium position. The child manages to displace the swing slightly – a small disturbance has entered the system. A parent can cause the oscillations to grow or diminish depending on how he times his pulses on the swing. Pushing exactly out of phase with the oscillations will cause them to die away; pushing in phase will make them grow. The effect of friction (damping) will be to reduce the effect of in-phase pushes.

A better example than the very slow acting central heating system might be a steel rolling mill (figure I38).

The mill produces sheets of steel by passing the white-hot metal between heavy rollers.

To maintain constant thickness the gap between the rollers is continuously monitored, in this case using a source of beta particles and a detector on opposite sides of the sheet of steel coming out of the mill. The signal from the β-gauge is compared with a preset value (reference signal). If there is any error signal then the control elements of the system – in this case powerful motors – rack the rollers up or down to correct the error. There are (at least) two reasons why this system – designed to give negative feedback – could go wrong. Firstly, there is clearly a time delay (transport lag) between the occurrence of any disturbance at the rollers and its detection by the β-gauge. Delay is also introduced by the inertia of the system: it takes a small but finite time for the servo-valve to operate, and probably a much greater time to set the power piston in motion. This delay could be so great that when a disturbance enters the system and the roller gap departs from its proper level, the feedback is positive – the 'correcting' action adds to the initial disturbance. Oscillations begin to build up – just the reverse of the intention. One way to reduce this is to introduce some mechanical damping (*i.e.* friction) into the system.

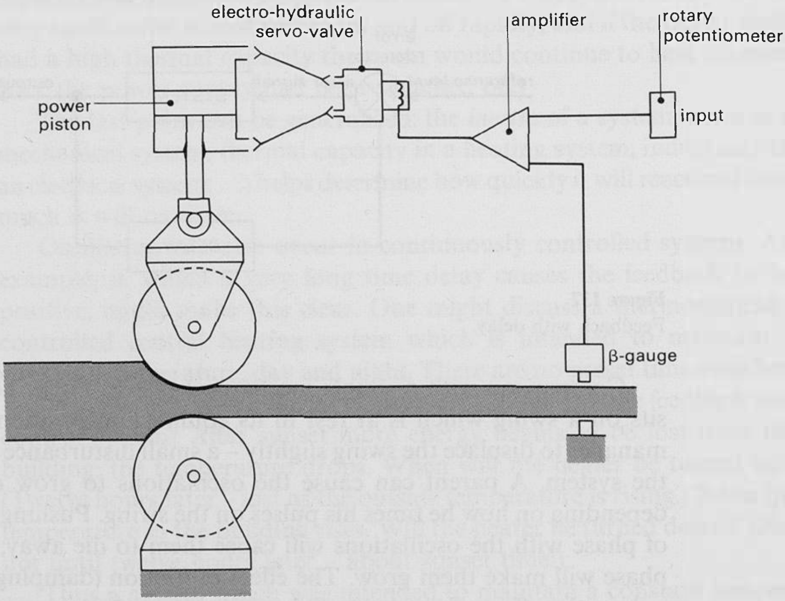


Figure I38
Rolling mill.
Based on HASLAM, *J. A.,* SUMMERS, *G. R., and* WILLIAMS, *D.* Engineering instrumentation and control. *Edward Arnold, 1981.*

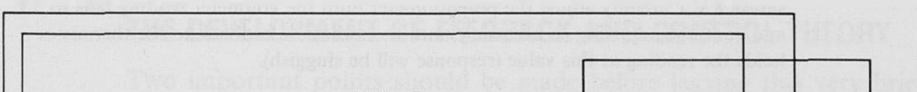
The way in which a system's response depends on damping and sensitivity (gain) is a study involving precise mathematical analysis (see for example HASLAM, SUMMERS, and WILLIAMS, *Engineering instrumentation and control* or (more advanced) JONES, *Instrumentation, measurement and feedback*). Some idea of this may be gathered from the article 'The scope and relevance of control engineering' in the Reader *Physics in engineering and technology.*

We can illustrate some of the aspects in the following demonstrations I14 and I15.

DEMONSTRATION
I14 Temperature control with thermal inertia

ITEM NO.	ITEM
1519	operational amplifier with power supply and resistors, 1 kΩ, 100 kΩ, 100 kΩ, 100 kΩ
1510	potentiometer, *e.g.* 1 kΩ linear
	either
1051	capacitor, *e.g.* 0.001 μF
	or
1018	capacitance substitution box
1017	resistance substitution box
1151	bead thermistor, *e.g.* GL23, 2 kΩ–115 Ω
1153	aluminium foil
10Y	dull black paint (Aquadag)
94A	lamp, holder, and stand
1507	voltmeter to read up to 12 V
1520	power amplifier
59	l.t. variable voltage supply
1000	leads


(b)



(a)

Figure I39
(a) Temperature control with thermal inertia.
(b) Thermistor mounting.

Preparation

Cut a piece of cooking foil about 2 cm × 1 cm, fold it in half and press it tightly round the thermistor as shown in figure I39(b). Paint one side dull black. Mount the thermistor level with the 24 W lamp, black surface facing the lamp and about 1 cm away from it.

Notes

i The behaviour of this circuit will inevitably depend on such local factors as the degree of thermal contact between thermistor and foil, and on the precise characteristics of the power amplifier, so some modifications may have to be imposed on the notes given below.
ii A possible variation is to cover the thermistor and lamp with an inverted beaker, representing a miniature 'room' being heated. The transfer of energy from lamp to thermistor will then probably depend more on convection than on radiation.

Circuit

The circuit is shown in figure I39(a). The operational amplifier supply voltage can be any value down to +3, 0, −3 V. R_A should be the resistance substitution box, set initially at 100 kΩ. R_B remains at 100 kΩ. The 'set level' adjustment of the potentiometer is very sensitive and can be disturbed by capacitive effects due to hand proximity: the feedback capacitor serves to eliminate this disturbance, and its value is not critical. Long leads will be needed to the thermistor and to the lamp.

Adjustment

Set the power amplifier supply to zero volts. Turn the potentiometer to maximum positive potential, then raise the power amplifier supply voltage until the voltmeter shows 8 V. Carefully adjust the potentiometer until the voltmeter reading falls to 5 V, and continue adjusting as necessary until the radiation feedback to the thermistor holds the reading at this value (response will be sluggish).

Demonstration

Now step up the power voltage by a few volts. The voltmeter reading should rise, then be corrected within a few seconds back almost to 5 V. Further upward steps of power voltage can now be made up to 25 V, and the feedback should continue to return the voltmeter reading fairly closely to 5 V.

The sequence should now be repeated with R_A at 47 kΩ, 22 kΩ, and 10 kΩ, and this time the system should show damped oscillation at each change of power voltage, and also if the set level is changed from 5 V to an adjacent value.

If the gain is raised to about 100 ($R_A = 1$ kΩ), the system should simply switch on and off indefinitely – it has become 'on–off' instead of 'continuous'.

Explanation (*for class discussion. This might be the task of a student, as a result of the questions posed in the students' laboratory notes*)
Radiation falling on the thermistor warms it. If the power of the lamp is increased by a change in its supply voltage, the thermistor becomes

hotter and its resistance falls. This makes the potential applied to R_A become more positive, lowering the output of the operational amplifier and so dimming the lamp. This continues until the potential is restored to its original value, *i.e.* until the output of the lamp is restored. Changing the setting of the potentiometer alters the temperature at which change ceases.

When the lamp is made brighter, the thermistor does not at once respond – time is needed for its temperature to rise. This introduces a delay, and with high gain results in oscillation.

The same comments apply here as are printed at the end of the notes on demonstration I1b, under Note *ii* (page 110).

DEMONSTRATION
I15 Light follower

See demonstration I1c on page 110, and the reference to satellite orientation on page 106.

Question

Question 49 is about oscillations in control systems.

THE DEVELOPMENT OF FEEDBACK AND CONTROL THEORY

Two important points should be made before leaving this very brief glimpse at a very broad subject.

The attempt presented here to give an overview of a large area has been entirely qualitative. This tends to misrepresent the subject: the mathematical aspects of it are now well developed. (It is interesting to note that it was started by Maxwell, who analysed mathematically the performance of the well known steam engine governor invented by Watt. The device had been successfully used for years before its performance was understood mathematically. But years later the mathematics was available to be applied to a very different system – repeaters for telecommunication cables; an interesting example of the interplay between invention, theory, and development.) The rapid growth in the subject, now known as cybernetics, started after the Second World War. The fact that the same mathematics applies to quite different systems illustrates the wide ranging power of the ideas involved.

It is also important to realize that although the discussion in this Unit has emphasized linear systems, the tendency nowadays is more and more towards digital or numerical control systems. These points come over very strongly in the article 'The scope and relevance of control engineering' in the Reader *Physics in engineering and technology.*

THE SCOPE OF FEEDBACK AND CONTROL

As a way of rounding off the discussion of feedback and control and showing how the same ideas apply in a wide variety of fields, we give some brief and simplified notes on some examples that might be discussed, perhaps as a team teaching venture, or perhaps by students who are also taking biology or economics being invited to lead the discussion of the examples. (NUFFIELD ADVANCED PHYSICS Students' book *Unit 6*, pages 14 to 27, provides a useful basis.)

Biology

Organisms, and particularly their constituent cells, can function normally under only a fairly restricted range of conditions. Maintaining these conditions (control) in spite of changes in environment (disturbance) is therefore vital. The process is called homeostasis, 'staying the same'; it depends on negative feedback. In extreme conditions the negative feedback may break down, positive feedback taking over with disastrous results, as described below. Biologists tend to talk about receptors, control mechanism, and effectors rather than error detectors or comparators, control elements, and plant.

Temperature control The hypothalamus (a structure in the brain) responds to changes in blood temperature and to information from thermoreceptors in the skin. If the temperature rises above the set point or reference signal, several changes occur: blood flow to skin capillaries increases, increasing the rate of cooling; sweating starts, causing cooling by evaporation; metabolic rate decreases, reducing energy conversion in the body. If the temperature falls below normal, the reverse changes occur: less blood to the skin; no sweating; increased metabolic rate; shivering. At low temperatures, mammals' fur and birds' feathers rise, increasing insulation; dogs (no sweat glands) pant when hot, increasing the rate of evaporation of water via the lungs. Above 'high critical temperature', regulation breaks down, body temperature rises; metabolic rate increases (approximately doubling for 10 °C temperature rise, like most chemical processes); faster metabolism releases more energy which raises the temperature further.

The system is remarkably efficient: humans can survive for 30 minutes even at 120 °C if the air is *dry* (*e.g.* sauna); but the lethal limit for *body* temperature is 42 °C.

Breathing and circulation Functioning of enzymes is very sensitive to the pH of the blood; excess CO_2 in the blood makes it more acidic (danger of breathing expired air, *e.g.* from polythene bag). Respiration and cardiovascular centres in the brain are sensitive to the level of CO_2 in blood (chemoreceptors in certain arteries, *e.g.* carotid bodies in the aorta close to the heart, are also sensitive to the level of O_2, but this is apparently less important). Efferent (motor) nerves from these centres control breathing rate and heart beat respectively. Increase in either increases the rate at which CO_2 is removed from blood in the lungs. The rate of breathing doubles if the concentration of CO_2 in blood rises from 0.04 % (normal) to 3.0 %. Positive feedback occurs when breathing oxygen at pressure above atmospheric: increase in oxygen in the blood leads to an increase in metabolic rate (to lower oxygen concentration). But faster metabolism means more CO_2 in the blood, hence faster breathing and *more* oxygen in the blood, and so on.

Blood sugar and insulin The normal level of glucose in blood plasma is about 100 mg per 100 cm^3. Higher levels stimulate secretion of insulin from the pancreas; insulin promotes conversion of glucose to glycogen in the liver; and the blood sugar level falls. If blood sugar falls too low other hormones (glucagon, adrenalin) are secreted. These act on glycogen stored in the liver to produce glucose which enters the blood stream. In the absence of insulin, blood sugar rises. Above about 160 mg per 100 cm^3, sugar is excreted in urine: a symptom of diabetes mellitus, treated by injection of insulin.

Many control processes depend on hormones like insulin. Often a complex series of interactions between enzymes is involved.

Population growth and control Simple organisms (*e.g.* bacteria) in a rich environment (no limit on food, etc.) may show exponential growth (rate of growth $\propto$ number present, an example of positive feedback, modelled in demonstration I8b, Integrator with feedback). In real environments growth is less rapid when competition for food, worsening of conditions, etc. become important factors. Then negative feedback keeps the population nearly steady: increase in population means increase in competition so the population falls; decrease in population means less competition, and the population rises. Real situations are invariably complicated, involving several species. Predator–prey interaction is another negative feedback mechanism. Removal of a check to growth (*e.g.* by improving health or increasing food supply) can lead to exponential growth until a new limit is reached. Human population appears to be a classic example of this.

Reading

ELECTRONIC SYSTEMS TEACHING PROGRAMME ESP700 Book 4 *Feedback systems.*
HARDY *Homeostasis.*
NUFFIELD REVISED ADVANCED BIOLOGY *Study guide I.* Chapter 2, and sections 10.5 and 11.1.
ROBERTS *Biology, a functional approach.* 2nd edn. Chapters 13, 15, 16, and 32.

Economics

Economic theory developed as an attempt to understand the periodic booms and slumps that seem to be unavoidable. One early theory related trade cycles to sun spots (sun spots affect weather, hence agricultural production, hence food supply, hence prices, ...). The notes below relate to some mechanisms operating in capitalist economics.

Supply, demand, and prices: the cobweb theory If demand for a particular commodity exceeds supply, the price tends to rise. Higher prices attract more firms to make (or farmers to grow) the product. Some delay occurs before the increased supply becomes available. When increased supply reaches the market, it may exceed demand: prices fall if other circumstances (*e.g.* consumers' tastes, price of other goods) have stayed the same; production is cut back. If the supply is now less than the demand, prices tend to rise again. Fluctuating prices are the result of imbalance between supply and demand.

Capital and consumer goods: the accelerator Capital goods (*e.g.* machinery, tractors, ships) are used to manufacture and move consumer goods (*e.g.* toothpaste, tomatoes, textiles). Suppose a factory needs one machine to make 1000 pairs of trousers per year. Total output is 10 000 pairs which matches demand, so ten machines are needed. If each machine lasts 10 years, and the factory replaces one per year, the machine-making factory has a regular order for one trouser-making machine per year. If the demand for trousers drops by 10 % (imports are cheaper, fashions change, ...), the trouser factory will produce 9000 pairs and need only nine machines. For one year it need order no new machine, so there is no work from it for the machine-making factory. Small instability (10 % reduction) in one part of the system has a major effect (100% reduction) in another. Ship building is an example: a small change in the amount of goods to be shipped results in a big change in orders for new ships. The reverse is also true: a small increase in production of consumer goods can cause big increase in demand for capital goods.

The circular flow of income Goods and services are exchanged for money. In a very simplified view, firms produce goods and services, individuals provide labour. Workers are paid for their labour, and spend money on goods and services produced by firms. Money flows one way, goods and services the other (figure I40).

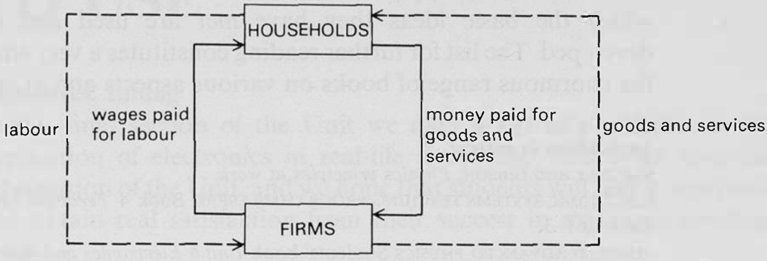


Figure I40
Based on LIPSEY, R. G. An introduction to positive economics. *George Weidenfeld & Nicholson, 1963.*

Of course the system is much more complicated: for example, firms require investment (new capital equipment, etc.); individuals are paid not only wages but profit, interest, etc., and may save (figure I41).

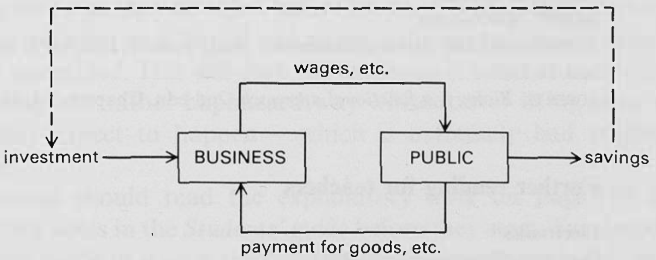


Figure I41
Based on SAMUELSON, P. A. Economics. *McGraw-Hill Book Co., 1961.*

Investment leads to increased production: more money in the system; savings decrease the amount of money in circulation until returned as investment. Equilibrium is only possible if savings match investment.

Real systems are, of course, very much more complex than suggested here.

Television or video

The programme 'Control systems' (television programme 1 of Open University course T391 *Control engineering*) is very relevant. It shows examples of control systems in glass manufacture, a power station, and in satellite tracking.

Question

Question 50 is about the 'self-fulfilling prophecy'.

Reading for students

This Unit is a brief introduction to a very wide field. We therefore recommend that *all* students should at least glance at the books named in the list below, simply to ensure that they are clear about the ways in which the basic ideas they have met are used and subsequently developed. The list for further reading constitutes a very small sample of the enormous range of books on various aspects and at various levels.

Applications in general

BARCLAY and GIBBON, *Physics principles at work.*

ELECTRONIC SYSTEMS TEACHING PROGRAMME ESP700 Book 4 *Feedback systems*, Sections 1–3.

NUFFIELD ADVANCED PHYSICS Students' book *Unit 6 Electronics and reactive circuits*, pages 14–34.

Control engineering

BOLTON, *Engineering instrumentation and control.*

ENGINEERING CONCEPTS CURRICULUM PROJECT, *The man-made world.*

ENGINEERING SCIENCE PROJECT, *Electronics, systems, and analogues.*

PROJECT TECHNOLOGY Handbook 14, *Simple computer and control logic.*

RAMSEY, *Engineering instrumentation and control.*

Biological systems

HARDY, *Homeostasis.*

NUFFIELD REVISED ADVANCED BIOLOGY *Study guide I.* Chapter 2, and sections 10.5 and 11.1.

ROBERTS, *Biology, a functional approach.* 2nd edn. Chapters 13, 15, 16, and 32.

Further reading for teachers

Electronics

CLAYTON, *Operational amplifiers.*

FOXCROFT, *Operational amplifier.*

GOUGH *et al.*, *Notes for guidance on the electronics option.*

HOROWITZ and HILL, *The art of electronics.*

Control engineering

HASLAM, SUMMERS, and WILLIAMS, *Engineering instrumentation and control.*

JONES, *Instrumentation, measurement and feedback.*

SECTION I3

PUTTING ELECTRONICS TO USE

Objectives, timing

In this final section of the Unit we offer a list of problems on the application of electronics in real-life situations. This is an essential culmination of the Unit, and we hope that students will find it enjoyable and obtain real satisfaction from their success in making something work.

For this reason it is important that sufficient time be allowed. As a minimum, each student or pair of students should be able to tackle two problems, and also have time to see something of what others have achieved. Problem-solving involves planning and usually a considerable amount of trial and error, so it would be quite wrong for this exercise to be rushed.

We can also help students to make the most of this section by making sure that they do think before they act. Some teachers may well consider insisting on a 'pencil and paper' solution before any apparatus may be assembled. This will curb the tendency of some students to start assembling in a rather haphazard way without ever being clear about what they expect to happen – which is extremely bad engineering practice!

Students should read the explanatory note on page 119 of the laboratory notes in the *Students' guide* before they start. Two important points are made in it, one emphasizing 'plan before you build', and the other pointing out the status and function of the 'engineer' (an often misunderstood title). We suggest that students' attention be drawn positively to this note before they start to select their problems.

Choice of problems

The selection of problems offered has been based on two considerations. The first is that of apparatus availability. It would be very tempting at this point to suggest an extensive list of extra items of apparatus to support a range of ambitious control experiments, but schools' resources are limited. The decision therefore has been to restrict the problems to ones which can be solved using only apparatus which is required elsewhere in the course. The second consideration concerns the range of difficulty of the problems. Some must be sufficiently straightforward so that even weaker students have the chance of experiencing success, whilst others need to offer more of a challenge. However, the challenge must be kept within bounds, and it would be wrong to include in the list problems which could only be solved on a basis of knowledge beyond what has been included in this Unit and earlier parts of the course.

There is no rigid grading of difficulty in the list, but the division into 'A: easier' and 'B: harder' should help in the making of appropriate choices. It is to be hoped, of course, that teachers will add to the list in the light of their knowledge of individual students, and possibly also because of the availability of extra equipment in the school.

In the list given in the students' laboratory notes we have tried to provide essential hints where it was thought necessary. The list below, for the teacher's use, provides:

possible solutions;
notes on special apparatus requirements;
notes on some special precautions or difficulties; and
some suggestions for extending problems.

Reading for students and teachers

MARSTON, *110 operational amplifier projects.*
PLANT, *Operational amplifier applications.*

INDIVIDUAL TASKS

I16 Putting electronics to use

All tasks require

ITEM NO.	ITEM
1519	operational amplifier unit and power supply
1000	leads

Access to the following

1033	cell holder with dry cells
1510	potentiometers, *e.g.* 1 kΩ linear
1151	selection of resistors
1151	selection of capacitors
1040	clip component holder
1017	resistance substitution box
1018	capacitance substitution box
1147	light-dependent resistor
1507	voltmeters, 1 V d.c., 10 V d.c.
1507	ammeter, 100 μA d.c.
92R	m.e.s. lamp, 2.5 V
92T	m.e.s. holder
1151	diode
1151	thermistor
1059	earpiece
157	microphone

1151	small loudspeakers
1039	search coil
1058	coil with 120 + 120 turns
1030	coil with 1100 turns
92G	C-core
1109	signal generator
504	iron rod
27	transformer
1511	oscilloscope
9B	small electric motor
1513	relay
52L	press-switch or mounted bell push

The above list clearly cannot include all the things that a student doing a particular task might ask for, especially if, as we hope, the tasks suggested are treated only as examples.

List A: easier

A1 Digital-to-analogue converter

This will require four resistors with values in the ratios 1:2:4:8.

A2 Light meter

An inverting amplifier with a light-dependent resistor as input resistance (figure I42).

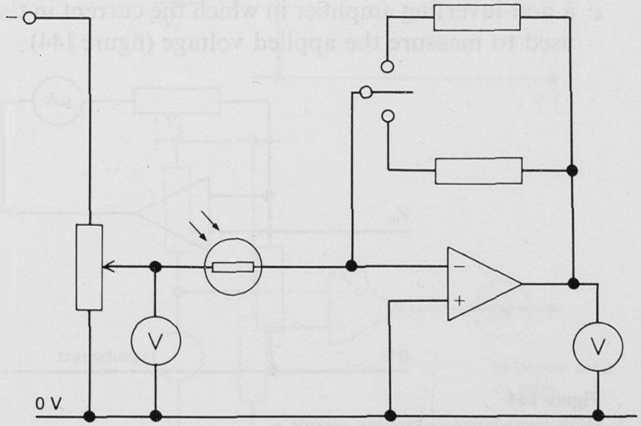


Figure I42
Light meter.

The input supply could be the negative rail or a separate battery. The input voltmeter is for setting a standard input voltage. The potentiometer should take as little current as possible, but its resistance should not be higher than that of the meter.

A3 Linear ohmmeter

The unknown resistor is used as the feedback resistor in an inverting amplifier (figure I43).

The same remarks apply as for *A2*.

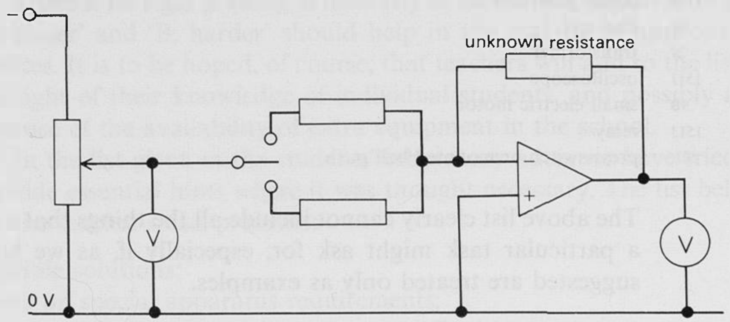


Figure I43
Linear ohmmeter.

A4 High-impedance voltmeter circuits

Three possibilities are suggested in the students' laboratory notes:

a a simple follower;
b an inverting amplifier (with three ranges);
c a non-inverting amplifier in which the current in the feedback loop is used to measure the applied voltage (figure I44).

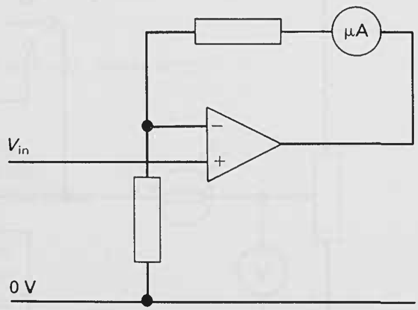


Figure I44
High-impedance voltmeter, circuit *c*.

A5 The comparator as a switch

The student is referred to question **37** for the comparator circuit. The suggestions are:

a switching on bollard lights as daylight fades;
b high- or low-temperature warning;
c high water-level warning;
d steam warning for a kettle.

For all four, the resistive transducer and 'set level' control could be connected across the + and − supply to the operational amplifier in one of the ways suggested in figure I45.

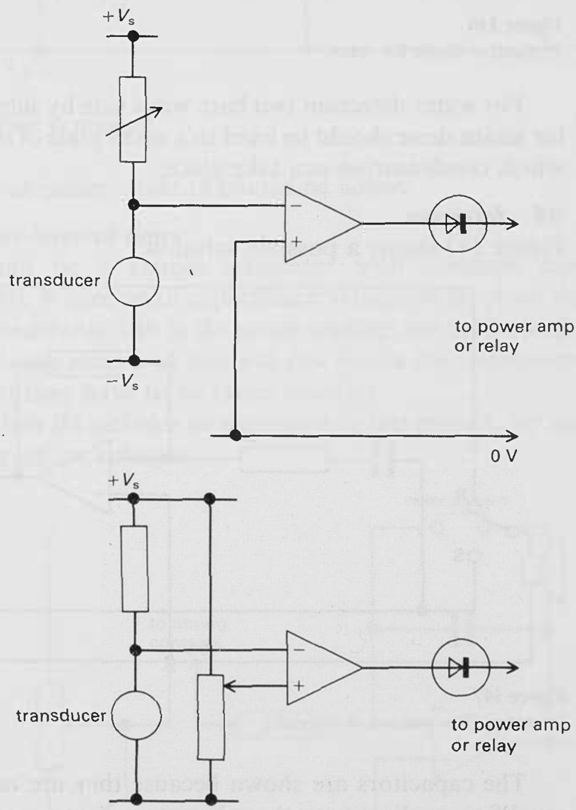


Figure I45
Comparator switching circuits.

Note the need for a diode in the output to the power stage: a relay must operate on only one side of the set level, and a power amplifier may need to be protected from reverse input to its transistor(s).

Remember also that a relay should have a diode connected across its coil to accommodate the inductive surge which occurs when it is switched off suddenly (figure I46).

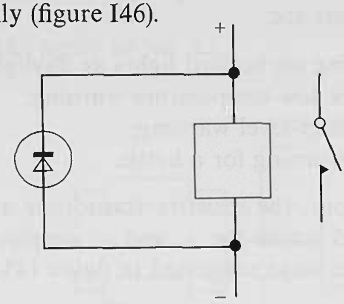


Figure I46
Protective diode for relay.

For water detection two bare wires side by side should be tried, and for steam these should be fixed to a small plate of insulating material on which condensation can take place.

A6 Intercom

Figure I47 shows a possible solution.

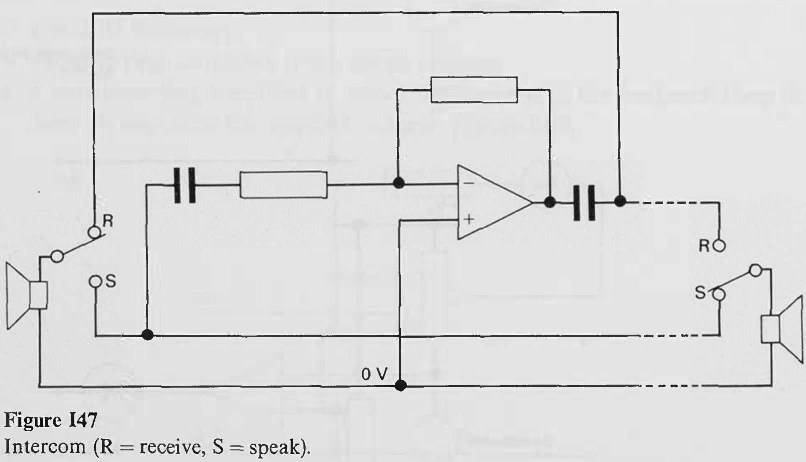


Figure I47
Intercom (R = receive, S = speak).

The capacitors are shown because they are necessary in most a.c. amplifier applications, though not in fact essential here. 0.1 μF is suggested, though the output capacitor could be of higher value. The amplification factor needed can readily be found by trial and error.

A7 Mixer

See figure I48 and the comments concerning capacitors under *A6* above.

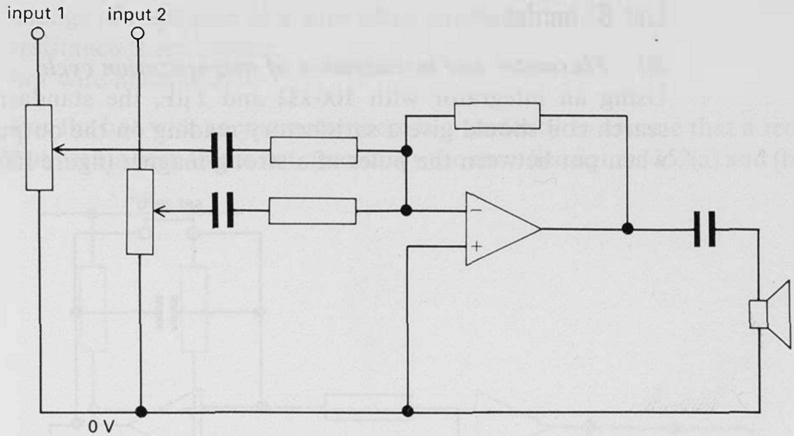


Figure I48
Mixer amplifier in principle.

A power amplifier could of course be added.

A8 Short-interval timer

This could be a simple integrator with constant input voltage (figure I49). Where small capacitance values are involved there is liable to be considerable drift in the meter reading, but on the scale of the time interval being measured this will not vitiate the measurement reading (though it may have to be taken quickly).

Problem *B5* includes an extension to this project, for automatically switching off an enlarger.

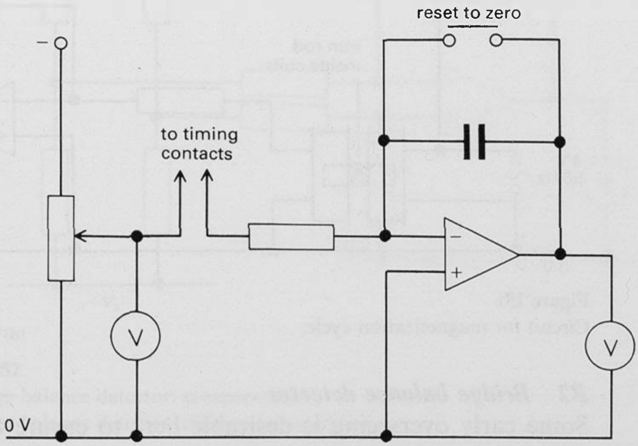


Figure I49
Short-interval timer.

List B: harder

B1 Fluxmeter and investigation of magnetization cycle

Using an integrator with 100 kΩ and 1 μF, the standard 5000-turn search coil should give a satisfactory reading on the output voltmeter when put between the poles of a strong magnet (figure I50).

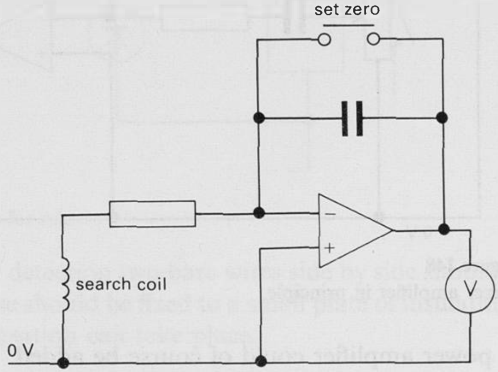


Figure I50
Fluxmeter.

The main point of this task, of course, is for the student to realize the need to perform an integration operation.

For the magnetization cycle, the information in the students' laboratory notes should be sufficient guide. See figure I51.

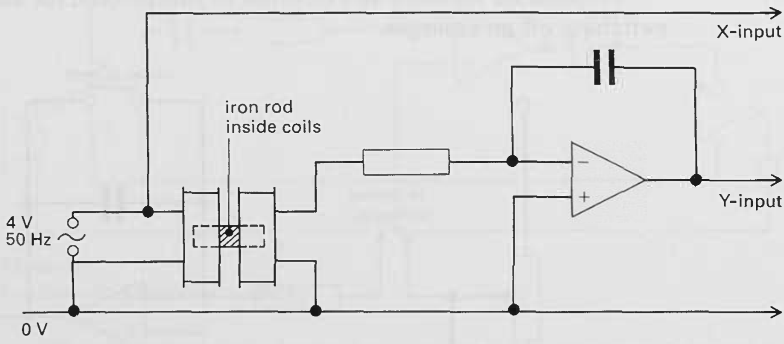


Figure I51
Circuit for magnetization cycle.

B2 Bridge balance detector

Some early overseeing is desirable here to ensure that unrealistically low resistances are not connected so as to give damagingly large currents. The suggested applications are:

change of resistance of a wire when strained;
resistance thermometer;
hot wire anemometer.

For the hot-wire anemometer there is a tendency to assume that a red-hot wire is best: this is not necessarily the case. See figures I52(a) and (b).

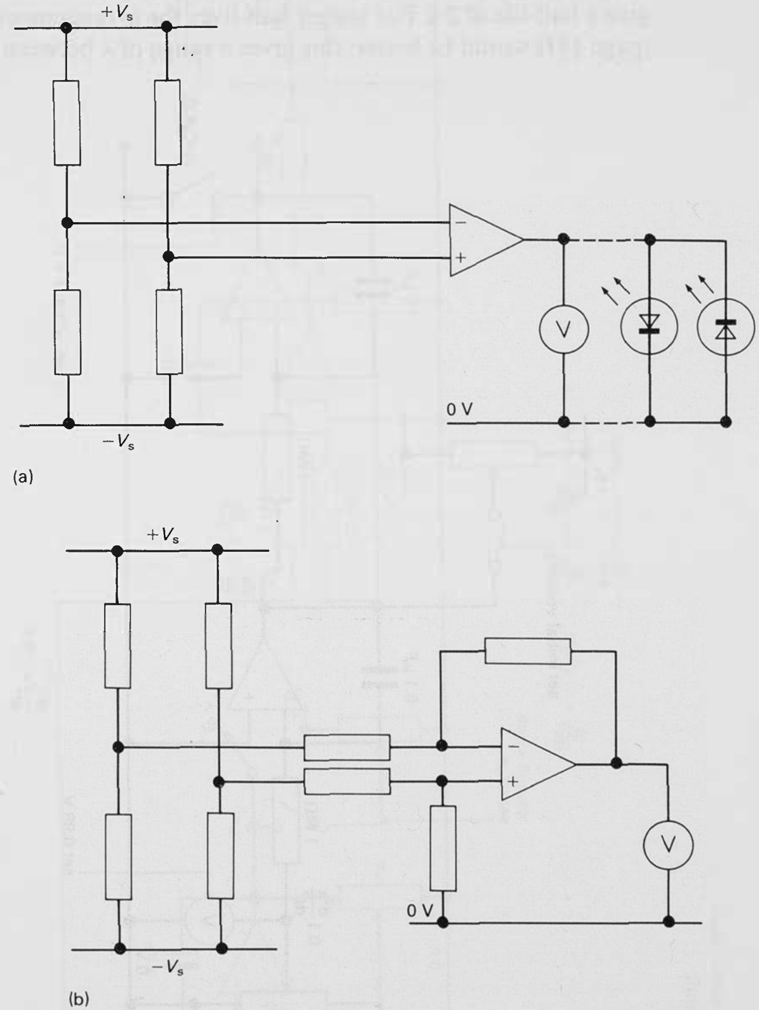


Figure I52
(a) Bridge balance detector: comparator.
(b) Bridge balance detector: differential amplifier.

B3 Analogue computing

Correct scaling (*e.g.* to give $g = -9.81\,\mathrm{m\,s^{-2}}$ exactly) should be a secondary consideration here; the important thing is to get the correct functional relationships. Possible solutions are shown in figure I53. In (b) and (c) (pages 179 and 180) the feedback fractions, λ, are the corresponding decay constants. Using $t_{\frac{1}{2}} \approx 0.7/\lambda$, a value of 0.35 for λ would give a half-life of 2 s. For longer half-lives the arrangement of figure I54 (page 181) would be better: this gives a range of λ between 0.1 and zero.

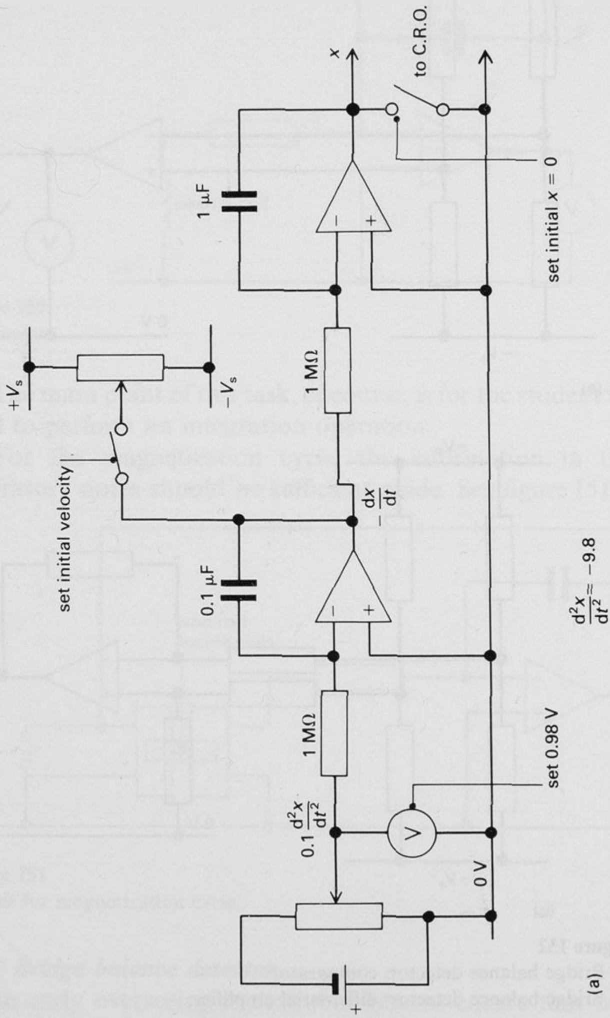
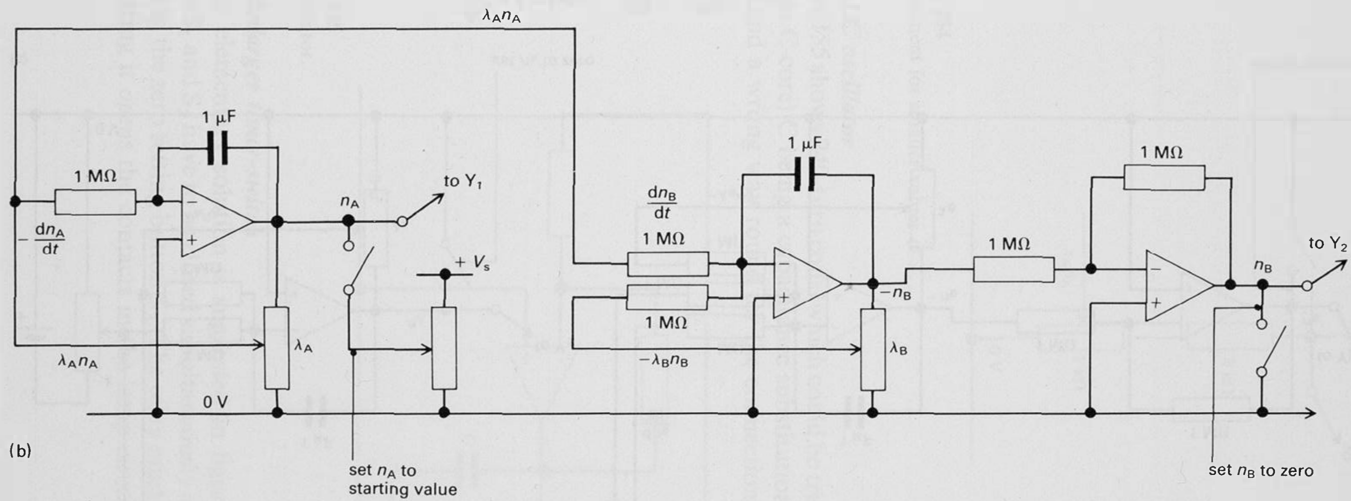


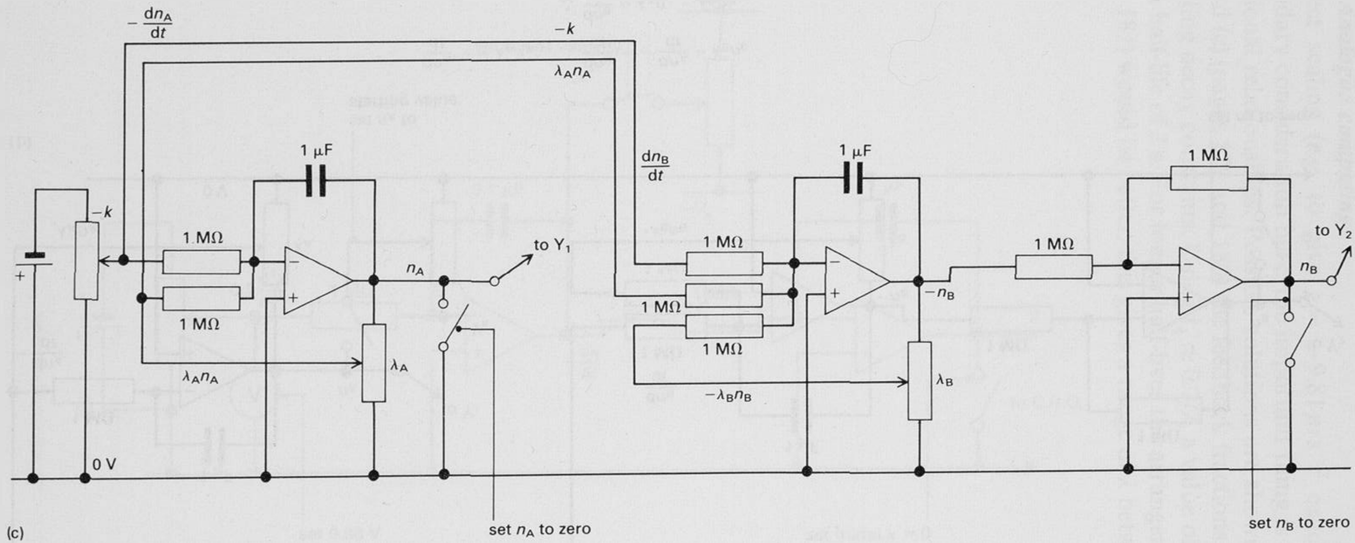
Figure I53 (part)
Analogue computing:
(a) Motion under gravity.



$$\frac{dn_A}{dt} = -\lambda_A n_A; \qquad \frac{dn_B}{dt} = -\frac{dn_A}{dt} - \lambda_B n_B$$

$$\therefore \frac{dn_B}{dt} = \lambda_A n_A - \lambda_B n_B$$

Figure I53 (part)
Analogue computing:
(b) Two-member decay series.



$$\frac{dn_A}{dt} = k - \lambda_A n_A; \qquad \frac{dn_B}{dt} = -\frac{dn_A}{dt} - \lambda_B n_B$$

$$\frac{dn_B}{dt} = -k + \lambda_A n_A - \lambda_B n_B$$

Figure I53 (part)
Analogue computing:
(c) Build-up of two daughter products.

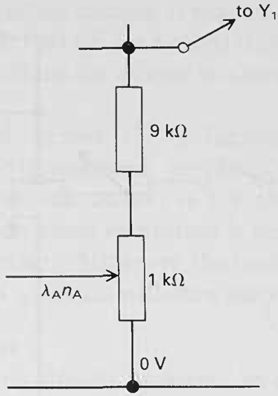


Figure I54
Arrangement for smaller values of λ_A.

B4 LC oscillator

Figure I55 shows 240-turn coils (which could be tried with and without a double C-core) C being a capacitance substitution box. There will be a right and a wrong way round for the connections to the transformer.

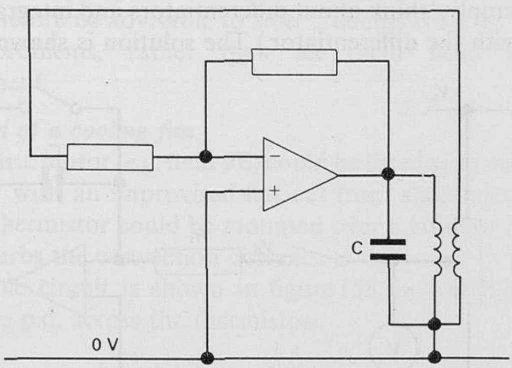


Figure I55
LC oscillator.

B5 Enlarger timer-switch

A very elementary solution is suggested in figure I56 (page 182), in which S_1 and S_2 have to be closed simultaneously and immediately after pressing the zero setting button S_3. The relay must be connected so that energizing it *opens* the contacts in the lamp circuit.

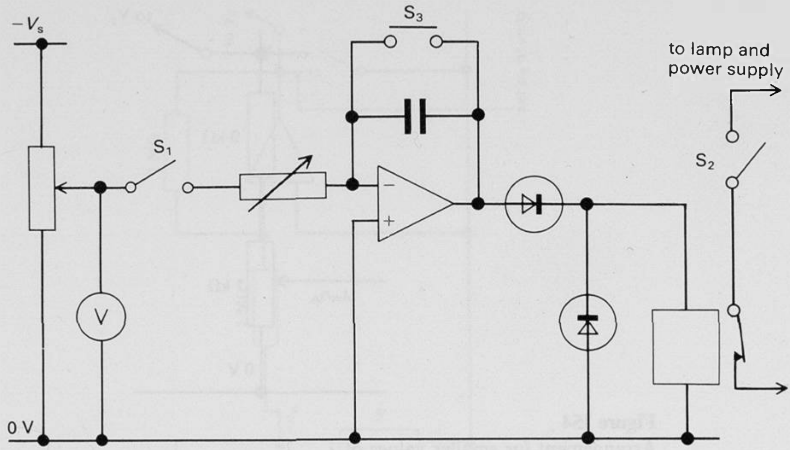


Figure I56
Enlarger time switch.

B6 Linear capacitance meter

This is a challenge for the more mathematically-minded students, and it is likely that some graded hints may be needed. The first might be simply, 'think about differentiators and integrators'. (Question **27** deals with the differentiator.) The solution is shown in figure I57.

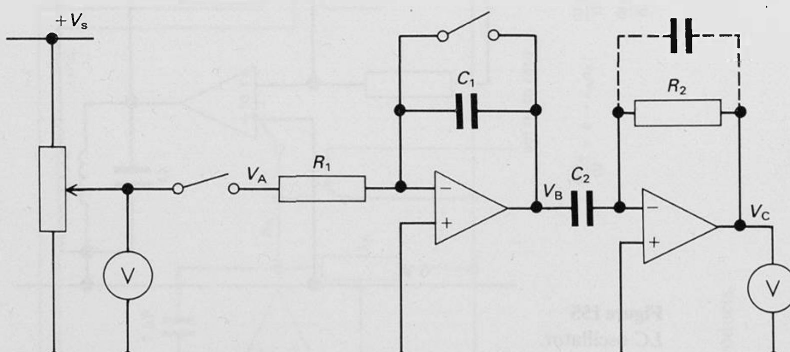


Figure I57
Linear capacitance meter.

When V_A is switched on, V_B falls at V_A/R_1C_1 volts per second. V_C is (rate of fall of V_B) $\times$ R_2C_2, and so is proportional to C_2. C_2 is therefore the capacitance being measured. Ranges can be chosen by choice of the values of the other components.

As with all differentiating circuits, it may be advisable to include the extra capacitor (perhaps 0.01 μF for a start) shown dashed in figure I57, for reasons of stability. Since the output is a steady voltage this will not affect the reading.

Note that the p.d. across C_1 is increasing steadily and will eventually reach the operational amplifier's limiting value. For example, if R_1C_1 is 1 second and V_A is 1 V, then V_B will be falling at $1\,\mathrm{V\,s^{-1}}$. In a few seconds, when saturation is reached, the output meter reading falls abruptly to zero. However, the reading can be recovered by momentarily shorting C_1, which will then start to charge up as before.

B7 Control of a heater

One example of this has already appeared as demonstration I14. The difficulty in a school laboratory is that one can work only on a small scale. The standard 12 V immersion heater can give a rapid rate of temperature change only in a very small volume of water. In a static tank, convection currents give large temperature inequalities, and for constant flow work the heat capacity of the heater itself makes for very sluggish response. Heating air by means of a 12 V lamp is faster, but again convection currents complicate the situation, and a thermistor near to a lamp will respond to radiation rather than to air temperature. It might therefore be advisable to steer students towards an awareness of these problems, rather than let them head for probable disappointment.

B8 Control of a cooling fan

A small electric motor (*e.g.* item 9B) could be fitted with a model aircraft propeller or with an improvised fan cut from sheet metal. For quick response a thermistor could be mounted over a lamp or heater so that the fan disturbs the convection currents.

A possible circuit is shown in figure I58, in which the voltmeter monitors the p.d. across the thermistor.

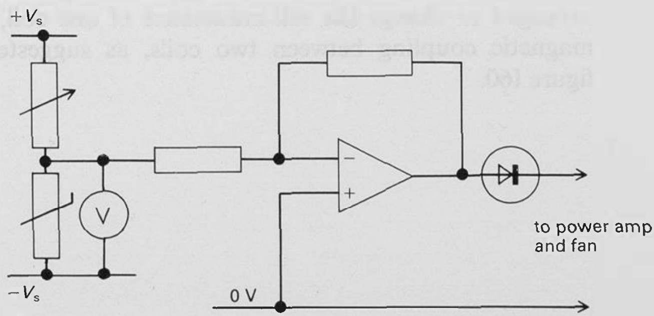


Figure I58
Controlling the temperature of a thermistor using a fan.

B9 Linear variable transformer

'Linear' here implies that the device responds to straight-line motion, but not necessarily strictly proportionally. With the circuit shown in figure I59 and using an iron rod about 6 cm long, readings from almost zero to several volts are obtainable as the rod is moved. The signal generator should be set at about 100 Hz, using the low-impedance output at a few volts. Any general-purpose diode is suitable as the rectifier. A possible application might be to have the rod vertical and suspended on a spring with a scale pan attached, to display weight on a meter.

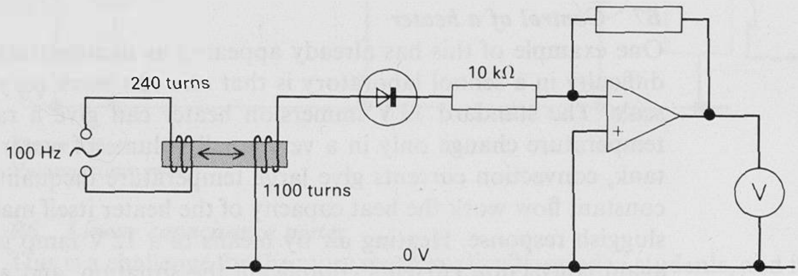


Figure I59
Linear variable transformer.

The linear variable differential transformer (l.v.d.t.), which is widely used in industry, has two secondary coils, one on each side of the primary and connected in antiphase; they are long coils, giving a truly proportional output. This requires a phase-sensitive detector, *i.e.* one which gives positive or negative output according to the phase of its input. An interested student might try to devise such a detector.

B10 Variable-reluctance transducer

Basically this means a variable air gap in a magnetic circuit. It could be arranged to change the self-inductance of one coil, or to change the magnetic coupling between two coils, as suggested in principle in figure I60.

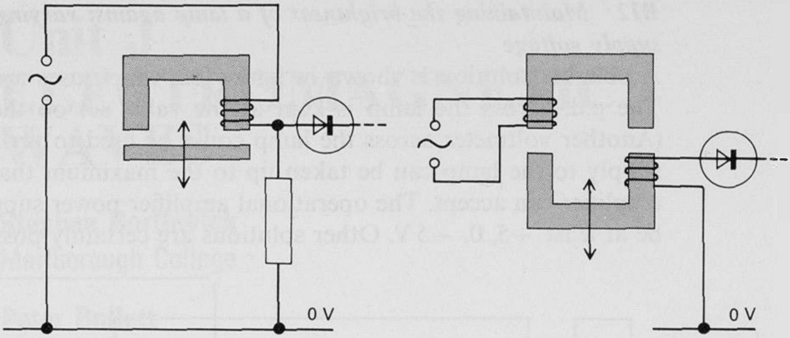


Figure I60
Variable-reluctance systems.

B11 Astable multivibrator, transducer-controlled frequency

The student is referred to question **38**, which leads to the circuit of figure I61.

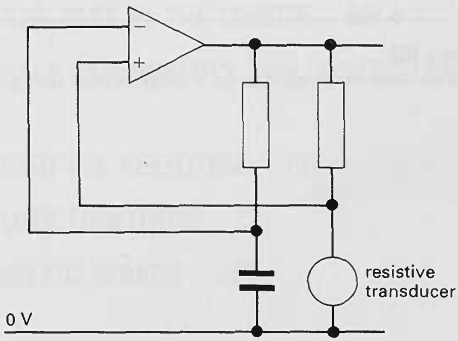


Figure I61
Transducer-controlled astable multivibrator.

B12 Maintaining the brightness of a lamp against varying supply voltage

A possible solution is shown in figure I62, which uses a comparator. The p.d. across the lamp is kept at the value set on the voltmeter. (Another voltmeter across the lamp could be used to verify this.) The supply to the lamp can be taken up to the maximum that the power amplifier can accept. The operational amplifier power supply needs to be at least +5, 0, −5 V. Other solutions are certainly possible.

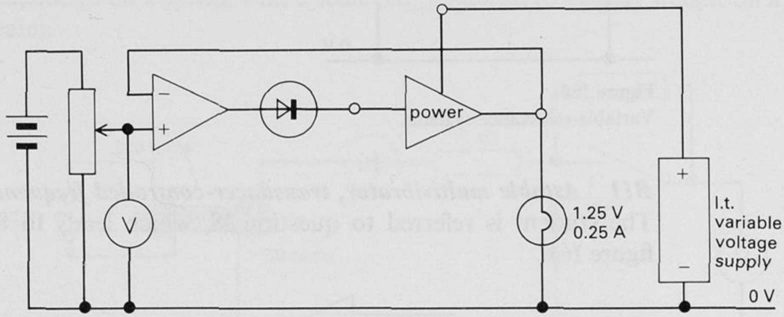


Figure I62
Maintaining the brightness of a lamp against changes in supply voltage.

Unit J
ELECTROMAGNETIC WAVES

Stephen Borthwick
Marlborough College

Peter Bullett
Rugby School

Suggested time allocation: four weeks

PLAN OF THE UNIT

Section J1
Waves through an aperture

using electromagnetic wave apparatus (Unit D) ►	**Experiments with diffraction at an aperture for a variety of waves**	
(GCSE) ripple tank work ►	**Huygens's hypothesis**	
superposition, path difference (Unit D); energy of simple harmonic oscillator $\propto A^2$ (Unit D); energy of a.c. $\propto I_0{}^2$ (Unit H) ►	**Intensity (energy) of a wave $\propto A^2$**	► photons, electrons: Unit L, 'Waves, particles, and atoms'
	$n\lambda = b \sin\theta$ for minima in a single-slit diffraction pattern	
	Variation of intensity in a single-slit diffraction pattern	
	Diffraction patterns with different shaped apertures	► Section J2: holography
$\sin\theta \approx \theta$ for small angles (Unit D) ►	**Resolution of two point sources, Rayleigh's criterion**	
	Interferometry and radio astromony	

Section J2
Waves through gratings

(GCSE) ripple tank, Young's fringes, and grating experiments; estimation of λ	►	**Experiments with diffraction at a grating**		
		Observation of spectra	►	$E = hf$ and $mv = h/\lambda$: Unit L, 'Waves, particles, and atoms'
		$n\lambda = s \sin\theta$	►	electron diffraction: Unit L, 'Waves, particles, and atoms'
single-slit pattern: Section J1	►	**Resolving power of a grating**		
		Measurement of λ for light with Young's slits and with a grating		
		Reflection gratings		
		Diffraction at complex gratings		
		Holography		

Section J3
Electromagnetic waves

E-field between plates (Unit E)	►	**Geometry of guided (or 'tied') and free electromagnetic waves**
parallel plate capacitor (Unit E); flat solenoid (Unit H)	►	**The speed of a 'tied' pulse**
		The speed of light
transverse and longitudinal waves (Unit D)	►	**Polarization of a variety of electromagnetic waves**

INTRODUCTION

This Unit is concerned with electromagnetic waves: their nature and how they allow us to gain information about the world around us. It deals with the traditional topic of physical optics, though the emphasis, and to some extent the content, differ from those of many A-level courses. The Unit sets out to achieve two distinct, though complementary goals.

The first goal is to develop an appreciation of the image-forming process – the scattering of an incident wave by an object and the subsequent recombination of the scattered wave to form an image. The diffraction of light, X-rays, radio waves, water waves, etc. is discussed within the same conceptual framework. This provides a basis for an appreciation of modern techniques such as holography and image processing.

Such an approach may seem novel, but the major concerns of the first two Sections of the Unit are in fact familiar: the diffraction of a variety of waves at single and multiple apertures.

Students observe a variety of diffraction effects. An interpretation in terms of Huygens's hypothesis suggests a reason for the spreading of waves and prepares for the analysis of single-slit and grating diffraction patterns in terms of superposition and path difference. The treatment contains little mathematics, but allows a quantitative treatment of resolution and the resolving power of a grating. The qualitative understanding of diffraction and image recombination with complex apertures is extended to the principles of X-ray crystallography and holography.

So although the perspective of physical optics has been shifted, the content is conservative. The omissions forced by the limitations of three weeks' teaching time for the first two Sections of the Unit include geometrical optics (it is assumed students will have covered enough in earlier courses); lens defects (which can be described rapidly and are of less practical importance with the advent of inexpensive, high quality lenses); and the phasor analysis of the detail of single-slit and grating diffraction patterns. Students with mathematical fluency may appreciate the phasor approach and reference is made to a computer program which may be useful.

The second major goal is to give students an appreciation of the nature of the electromagnetic wave. It is often argued that such work should wait for a university course, when Maxwell's equations can show the 'mechanism' whereby an electromagnetic wave can propagate in free space. It seems a pity, however, to deny students the opportunity to see how a number of ideas developed earlier in the course can be brought together to provide a picture of the nature of electromagnetic radiation. This last Section of the Unit is intended to last no more than a week (about one-quarter of the suggested time for the Unit), and it would be wrong to claim too much for this very simplified treatment. We hope students will gain a sense of satisfaction that the two constants ε_0 and μ_0 met in earlier Units can be combined to predict the speed at which electromagnetic waves propagate along a rectangular waveguide. This treatment provides an opportunity to revise the work on parallel plate capacitors (Unit E, 'Field and potential') and the solenoid (Unit H, 'Magnetic fields and a.c.'), together with the idea that an e.m.f. can cause an electric current and hence generate magnetic flux within a system. In this sense, predicting c represents the end-point of a thread of argument which runs throughout the course.

No mechanism is suggested for the transition of the 'tied' wave of the waveguide to the freely propagating waves radiated by the dipole. Instead, the two waves are shown to have many properties in common, which suggests that the picture established of one may be a good way to think of the other.

Having shown that the speed of light is related to the fundamental constants of electricity and magnetism, it seems desirable to give students a taste of some of the consequences of the finiteness and constancy of c. The reading passage 'Relativity' could be assigned at the end of the Unit. A thought experiment and the 'k-calculus' show that two observers will not always agree about the interpretation of an event and explain the unexpectedly long life of rapidly moving π-mesons. It is a rare student who is not confused by Special Relativity, but all should appreciate the striking intellectual achievement of a theory which revolutionized physics in the early part of this century.

THE PLACE OF THE UNIT IN THE COURSE

We hope and expect that teachers will find their own way of using the material in this Unit. The detailed teaching programme suggested seems a logical way to present the topics: Section J2 depends on many of the ideas developed in Section J1; Section J3 is in many ways independent of the other two and could precede them.

Sections J1 and J2 depend on the work of Unit D, 'Oscillations and waves', while J3 uses ideas from Unit B, 'Currents, circuits, and charge', Unit E, 'Field and potential', and Unit H, 'Magnetic fields and a.c.'. While there could be an argument for covering the work of J1 and J2 earlier, the synthesis of J3 makes it unlikely that it will be tackled until well into the second year of the course.

The work on spectra in Section J2 and the amplitude–energy relationship for a wave in J1 are taken up again in Unit L, 'Waves, particles, and atoms', and so must necessarily precede it.

LIST OF SUGGESTED EXPERIMENTS AND DEMONSTRATIONS

SECTION J1
WAVES THROUGH AN APERTURE

DIFFRACTION

Diffraction effects inevitably accompany the passage of a wave through an aperture. We suggest starting with some simple class experiments. Suggestions from students that what they see has to do with 'waves' or 'superposition' are welcome, but not yet essential. The main point is to use the experiments to suggest what may be worth exploring later, by experiment and theoretical argument.

We see everything through a hole: the pupil of the eye. Some of the implications of having to look through a hole a few millimetres wide can be gained by looking through a yet smaller hole.

EXPERIMENT
J1 Looking at a lamp through a slit and through a pin-hole

ITEM NO.	ITEM
1167/1J	holder with two halves of a razor blade, to be used as a single slit
1167/3R	set of three colour filters (red, blue, green)
1153	aluminium foil
1153	35 mm slide mounts
1501	copper wire, 0.2 mm diameter, bare
1501	steel or nichrome wire, 0.2 mm diameter, bare
94A	lamp, holder, and stand
27	transformer
1167/3Q	matt white reflecting screens or white cards
1153	mains lamp with 30 cm single filament (clear tube) and holder

Safety note: The mains lamp must be safely mounted: no part of the fitting which is connected to the mains supply should be accessible to students; nor should they be able to remove the lamp easily.

Set up one or more single-filament, clear tube mains lamps around the darkened room with the colour filters attached, but not touching the glass envelope. See figure J1.

Students adjust the razor blades (figure J2) so that their edges form a parallel-sided slit, which is as narrow as possible, by viewing them against the background of an illuminated matt white screen. They then look through the slit at the single filament lamp from a distance of about three metres. Light from the filament is seen to spread widely.

The copper wire is stretched until it breaks; then one of the pieces is held a few millimetres from the break and the point is pressed slowly but firmly right through the aluminium foil, which may conveniently be held in a slide mount. The filament is viewed again through this small hole.

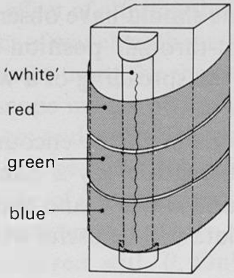


Figure J1
Single-filament lamp with colour filters.

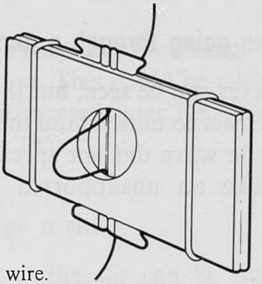


Figure J2
Adjustment of slit width to diameter of steel or nichrome wire.

If steel or nichrome wire of the same gauge is used to fix the separation of the razor blades (a broad loop should be gently clamped by the blades at the top and bottom of the slit on the viewing side and bent back, figure J2) the slightly wider diffraction pattern of the hole compared with the slit can point to the $\theta \approx 1.22\lambda/b$ of later work.

Students should be able to report that the pattern seen through a small aperture consists of light and dark regions (or colours, where no filters are present); that the shape of the pattern depends on the shape of the aperture (fringes parallel to the slit or concentric to the hole); that the smaller the aperture, the wider the pattern; and that long wavelength (red) light gives a broader pattern than short wavelength (blue) light.

In the following circus of experiments, students take a closer look at the diffraction of a variety of waves and make some rough measurements.

CIRCUS OF EXPERIMENTS

J2 Diffraction

The experiments are best shared out, each student investigating one (or two). After a double period of experimenting, students report what they did and what they saw to the rest of the class.

All students should have observed the first minimum on either side of the 'straight-through' position and should have gained an appreciation of how the spreading of a wave depends on wavelength and the size of the aperture.

Some groups might be encouraged to make measurements of slit width, b, wavelength, λ (by a standing wave method, refer to Unit D, 'Oscillations and waves'), and the angle θ at which the first minimum occurs. This data will be useful when the result $\theta = \sin^{-1} \lambda/b$ is derived later.

J2a Water waves going through a gap

As water waves can be seen, but the waveform of light cannot, students may find it easier to understand the diffraction of water waves than that of light. If the wave did *not* spread beyond the gap, each wave crest would be like an unsupported hill of water and would collapse outwards.

ITEM NO.	ITEM
90	ripple tank kit
47	illuminant
27	transformer
1033	cell holder with two cells
541/1	rheostat, 10 to 15 Ω
105/1	hand-held stroboscope

For details of the use of the ripple tank, see REVISED NUFFIELD PHYSICS, *Teachers' guide Year 3*, Class Experiment 4 (particularly 4q).

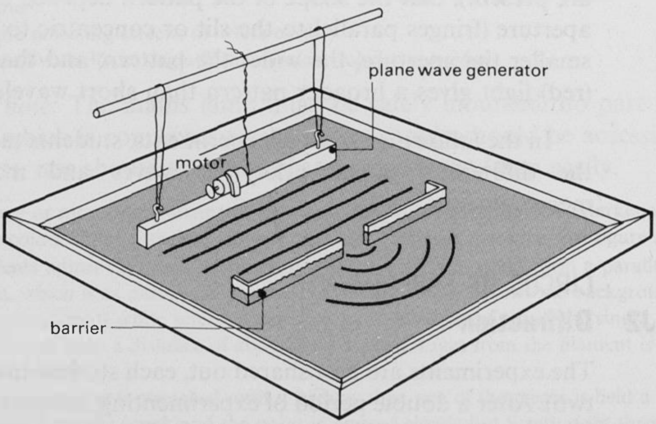


Figure J3
Diffraction of water waves at a gap.

Students first observe the effect of altering the gap size on the spreading of the wavefront. The effect of wavelength may be observed by changing the frequency of the generator using the rheostat. 'Freezing' the pattern with a hand-held stroboscope makes observations easier.

With a gap of about 5 cm and a wavelength of about 1 cm, directions of minimum disturbance can be seen radiating from the gap. By laying rulers on the screen the value of θ for the first of these directions may be measured. λ and b can be measured *as on the screen* to simplify analysis and the value of λ/b compared with θ. Only an approximate comparison can be made, but it should indicate that the relationship is worth pursuing further.

Very careful observers may spot the change of phase across the line of destructive superposition. This could be useful when the theoretical detail of the pattern is discussed later (see 'Explaining the single-slit diffraction pattern', page 205).

J2b Microwaves going through a slit

For microwaves ($\lambda \approx 3$ cm) the slit can be quite wide and still give substantial angles of diffraction.

ITEM NO.	ITEM
184/1	microwave transmitter
184/2	microwave receiver
181	general purpose amplifier
183	loudspeaker (if not part of item 181)
	wax lens or plastic lens filled with paraffin (if available)
1153	2 metal screens about 0.3 m × 0.3 m
501	metre rule
1000	leads

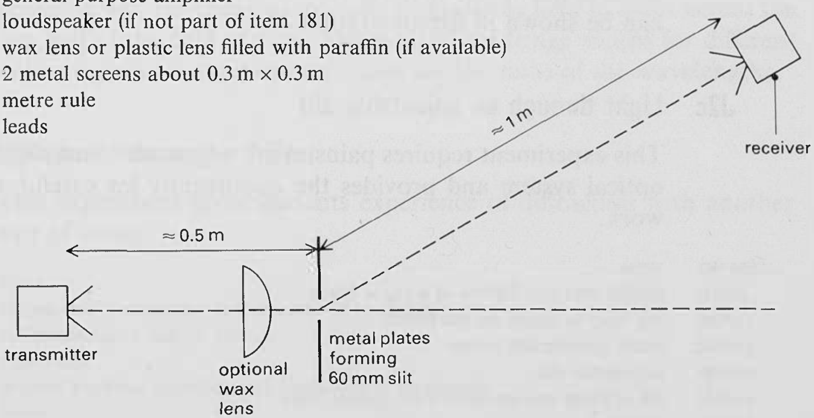


Figure J4
Diffraction of microwaves.

The transmitter should be behind the two metal screens which form a 60 mm slit. A wax or paraffin lens may be used to produce a more parallel beam by placing the transmitter at its focal point (about 0.5 m), but is not essential. As the receiver is moved around on either side of the straight-through direction the signal falls off and reaches a low value at about 30°.

Although the precision of the test is poor and the geometry approximate, rough agreement with the expected λ/b value can be shown.

Students can try a further, striking experiment shown in figure J5.

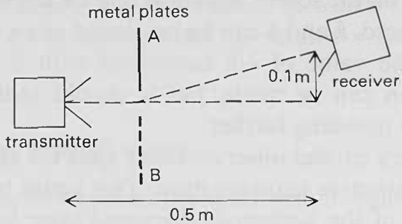


Figure J5
Superposition experiment.

The transmitter and receiver are placed about 0.5 m apart with the receiver offset some 0.1 m. Plate A is slid between them until it cuts off the signal. Then plate B is gradually slid into the beam until it is within 30 mm to 60 mm of plate A. The received signal will grow to a maximum.

As the second plate is pushed closer, *less* wave energy can pass through the gap, but the receiver output *increases*. It seems that waves from part of an extended wavefront can superpose destructively, since removing some of them *adds* to the net effect. A similar 'less can mean more' effect can be shown in demonstration J4.

J2c Light through an adjustable slit

This experiment requires painstaking adjustment and alignment of the optical system and provides the opportunity for careful, quantitative work.

ITEM NO.	ITEM
1167/1J	holder and two halves of a razor blade
1167/1G	big stop to stand on the bench
1167/1H	small translucent screen
1167/2M	adjustable slit
1167/3R	set of three colour filters (red, green, blue)
1167/2L	eyepiece
1167/1I	2 holders (for eyepiece and adjustable slit)
94	lamp, holder, and stand
27	transformer
1153	transparent ruler or 0.5 mm graticule
1167/1C	plano-convex lens, +2D, diameter 37 mm
1167/1B	holder for lens of diameter 37 mm

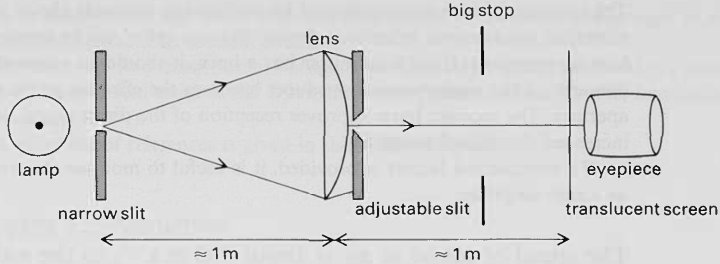


Figure J6
Diffraction of light by a single slit.

The lamp and screen are just over 2 m apart; the lens forms an image of the lamp filament on the screen. The adjustable slit is placed close to the lens, its length parallel to the lamp filament. As the slit is narrowed the image dims and then widens. The lamp may be over-run for short periods to make the image brighter and a big stop used to cut out light which has not gone through the slit. Its best position, somewhere between the slit and the screen, must be found by trial and error.

The pattern will be sharper if the razor blade slit (about 1 mm wide) is propped against the lamp, and the screen moved back to focus the image of this slit rather than the filament. Light from the razor blade slit must fall on the adjustable slit.

Students measure the fringe width for red, green, and blue light by measuring the separation of a number of maxima either side of the central fringe using an eyepiece and graduated scale. The eyepiece should be focused on the fringes formed on the screen; then the screen is removed and the transparent ruler or graticule held in focus across the top half of the field of view. The ratio of the fringe widths for different colours gives an approximate value for the ratio of the wavelengths.

J2d Ultrasound through a hole

This experiment gives students experience of diffraction with another sort of wave.

ITEM NO.	ITEM
	ultrasound transmitter and receiver (if available)
1507	milliammeter, 1 mA or 10 mA
501	metre rule
181, 183	general purpose amplifier and loudspeaker (optional)
1000	leads

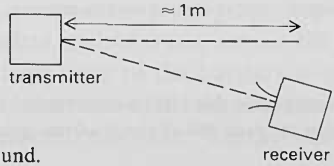


Figure J7
The diffraction of ultrasound.

The transmitter and receiver should be on benches or stools about one metre apart to minimize troublesome reflections. Some trial and error will be necessary to find the best arrangement. If the transmitter has a horn, it should be removed so that the diameter of the piezoelectric transducer becomes the effective width of the diffraction aperture. The receiver horn improves reception of the direct signal and gives some increased directional sensitivity.

If a modulation facility is provided, it is useful to monitor the received signal with an audio-amplifier.

The speed of sound in air is about 330 m s^{-1}, so the wavelength of the sound waves at a frequency of 40 kHz will be about 8×10^{-3} m, or 8 mm. If the measured width of the transducer (aperture) is 24 mm, the first minimum should occur at

$$\sin^{-1} \theta = 1.22\lambda/b = 1.22 \times 8 \text{ mm}/24 \text{ mm}$$

$$\Rightarrow \theta = 24°$$

In practice it is somewhat easier to measure the angle 2θ from the first minimum on one side to the first minimum on the other.

Discussion of results

When students report back on the experiment they tried, the important points to draw out are:

1 That in order to get an appreciable diffraction effect, b must be of the same order of magnitude as λ, as shown for a wide variety of types of wave and wavelengths. If $b \gg \lambda$, then there is little spreading. If $b \ll \lambda$, then the wave energy passing through the aperture is too small to observe easily, though it is sent out over almost 180°. In summary, spreading increases as λ/b increases.

2 That there are angles at which no energy appears to emerge; *i.e.*, the effect of an aperture for which $b \approx \lambda$ is to *redistribute* the incident wave energy.

Questions

Question 1 revises some ideas about superposition. Question 2 deals qualitatively with an example of diffraction of sea waves.

Home experiment

In home experiment JH1, A homemade slit, students look at a distant light source through a narrow gap between two fingers.

Reading

Many of the topics in this Unit are covered in the general list of text books for students given on page 348 of the *Students' guide*. Although references to specific pages or chapters are not given here, students should of course be encouraged to read the appropriate sections of their books.

Teachers will find much useful background material in HECHT and ZAJAC, *Optics*, TAYLOR, *Images*, and units of the Open University, Second Level Course, *Images and Information* (ST291).

A fuller list of references is given in this *Teachers' guide*.

Huygens's construction

The treatment suggested here is a *brief* introduction to Huygens's construction to make the spreading out of a wave when it passes through an aperture seem reasonable and to indicate a fruitful way to proceed in explaining the observed minima and maxima.

It is fairly simple 'rule-of-thumb' which predicts successfully how waves behave in a variety of circumstances, including reflection, refraction, and diffraction, and prepares students for dividing a wavefront into elements in the derivation of the single-slit diffraction formula.

It also provides a good example of the way that hypotheses can become principles as they are tested and modified against experimental observations. It might also begin a discussion of the seventeenth century controversy between the wave and corpuscular theories of light. The dual nature of light will be taken up in Unit L, 'Waves, particles, and atoms'. Reference to an historical outline is given on page 203.

The discussion can start from the diffraction of water waves in a ripple tank, experiment J2a, extended here as a similar experiment. Huygens assumed that each point on a wavefront acts as a point source, giving rise to so-called secondary wavelets which travel out radially in all directions. He also assumed (with no satisfactory explanation) that the secondary wavelets destroy each other except where they touch their common envelope, and that they do not propagate backwards.

EXPERIMENT
J3 Huygens's construction

Apparatus as for experiment J2a plus:

ITEM NO.	ITEM
	A4 graph paper (*e.g.* 2 cm squares in 2 mm graduations)
556	pair of compasses
1153	plastic ruler

Students mark the barriers on a sheet of graph paper as in figure J8, and draw secondary wavelets, radius 3 cm, across the whole width of the aperture, extending them back to the barriers at the ends. It does not matter how far apart the centres of the wavelets lie – more can be added afterwards if the envelope is not clear. Further secondary wavelets are drawn from the envelope(s) to produce three progressive positions of the wavefront which can then be drawn in ink.

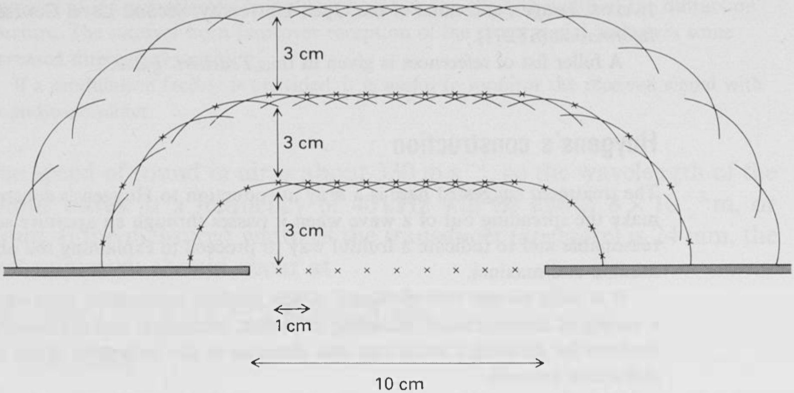


Figure J8
Huygens's construction.

Put the sheet of graph paper on the screen of the ripple tank and adjust the height of the lamp until the barriers' shadows lie over the barriers drawn on the paper. Thorough 'wetting' of the barriers will make their shadows sharper, as will ensuring that the lamp's filament is parallel to them. When a pulse is sent through the aperture, the wavefront will follow the outline of the envelopes drawn on the graph paper. The wavelength can be adjusted so that waves lie over each of the three wavefronts when the screen is viewed through a stroboscope.

There is, of course, one important phenomenon not predicted by Huygens's construction: the amplitude of the wave falls off as the direction of travel moves away from the straight-through direction. It requires the more detailed analysis of wave propagation by Fresnel, and its rigorous mathematical formulation by Kirchhoff, to account for this 'obliquity factor' and to justify discounting the backwards propagation of the secondary wavelets. If these points do arise, one can say that there is a more sophisticated theory which tackles these points, but it is beyond the scope of an A-level course.

Questions

Question 3 applies Huygens's construction to reflection and refraction, and question 4 applies it to diffraction at a straight edge. Question 3 also introduces the refractive index.

Reading for students

MASON, *The light fantastic.*

Reading for teachers

LONGHURST, *Geometrical and physical optics.*
DAMPIER, *A history of science.*
FEYNMAN *et al.,* The Feynman lectures on physics *Volume 1.* Chapter 26.

Wave amplitude and energy

The energy of an harmonic oscillator was seen to be proportional to $(\text{amplitude})^2$ in Unit D, 'Oscillations and waves', and a similar relationship gives the rate at which electrical energy is delivered to a resistor by an alternating current (Unit H, 'Magnetic fields and a.c.').

The idea that intensity is proportional to $(\text{amplitude})^2$ will be needed again in Unit L, 'Waves, particles, and atoms'. The next demonstration shows this for sound waves. It can also be used to revise ideas of superposition and path difference which will be needed for the theory of single-slit diffraction.

DEMONSTRATION
J4 Wave amplitude and energy when waves superpose

ITEM NO.	ITEM
1109	signal generator
1035	pre-amplifier
1511	oscilloscope
183	2 loudspeakers
157	microphone
541/1	rheostat, 10 to 15 Ω
501	metre rule
503/6	3 retort stand bases, rods, bosses, and clamps
1000	leads

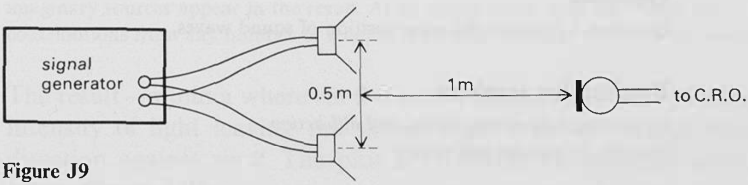


Figure J9
Amplitude and energy of a signal.

The signal generator drives the two loudspeakers, connected in parallel, which are placed about 0.5 m apart. The microphone is about 1 m from them on the edge of another bench (figure J9). Connect the microphone, via the pre-amplifier if necessary, to the oscilloscope switched to its most sensitive range, and put it opposite the point midway between the two loudspeakers, that is, at the central maximum. Set the signal generator to about 4 kHz and increase the gain of the oscilloscope (and pre-amplifier) until the trace nearly fills the screen. If the signal is a minimum in this position, the loudspeakers are in antiphase and the connections to one should be reversed.

Cover each loudspeaker in turn with a coat or box to show that each contributes half the total amplitude. The rheostat can be used to ensure each emits a wave of equal amplitude.

Discussion

The oscilloscope trace indicates potential difference. When both loudspeakers are emitting the p.d. is doubled. Suppose the microphone were connected to a resistor instead of to the oscilloscope. When both loudspeakers are emitting the power dissipated in the resistor would be *four* times as great as when only one is emitting. This is because at every moment during each cycle the potential difference across the resistor, and therefore the current through it, would be doubled. Since power dissipated $= I^2R$, the energy of the sound wave in front of the microphone must be four times as great. The *total* power emitted by two loudspeakers can only be twice the power emitted by one, so this extra power must have been obtained from the places of minimum signal. Energy is not created or destroyed when two waves superpose, it is just distributed differently, being proportional to the square of the amplitude of the resulting wave at each point.

The microphone detects amplitude, as do aerials and radio receivers, while the eye and photographic processes (amongst others) respond to intensity (rate of arrival of energy per unit area). As we have seen, intensity varies as $(\text{amplitude})^2$. This apparatus can also be used to show that 'less can mean more' (*cf.* experiment J2b). Move the microphone to one of the first minima (unlikely to be zero because the microphone is nearer to one loudspeaker than the other). If one loudspeaker is now covered the signal received by the microphone will rise.

Question

Question 5 is about the superposition of sound waves.

Reading for teachers

BRADDICK, *Vibrations, waves, and diffraction.*
FRENCH, *Vibrations and waves.*

In Unit L, 'Waves, particles, and atoms', the significance of $(\text{amplitude})^2$ will be extended to the photon model of light and to the wave model of the electron. There we will say that for light the number of photons arriving per second, and therefore the chance of a photon arriving in a short time interval, is proportional to $(\text{amplitude})^2$. And the chance of finding an electron near a particular place (for example in a hydrogen atom) depends on $(\text{amplitude})^2$ of the electron standing wave.

Explaining the single-slit diffraction pattern

Fresnel and Fraunhofer diffraction
In their reading students may come across the distinction between Fresnel or *near field* diffraction, and Fraunhofer or *far field* diffraction. Teachers may want to give a word of explanation.

Most of the situations investigated in experiments J1, J2, and J3 are in fact examples of Fresnel diffraction, while the derivation given below for a single slit is for a Fraunhofer pattern, which is much more easily analysed.

The simple analysis (Fraunhofer diffraction) applies if either *i* the screen is very far from the slit, or *ii* a lens or eyepiece is used to focus parallel rays. Students may question how parallel rays all leaving the slit in the same direction can superpose. If the screen is very distant the calculated path difference is, to a very good approximation, the same as if the rays were drawn to meet on the screen. If a lens is used the optical path length for all parallel rays entering the lens is the same, so only the initial path difference is relevant.

Question 6 is a structured question, which leads students through the analysis of single-slit diffraction by considering the wavefront to be a row of secondary sources and summing their contributions at an angle θ to the straight-through position. The question could be set as homework, or used in class to derive the expression $n\lambda = b \sin\theta$ for the positions of minima in the pattern. A demonstration of the effect being discussed may be helpful. Demonstration J5a has the advantage of being visible in a shaded corner of the room in daylight and so can be referred to conveniently throughout the discussion; J5b may also serve.

Students may object that the number of secondary sources chosen in the argument of question 6, and the way they are paired, appears somewhat arbitrary. Certainly the analysis is made as convenient as possible, but if at an angle θ the *total* amplitude is zero, then neither changing the number of sources, nor the way the contributions are added will alter the result, and of course neither the number nor the width of the imaginary sources appear in the result. At all angles other than $\sin^{-1} n\lambda/b$ the contributions from any number of secondary sources will sum to give a net amplitude.

The result – minima where $\sin\theta = n\lambda/b$ – allows us to start plotting the intensity of light leaving the slit at angle θ to the straight-through direction against $\sin\theta$. The plot is of course symmetrical about the centre (figure J10).

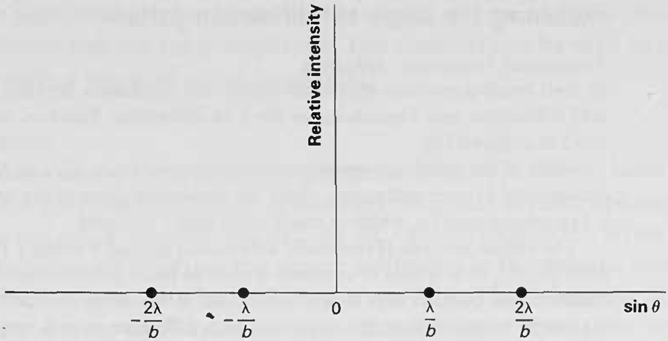


Figure J10
Minima of intensity.

At $\theta = 0$, there is no path difference between secondary wavelets and so the greatest intensity of light is in this direction. At all angles greater than $\sin^{-1} \lambda/b$ it is possible to 'pair' some secondary wavelets with a path difference of $\lambda/2$, so the maxima of intensity are not so great. This suggests that the graph of intensity would be like figure J11: a bright, central band, twice the width of less intense bands either side.

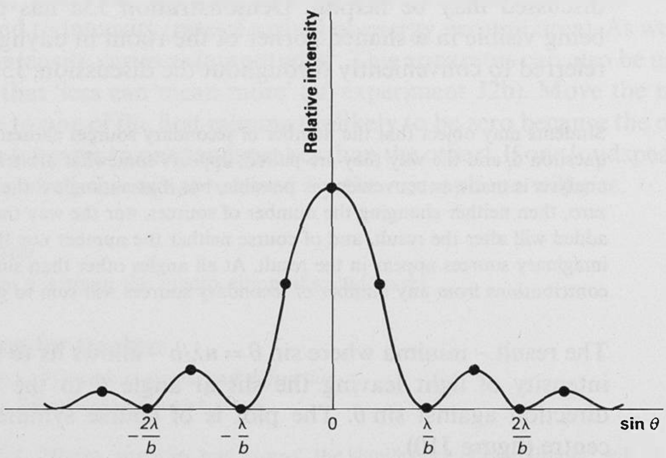


Figure J11
Suggested intensity pattern for a single slit.

Testing the relationship $n\lambda = b \sin \theta$

Students' measurements from the circus of single-slit diffraction experiments J2 could now be compared with the predictions of $n\lambda = b \sin \theta$, but agreement will be poor because of the rough measurements. A demonstration with more careful measurement gives better support for the relationship. Two alternatives are suggested. J5a, which uses a laser, may produce a more striking result and lead to more precise measurements. But if it is used in preference to J5b, in which the light source is a slide projector, teachers should stress that the effect does not depend on the laser, and remind students that they have already seen diffraction patterns using more conventional light sources.

Phasor treatment of single-slit diffraction

A phasor treatment has advantages for students who would enjoy a more rigorous and complete discussion, and who would profit from it. It predicts the reduction of intensity maxima with distance from the centre, and their displacement inwards from the position midway between minima, but it is *not* a necessary part of the course. References to relevant reading are given on page 210.

Computer programs

Several programs are available which plot single-slit diffraction patterns as b and λ are changed, and one shows the construction of phasor diagrams. Details are given on page 210.

ALTERNATIVE DEMONSTRATION

J5 Measuring a single-slit diffraction pattern

J5a Using a laser

ITEM NO.	ITEM
1505	laser
1167/2M	adjustable slit
1167/1B	two holders for lenses
1167/1I	holder for adjustable slit
501	metre rule (or surveyor's tape)
1153	transparent ruler with mm graduations (or 0.5 mm graticule)
24	hand lens (or diverging lens, −20D)
1521/A	converging lens, +4D
1167/1H	small translucent screen
1501	steel or nichrome wire, 0.2 mm diameter, bare
	slide projector

Safety note: Before using a laser teachers should ensure that they are familiar with the recommendations issued from time to time by the D.E.S. (for instance, Administrative memorandum 7/70 'The use of lasers in schools and other educational establishments' – or a more recent one).

If the laser beam is broad enough to cover the adjustable slit at its widest opening, then it can be used undiverged. Otherwise it must be diverged: either a powerful diverging lens or a powerful converging lens – figure J12b – can be used. A converging lens (+4D) forms a bright spot of light on the screen several metres away.

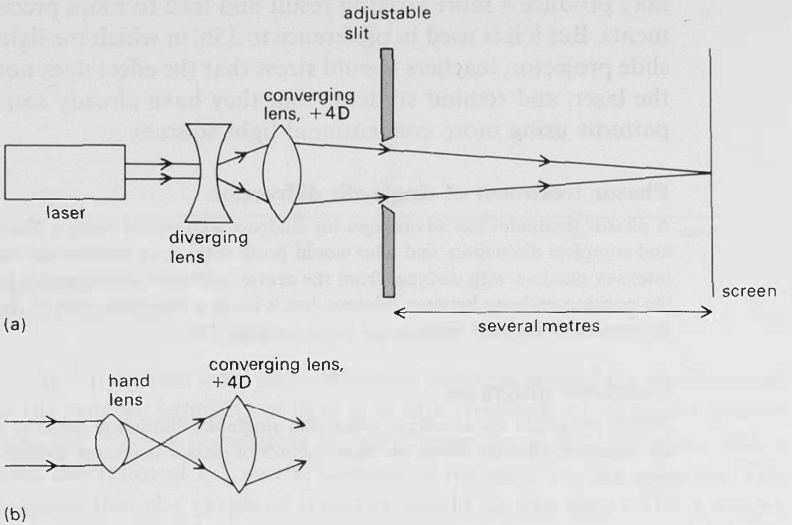


Figure J12
(a) Measurement of single-slit diffraction pattern using a laser.
(b) Use of a powerful converging lens to diverge the laser beam.

Open the adjustable slit to several millimetres and bring the light to a focus on the screen. Then close the slit until a diffraction pattern with a central maximum about 5 cm wide is formed. Measure the separation of the first minima and the distance from the adjustable slit to the screen to obtain a value of θ.

The slit width, b, may be measured by holding the slit in the slide carriage of the projector and marking the position of the edges of its image on a distant screen. The slit is replaced with the transparent ruler, and with its graduations held so that they are in focus on the screen, the width of the slit may be read off. Alternatively, steel or nichrome wire of known diameter may be used to fix the slit width, as in experiment J1.

For the small values of θ involved [for instance $(2.5 \times 10^{-2})/5 = 5 \times 10^{-3}$ radian], $\sin \theta \approx \theta$. Compare the values of $n\lambda$ and $b\theta$ for the first minimum ($n = 1$) using the value of λ given by the laser manufacturer.

J5b Using a slide projector

ITEM NO.	ITEM
	either
1167/1J	holder and two halves of a razor blade, to be used as a single slit
	or
1167/2O	single slit from set of slides
1167/2M	adjustable slit
1167/1H	small translucent screen
1153	transparent ruler with mm graduations
501	metre rule
	powerful slide projector

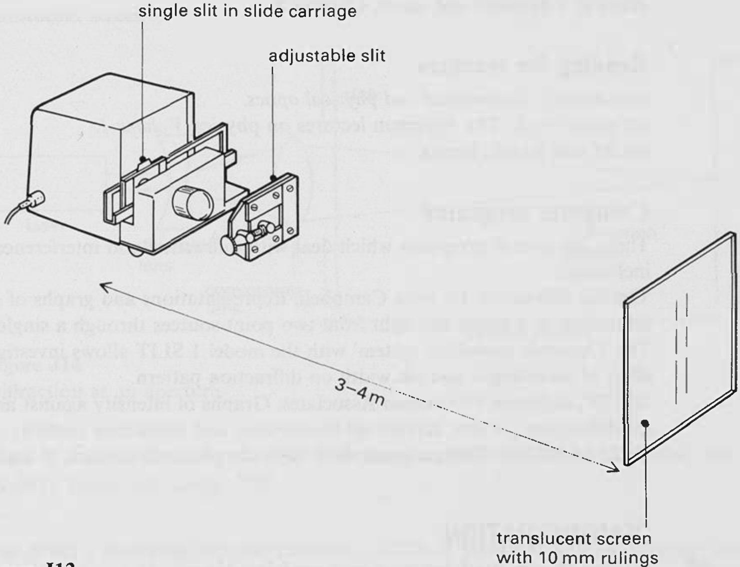


Figure J13
Measurement of single-slit diffraction pattern using a projector.

Put the razor blade slit or the single-slit slide in the projector and focus its image on the translucent screen. The screen should have two lines 10 mm apart ruled on it, and be positioned three to four metres from the projector. Ensure the slit's image is narrow compared with the 10 mm rulings, and that it is central and parallel to them.

Add the adjustable slit in front of the projector. With an assistant viewing the screen, narrow the slit until the central maximum just fills the 10 mm space. Record the distance from the adjustable slit to the screen. The width of the adjustable slit may be measured using the slide projector and transparent ruler as described in demonstration J5a.

$n\lambda = b \sin\theta \approx b\theta$, since θ is small. If the first minima are 10 mm apart, $\theta \approx (5.0 \times 10^{-3})/(\text{slit to screen distance})$.

Using the measured value for b and an 'average' wavelength for white light (5×10^{-7} m), the validity of the formula may be tested. (λ could perhaps be better defined if a green filter were used, though this would of course reduce the intensity.)

Questions

Question 6 is a structured derivation of $n\lambda = b \sin\theta$ for the minima of a single-slit diffraction pattern. Questions 7 and 8 apply the formula to microwaves and sound waves respectively. Question 10 is a coded answer revision question.

Reading for students

FRENCH, *Vibrations and waves*, Chapter 8.

Reading for teachers

LONGHURST, *Geometrical and physical optics.*
FEYNMAN *et al.*, The Feynman lectures on physics *Volume 1.*
HECHT and ZAJAC, *Optics.*

Computer programs

There are several programs which deal with diffraction and interference patterns, including:
'Optical diffraction' by John Campbell. Representations and graphs of intensity for diffraction at a single slit, light from two point sources through a single slit.
The 'Dynamic modelling system' with the model 1 SLIT allows investigation of the effect of wavelength and slit width on diffraction pattern.
'SLITS', Software Production Associates. Graphs of intensity against angle in a single-slit diffraction pattern, double-slit interference and diffraction patterns.
'RCL' by MUSE. This program deals with the phasor treatment of superposition.

DEMONSTRATION

J6 Diffraction and image recombination

Although diffraction at a single slit and at a circular hole has many important applications, apertures in general give rise to more complicated patterns. We show students some of these, briefly, to make the point that there are two possible ways of receiving and storing information about an 'object'.

The time spent on this work should be no more than a single period, with the emphasis on looking at something interesting and which is fun.

Reading for students and teachers

Many of the ideas in this and later parts of the Unit were stimulated by C. A. Taylor's seminal book *Images.* The book explores at greater length many of the themes in the article 'Imaging' in the Reader *Particles, imaging, and nuclei.*

DEMONSTRATION
J6a Diffraction at an aperture

ITEM NO.	ITEM
1505	laser
24	hand lens (or diverging lens, −20D)
1521/A	converging lens, +4D
1167/1B, 1I, 1K	holders for lenses, slide, and adjustable slit
1521/B	set of slides with letter apertures
1167/1E	stops
1167/2M	adjustable slit
1153	aluminium foil
1153	slide mount
1167/1H	translucent screen

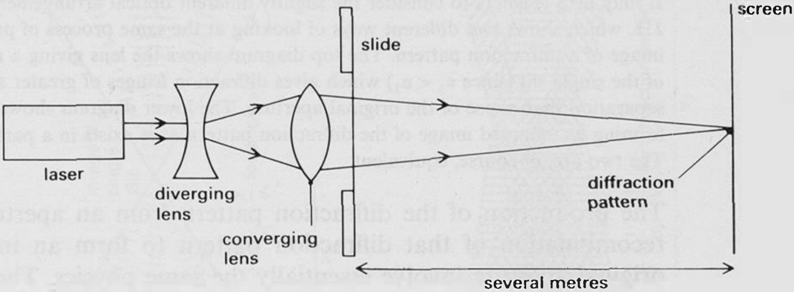


Figure J14
Diffraction at an aperture.

Use of lasers in schools and other educational establishments: see the 'Safety note' on page 207.

Use either a diverging lens (for example, −20D) as shown in figure J14, or the hand lens – see figure J12(b) – to produce a beam about 5 mm wide at a distance of 50 cm from the laser. The +4D lens at this point focuses a sharp point of light on a distant screen. The powers of the lenses are not critical, but the beam must be wide enough to cover the apertures in the slides at the converging lens, which should be weak enough to produce an image large enough to be seen readily by the whole class.

If a prepared set of photographic slides is not available, slides of a circular hole, a single slit, and the letters T, O, X, L, I, Q, A, and Y may be prepared by coating microscope slides with colloidal graphite and scratching the apertures using a stylus and a 2.5 or 3.0 mm stencil. Alternatively, the adjustable slit, a slide of aluminium foil with a circular hole (see experiment J1), and the stencil itself may be used as apertures and supported in place of the slide in figure J14.

Start by reminding students of the patterns produced by a single slit and a circular hole.

The 'T' and 'O' apertures give patterns similar to the characteristic single-slit and circular hole patterns, though there is additional detail, in particular a pattern at right angles to the upright and cross-piece of the 'T'. When these patterns have been discussed, insert the slide with an 'X' and ask students to guess the shape of the aperture from the diffraction pattern it gives.

'T', 'L', and 'I' have similar properties, whilst 'Q' begins to combine features of both 'T' and 'O'. 'A' and 'Y' are more testing and teachers will have to judge how far to take this series of quick demonstrations.

Question

Question 9 is on the recognition and interpretation of diffraction patterns. The aim is modest – that students develop a feel' for the relationship between an aperture and its diffraction pattern.

It may help teachers to consider the slightly different optical arrangement of figure J15, which shows two different ways of looking at the same process of producing the image of a diffraction pattern. The top diagram shows the lens giving a reduced image of the single slit (since $v_1 < u_1$) which gives diffraction fringes of greater angular separation than those of the original aperture. The lower diagram shows the lens forming an enlarged image of the diffraction pattern as it exists in a particular plane. The two are, of course, equivalent.

The production of the diffraction pattern from an aperture and the recombination of that diffraction pattern to form an image of the original aperture involve essentially the same physics. The diffraction pattern formed by shining light on a slide containing the diffraction pattern of an aperture is simply an image of that aperture itself. This can be shown using a slide of the diffraction pattern of a pin-hole or of one of the letters of the alphabet in the experimental arrangement shown in figure J14, or indirectly, as in demonstration J6b.

These ideas will prove useful in the discussion of holography in Section J2.

DEMONSTRATION
J6b Recombination of a diffraction image

Apparatus as for demonstration J6a plus:

ITEM NO.	ITEM
1521/C	converging lens, +10D
1167/1B	holder for lens
1167/2N	coarse grating
1521/D	slides of diffraction pattern of a pin-hole and the letter X

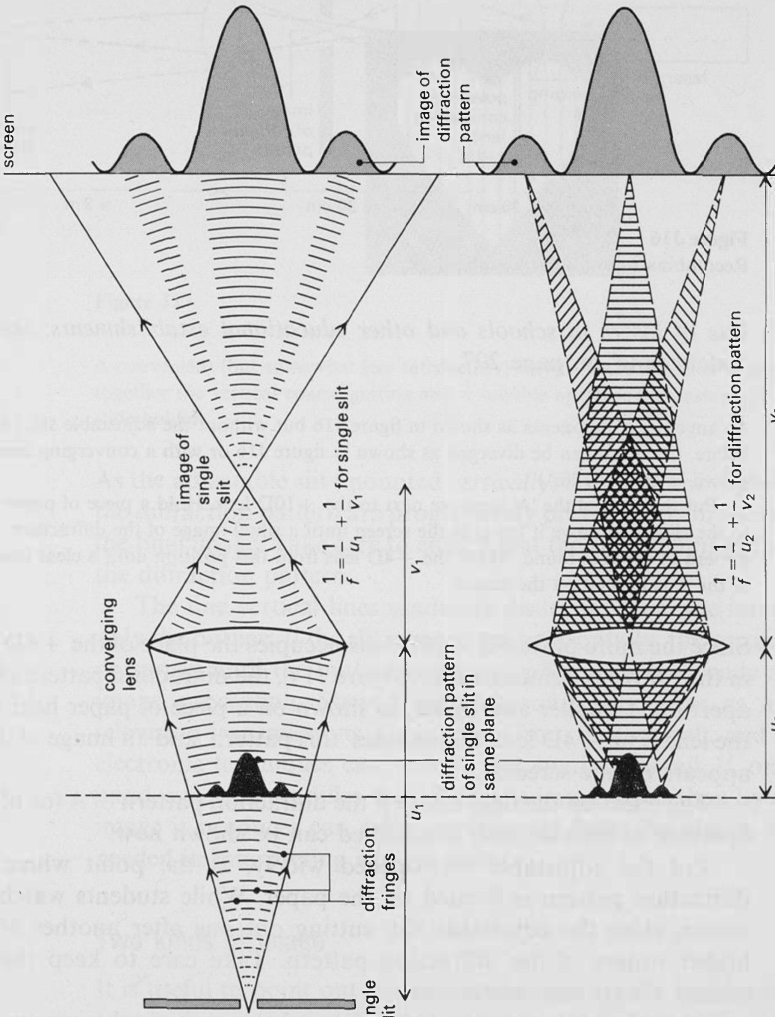


Figure J15
Forming an image of a diffraction pattern.

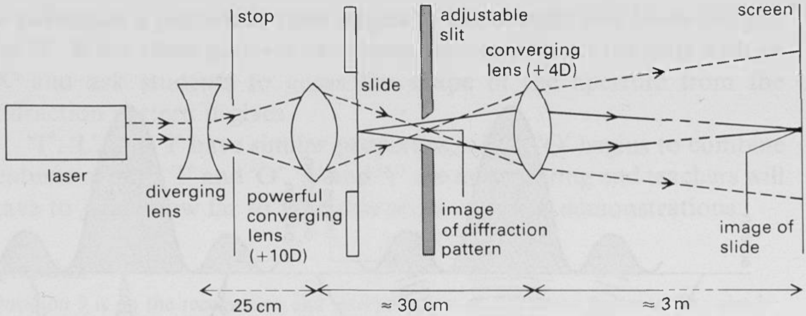


Figure J16
Recombination of a diffraction image.

Use of lasers in schools and other educational establishments: see the 'Safety note' on page 207.

Arrange the components as shown in figure J16 but without the adjustable slit. (As before, the beam can be diverged as shown in figure J16 or with a converging lens as shown in figure J12(b).)

Put the slide of the 'A' aperture next to the +10D lens. Hold a piece of paper next to the slide and move it towards the screen until a small image of the diffraction pattern of the A is found. Move the +4D lens from this position until a clear image of the A is formed on the screen.

Since the more powerful +10D lens occupies the place of the +4D lens in the previous demonstration (figure J14), the diffraction pattern of the aperture is smaller and closer, as shown on a piece of paper held near the lens. The +4D lens 'recombines' this pattern, and an image of the A appears on the screen.

The effect on the final image if the diffraction pattern of A (or of any aperture of suitable size) is modified can be shown now.

Put the adjustable slit, opened widely, at the point where the diffraction pattern is formed on the paper. While students watch the screen, close the adjustable slit, cutting out one after another of the higher orders of the diffraction pattern. Take care to keep the slit central if only one side moves.

As well as the expected dimming, the image on the screen also becomes much less well defined, showing that much of the detailed information about an object is 'stored' in the higher orders of the diffraction pattern. Removing these orders does not destroy one particular region of the image – it affects all parts. We shall return to this point in our brief look at the hologram (demonstration J17).

A possible *benefit* of this type of image degradation can be shown by substituting a slide of an object with a 'grating' of lines superposed on it, such as shown in figure J17.

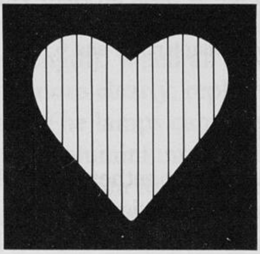


Figure J17

A convenient (but somewhat less satisfactory) way to make the slide is to clamp together the vertical coarse grating and a suitable aperture (for instance a letter) in the slide holder.

As the adjustable slit (mounted *vertically*) is closed, the higher orders of the diffraction pattern are progressively prevented from contributing to the final image. Take care to ensure that the slit remains in the centre of the diffraction pattern.

The fine vertical lines gradually disappear from the image, leaving only the outline. If the slit is mounted horizontally, the vertical lines are preserved until the outline begins to suffer. This is a crude example of image processing. Striped pictures commonly arise as a result of television transmissions, for example from a spacecraft, and analogous electronic techniques can remove this spurious detail. If, on the other hand, a system modifies the diffraction image other than by design, the image it produces can differ significantly from the object, and care is needed in interpreting what is seen.

Two kinds of image

It is useful to point out the differences between the two kinds of image seen in demonstration J6: the 'conventional' and the 'diffraction' image.

A slide projector forms a 'conventional' image where *each point* of the image corresponds to *one point* on the object (the slide). If part of the beam is cut off, part of the image vanishes. This is in contrast to demonstration J6b where *each point* on the diffraction pattern, acting as the object, contributes to *every point* of the final image. In this case the overall shape of the image is preserved, even if degradation reduces the fine detail.

Reading for students

'Imaging' in the Reader *Particles, imaging, and nuclei.*
CANNON and HUNT, 'Image processing by computer'.

Reading for teachers

TAYLOR, *Images.* Chapters 1.1 and 6.7.

RESOLUTION

Diffraction effects, together with other factors, determine how clearly an object is seen when viewed through an aperture. The following two experiments show that diffraction limits the angular separation of two objects which can be seen as distinct and separate.

EXPERIMENT
J7 Resolution

J7a Distinguishing lamps as separate

ITEM NO.	ITEM
1167/1J	holder for two halves of a razor blade, to be used as a single slit
1167/R	set of three colour filters (red, green, blue)
1153	aluminium foil
1153	slide mounts
1501	copper wire, 0.2 mm diameter, bare
1501	steel or nichrome wire, 0.2 mm diameter, bare
27	transformer
	either
1063	multiple light source with two lamps
	or
	2 mounted festoon lamps

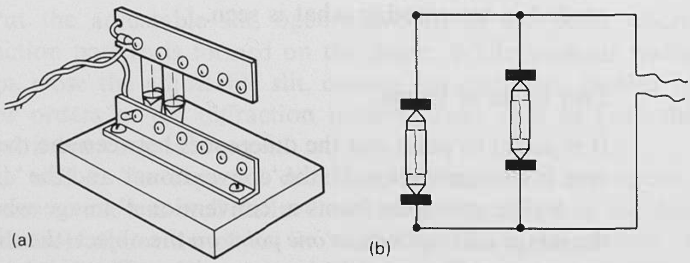


Figure J18

Set up the multiple light source with two festoon lamps and a green filter so that all the students can be more or less in front of it. The 'off-set' arrangement shown in figure J18(b) is better, since the two diffraction patterns and their combination can be seen.

Students adjust the slit width using steel or nichrome wire (as in experiment J1) and view the lamps through the narrow slit. They then increase their viewing distance until the central maxima of the two patterns can just (but only just) be seen to be separate. When all have found this position, replace the green filter (medium wavelength) with the blue filter (short wavelength). When all have seen that the maxima can now be distinguished more clearly, replace the blue filter with the red one (long wavelength). The lamps cannot now be resolved.

If a circular hole made in aluminium foil (see experiment J1) is used with the green filter, the hole is seen to be less successful at resolving the two sources, because the central maxima with the hole are about 20 % wider than with the slit.

Note on resolving power

Using several festoon lamps gives a sharper change from resolution to non-resolution than just two, but Rayleigh's criterion is no longer strictly applicable.

J7b Resolving detail with the eye

ITEM NO.	ITEM
94A	lamp, holder, and stand
1153	card
27	transformer
116	plane mirror
1153	transparent ruler with millimetre graduations

Draw two parallel black lines, 2 mm apart, on white card, and illuminate it brightly. Students measure the greatest distance at which these two lines can be seen as separate with the naked eye. A typical distance is 5 m. The diameter of the eye pupil can be measured by holding a ruler to the eye and looking at it in a mirror.

It is the angle the lines subtend at the eye which matters: the smallest angle for details to be resolved thus seems to be about $2 \times 10^{-3}/5$ or 4×10^{-4} radian. Later experiments and arguments will explain the size of this angle. It may be worth noting at this stage that the ratio of the wavelength of light to the pupil diameter is also a few times 10^{-4}.

Rayleigh's criterion for the resolution of two objects

In experiment J7 students saw that resolution depended on the size of the slit they looked through and on the wavelength. The theory of diffraction at a slit can be used to set a quantitative limit on the resolution of an optical system.

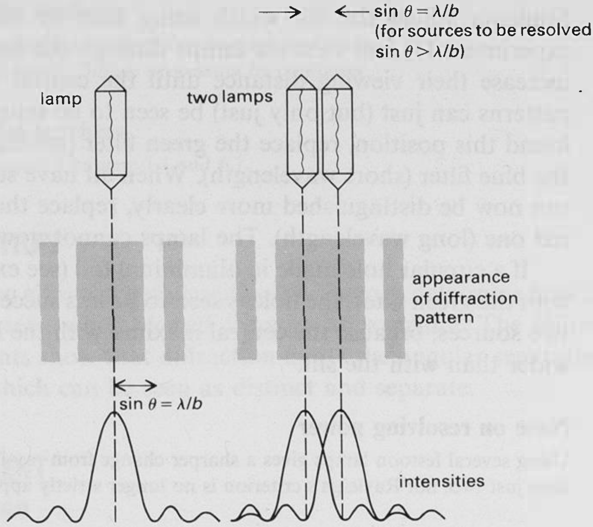


Figure J19
Overlapping diffraction patterns of two sources.

Rayleigh suggested as a guide that if the position of the central maximum due to one object coincides with the first zero of intensity in the diffraction pattern of the other, then in practice the objects may just be distinguished as separate. Where a slit of width b is responsible for the diffraction patterns, the angle between the central maximum and the first zero is λ/b ($\theta \approx \sin\theta$), so λ/b radian is a good, if rough guide to the limit of angular resolution of a system with an aperture of width b.

Resolution of the eye

Taking 2 mm for the diameter of the eye's pupil and 5×10^{-7} m for an 'average' wavelength of light, $\lambda/b = \theta \approx 3 \times 10^{-4}$ radian (taking $\theta = \sin\theta$). The agreement of this angle with that measured in experiment J7b is quite good for this kind of rough measurement, especially as the pupil is circular, not a slit, and θ for the first minimum in that case is $1.22\,\lambda/b$. The eye is a complex system, with defects such as lack of the proper curvature of the retina, which has a finite number of light receptors. It is important also to remember that the eye is an extension of the brain, and not a camera: visual perception is limited by neurological and psychological as well as physiological and physical factors.

Radio astronomy

Resolution plays an important role in radio astronomy. The Jodrell Bank radio telescope has an 80 m diameter paraboloidal mirror with a receiving aerial at its focus. The precision with which it can separate two neighbouring sources, or locate a radio source, is limited by diffraction. At a wavelength of 0.2 m, $1.22\,\lambda/b = (1.22 \times 0.2)/80 = 0.003$ radian (0.17°). The human eye can resolve several times better: about 0.000 4 radian as was seen in experiment J7b.

Interferometer type of radio telescope

It would be impracticable for mechanical reasons to make an aerial several times the width of the one at Jodrell Bank so as to approach the resolution of the naked eye. But a radio telescope can be made with smaller aerials far apart. High resolution is achieved by superposing the signals from each aerial. Two examples of such systems are mentioned in the students' laboratory notes on the following demonstration.

A model of a radio interferometer, using microwaves ($\lambda \approx 3$ cm), gives a clearer idea of the principle.

DEMONSTRATION
J8 Model of a radio interferometer

ITEM NO.	ITEM
184/1	microwave transmitter
184/2	microwave receiver
	2 wax lenses (if available)
181	general purpose amplifier
183	loudspeaker (if not part of item 181)
1153	2 metal plates, about 0.3 m square (*e.g.* capacitor plates; item 1025)
1153	metal plate (0.2 m square, bent to form two 0.2 m × 0.1 m surfaces at right angles)
	table on wheels, or trolley, or sheet of hardboard (see below)

Put the transmitter at one end of the room. At the other end set up the items on the lefthand side of figure J20 on a table, trolley, or board, so that they may be turned through small angles as a unit.

The demonstration works better if lenses are used, but they are not vital.

Position the lenses a metre apart so that the perpendicular bisector of the line joining them points towards the transmitter. Connect the receiver to the amplifier and put it on a support block along this bisector with the right-angled reflector standing symmetrically and pointing towards the horn. Angle reflector A to give maximum receiver response and mark its position. Remove A and adjust B in the same way. If the receiver response is not the same in the two cases, readjust the right-angled reflector until approximate equality is achieved.

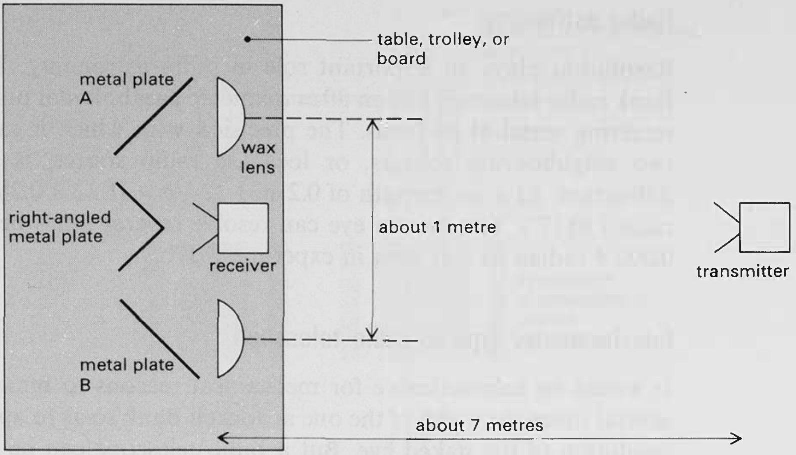


Figure J20
Microwave analogue of radio interferometer.

Turn the receiver around so it faces the transmitter and receives only direct radiation. Carefully rotate the table or board. The receiver response is seen to go through a main maximum with maxima and minima at an appreciable fraction of a radian on either side of it.

Replace the receiver in its original position, as in figure J20, and again rotate the table carefully. This time there are several maxima and minima within the main maximum of the single-beam arrangement and the direction of the transmitter may be determined much more accurately.

Questions

Question 11 is about the effect of a telescope's objective diameter on the image of a distant star.

Question 12 compares the resolution of optical and radio telescopes, while questions 13, 14, 15, 17, and 18 examine various aspects of radio astronomy and interferometry. The resolution of the eye is explored in question 16, and the important ideas of this Section as a whole are revised in question 19.

Reading

The passage on page 171 of the *Students' guide* is about a topical application of diffraction in medical physics.
HENBEST, 'Jodrell under Merlin's spell'.
KELLERMANN, 'Intercontinental radio astronomy'.
READHEAD, 'Radio astronomy by very long base line interferometry'.

SECTION J2
WAVES THROUGH GRATINGS

INTRODUCTION TO DIFFRACTION BY A GRATING

A circus of experiments introduces the grating as a useful tool. No measurements are expected, but the similarities and differences between multiple- and single-slit patterns should be identified, and the dependence of the pattern on grating spacing, wavelength, and type of source should be noted. The two demonstrations, J10 and J11, can be used to elicit and summarize observations and to ensure that students have experience with gratings before treating them theoretically.

The time spent on this work will depend on the students' previous experience and the predilection of the teacher. Some may wish to extend the work on spectra, but a double period probably represents the upper limit.

CIRCUS OF EXPERIMENTS

J9 Looking through gratings

ITEM NO.	ITEM
191/1	'coarse' grating, 100 lines mm^{-1}
191/2	'fine' grating, 300 lines mm^{-1}
1153	mains lamp with 30 cm straight filament (clear tube) and holder
1167/3R	set of colour filters (red, green, blue)
1167/3S	fine black chiffon (or umbrella material)
193/1 and 2	neon (and/or hydrogen) spectrum tube (fluorescent tube will do)
194	holder for spectrum tubes
14	e.h.t. supply
1071	mercury discharge lamp (optional)

Set up the spectrum tube(s) and the straight filament lamp (with colour filters attached as in experiment J1) at opposite ends of a darkened laboratory. Students observe them from about 3 m with one of the gratings held close to one eye.

Safety note: The mercury discharge lamp (item 1071) emits ultra-violet as well as visible radiation. A screen with a fairly small hole or slit should be placed in front of it, and students should be warned not to look directly at the lamp.

The uncovered filament will give a bright, continuous pattern which is more extended with the fine than with the coarse grating. The colour filters give comparatively localized, bright bands, with the red being most widely spaced. The fact that the light transmitted by the filters is far from monochromatic can be used to introduce the next experiment.

The apparently single colour of the spectrum tubes and mercury vapour lamp gives rise to characteristic line spectra. If neither of the spectrum tubes nor a mercury discharge lamp is available, students can look at a fluorescent strip light.

Spectra give evidence for the existence and spacing of atomic energy levels. It might be well to hint, in the middle of a Unit devoted to light as a wave motion, that atoms seem to emit and absorb radiation in parcels, or photons. This will be taken up again in Unit L, 'Waves, particles, and atoms'.

If the single filament is observed through the chiffon, not only are fringes seen at right angles to both weave directions as expected, but 'diagonal' fringes are also present. If the filament is observed through the material of an umbrella the same pattern may be seen, though some material has three weave directions at 120° to each other, and a more complicated pattern results.

Home experiments

In Home experiment JH2, Simple spectroscopy, a simple spectroscope is made from a cardboard tube, one end of which is covered with aluminium foil containing a narrow slit, and the other with a fine grating or piece of replica grating taped over a larger aperture. The lines of the grating and the collimating slit must be parallel. The spectroscopes are cheap enough to allow students to take them home and examine street lights and shop signs, or even to observe the Sun's spectrum *by reflection from a piece of white card.* Students should be warned against direct observation.

Home experiment JH3, Ear and eye, draws students' attention to the fact that our ears are able to detect that a sound (for example two notes played simultaneously) contains a mixture of frequencies, but our eyes cannot do the same for light containing several frequencies.

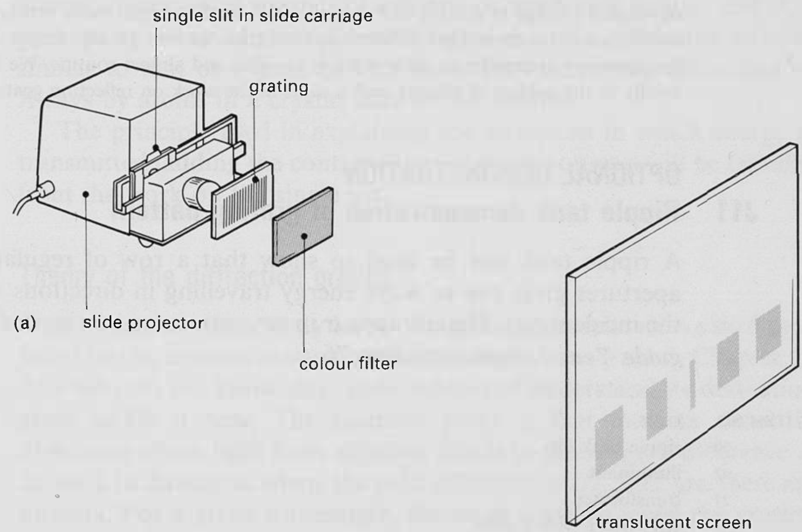
DEMONSTRATION
J10 The diffraction grating

ITEM NO.	ITEM
191/1	coarse grating
	either
1167/1J	holder for two halves of a razor blade
	or
1167/2O	single-slit slide
1167/2N	set of coarse gratings
1167/3R	set of colour filters (red, green, blue)
23	microscope
46/1	translucent screen
	powerful slide projector

Put a single slit in the slide carriage of the projector, and focus the image of the slit on a translucent screen about a metre from the lens.

Hold the coarse grating with about 100 lines per millimetre (item 191/1) over the projection lens, as in figure J21(a). Three or four orders of the pattern should fill the width of the screen. The fringes are best viewed from behind the screen. Add colour filters in front of the grating.

Replace the coarse grating with a very coarse one (from item 1167/2N) with about 10 lines per millimetre. The fringes on the screen are now a few millimetres apart instead of a few centimetres as before – figure J21(b).



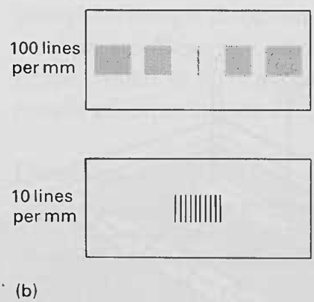


Figure J21
Diffraction pattern of a grating.
(a) Method of projection.
(b) Patterns produced on screen by coarse and by very coarse gratings.

In discussion, the following points should emerge: the fringe spacing depends directly on wavelength, λ, and inversely on grating spacing, s; the fringes in the patterns produced with a grating are *sharper*, as well as brighter, than the fringes in the single-slit pattern; all the maxima are equally wide (the central maximum is *not* twice the width of the others).

The lines ruled on a fine grating can easily be seen under a microscope. Students should see that they are not the opaque stripes of the text book, but imperfect, even kinked, scratched lines.

One obvious disadvantage of the gratings seen so far is that precious light is sent into many orders, when only one spectrum is needed for analysis. Spectroscopists have developed a range of grating-like devices which send the light more nearly where it is needed, and also spread the different wavelengths out over greater angles, so that in spectroscopy extremely accurate work is possible and almost routine. We shall return briefly to the subject of blazing and etalons in the work on reflection gratings.

OPTIONAL DEMONSTRATION
J11 Ripple tank demonstration of grating pattern

A ripple tank can be used to show that a row of regularly spaced apertures gives rise to wave energy travelling in directions other than the incident one. Details appear in REVISED NUFFIELD PHYSICS *Teachers' guide Year 5*, demonstration 76.

ITEM NO.	ITEM
90	ripple tank kit
47	illuminant
27	transformer
1033	cell holder with 2 cells
541/1	rheostat, 10 to 15 Ω
105/1	hand stroboscopes
1000	leads

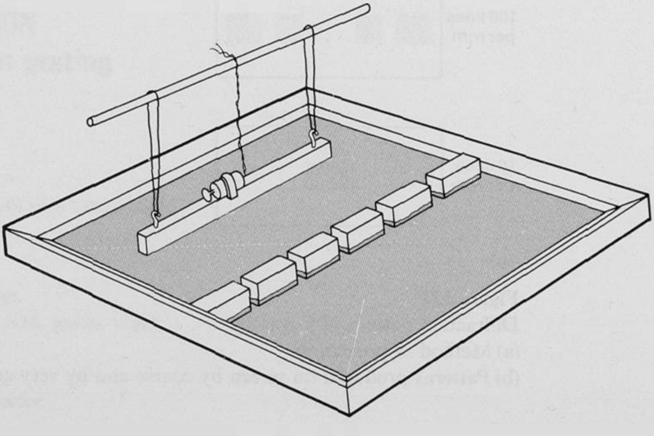


Figure J22

Arrange the barriers so that the gaps between them are about the same size as the wavelength of the water waves when the motor is rotating at the middle of its frequency range. Use large barriers to prevent waves passing round the sides of the row.

Students use hand stroboscopes to 'freeze' the pattern, and by looking along the water surface, see waves in directions on either side of the straight-through position. By altering the wavelength, the change of direction may be demonstrated.

A teacher may wish to point out in passing that as the gaps between the barriers approach the wavelength, the barriers themselves act almost as sources of secondary wavelets or 'scattering points', and that perhaps an array of obstacles would give rise to a diffraction pattern similar to that of a grating. This looks forward to the diffraction of X-rays by atoms in a crystal later in this Section.

The principle used in explaining the directions in which energy is transmitted (adding the contributions of many sources) will be familiar from the work on the single slit.

Theory of the diffraction grating

Some students will already know the grating formula: $n\lambda = s \sin\theta$ (see, for example, REVISED NUFFIELD PHYSICS, *Pupils' text Year 5*, Chapter 9). Any who do not know this relationship and understand its derivation must tackle it now. The essential point is that maxima occur in directions where light from adjacent slits is in phase (path difference λ, 2λ, etc.). In directions where the path difference is $\lambda/2$, $3\lambda/2$, etc. there are minima. For a given wavelength, the more closely spaced the grating elements (smaller s) the larger the angle the maxima make with the straight-through direction.

For small angles, $\sin\theta \approx \theta$, so the fringe separation, which is proportional to θ, is proportional to λ/s, which is consistent with the results of experiments J9, J10, and J11.

Questions

Question 20 shows how Huygens's construction can be used to predict the maxima in the diffraction pattern of a grating. Questions 21 to 23 are largely qualitative; question 24 requires a calculation using the grating formula.

At A-level we can show students how the fringe pattern also depends on the *number* of slits in the grating, and the *width* of each slit. This is investigated in demonstration J12.

DEMONSTRATION/EXPERIMENT

J12 Sharpness of maxima depends on number of slits

Demonstration

ITEM NO.	ITEM
1505	laser
1167/2O	set of parallel slits
1167/2N	set of coarse gratings
1167/1K	support for slits
1167/1G	big stop
501	metre rule
24	hand lens (or diverging lens, −20D)
1521/A	converging lens, +4D
1167/1B	lens holder
	projector screen or light-coloured wall

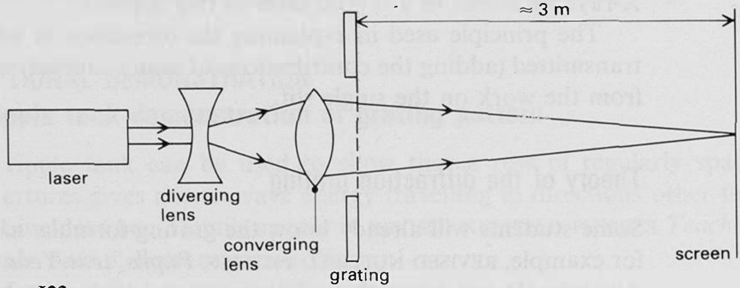


Figure J23

Use of laser: See the 'Safety note' on page 207.

The laser beam is diverged using either the hand lens – figure J12(b), or the diverging lens as shown in figure J23.

The powers of the lenses are not critical, but the beam should diverge sufficiently to cover about 1 cm of the grating. A balance has to be struck: the pattern will be less bright if the beam diverges too much and some light is lost from the system; but the beam must be wide enough to pass through appreciably more than six slits in the coarse grating to demonstrate that more slits give sharper maxima.

A weak, converging lens forms a focused image of the pattern on the distant screen (≈3 m away). Again, a balance between brightness and size of the image has to be achieved. Adjust the system for the 2-slit slide (which gives the dimmest pattern); the other patterns will then be readily visible.

The set of slides needed consists of different numbers of equally spaced, equal-width slits, 1, 2, 3, 4, 5, and 6, and a coarse grating of some fifteen slits.

Hold the slides in turn in front of the converging lens, starting with the 2-slit slide. The maxima will be quite broad, but as the number of slits, N, is increased, so the maxima stay in the same position on the screen, but become narrower (sharper) and brighter. With the coarse grating the fringes are very well defined, but have the same spacing as before.

As the number of slits is increased, fainter (secondary) maxima become visible between the brighter (principal) maxima. The empirical rule that there are $(N-2)$ secondary maxima with a grating of N slits may be established, and students may recall having observed them in demonstration J10. This may provide the stimulus to look into the phasor analysis later; but for the majority of students it will be enough to indicate that there are some features of these patterns which require a more detailed analysis than is appropriate at this stage.

With the coarse grating in position it will be obvious that, although the maxima are equally spaced, there are some gaps (or missing orders) in the otherwise regular pattern. Mark the positions of these on the screen with small pieces of PVC tape, 'Blu-tack', or drawing pins.

Now substitute the single-slit slide from the same set (1167/2O) for the coarse grating. The minima in the single-slit pattern fall exactly where the markers have been placed, confirming that it is the minima of the single-slit envelope which account for the missing orders. If no light is leaving *any* slit at an angle θ, then even if the path difference between slits is a whole number of wavelengths, there can be no maximum on the screen at this angle.

Experiment

ITEM NO.	ITEM
94A	lamp, holder, and stand
27	transformer
1521/E	2 lenses, + 1D or + 0.5D
1167/2L	eyepiece
1167/2N	set of coarse gratings
1167/2O	set of parallel slits
1167/1K	slide holder
1167/3R	set of colour filters (red, green, blue)
1167/1H	small translucent screen

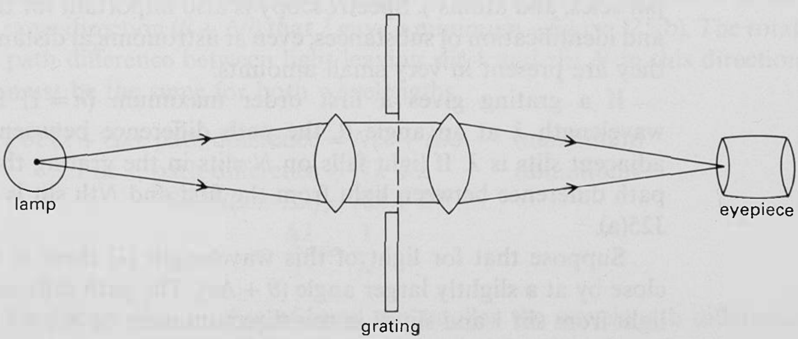


Figure J24

This alternative, which is a student experiment rather than a demonstration, is described in the students' laboratory notes. An axial filament lamp at the principal focus of a $+1$D converging lens forms a parallel beam. The gratings are positioned immediately after the first lens, and can be propped against it, and a second lens of equal power is put close to the grating to focus a real image of the lamp filament on the translucent screen at a distance equal to the focal length of the lens (figure J24).

The eyepiece is positioned so that a magnified image of the pattern on the screen can be seen (this is to aid what can be a difficult alignment), and the screen is then removed. If the image is uncomfortably bright with the larger number of slits, the lamp may be under-run.

Students observe the pattern produced by the coarse grating, then 6, 5, 4, 3, and 2 slits. Return to the grating. Some orders are 'missing' and are found to be coincident with minima produced by a single slit of the same width as the others, confirming the explanation for the grating pattern given above.

Question

Question 25 takes students through a numerical derivation of the width of a maximum and the resolving power of a grating. Teachers may prefer to set this before giving the algebraic treatment below, or perhaps instead of it.

Many slits: sharply defined spectra

Students should appreciate that gratings are important and useful because they can send a lot of light into *bright*, *narrow* fringes, which makes it possible to measure wavelengths precisely and to distinguish between two very similar wavelengths. Precise measurement of wavelength is important for the information it can give about the structure of atoms and molecules (an example will be seen in Unit L, 'Waves, particles, and atoms'). Spectroscopy is also important for the detection and identification of substances, even at astronomical distances or when they are present in very small amounts.

If a grating gives a first order maximum ($n = 1$) for light of wavelength λ at an angle θ, the path difference between light from adjacent slits is λ. If light falls on N slits in the grating then the total path difference between light from the first and Nth slit is $N\lambda$ – figure J25(a).

Suppose that for light of this wavelength (λ) there is a minimum close by at a slightly larger angle ($\theta + \Delta\theta$). The path difference between light from slit 1 and slit N in this direction must be $N\lambda + \lambda$ (so that the path difference between light from slit 1 and slit $N/2$, slit 2 and slit $N/2 + 1$, etc. will be half a wavelength, ensuring no light in this direction).

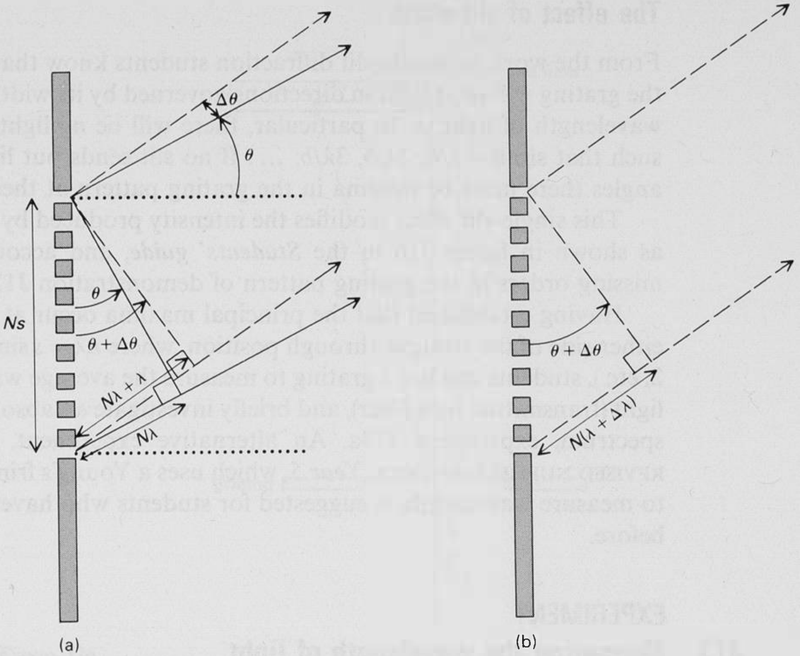


Figure J25
Resolving power of a grating.

From figure J25(a)

$$Ns \sin(\theta + \Delta\theta) = N\lambda + \lambda$$

Adopting Rayleigh's criterion, two spectral lines will just be resolved if the maximum of one coincides with the first minimum of the other. Suppose a slightly larger wavelength $(\lambda + \Delta\lambda)$ gives a maximum in the same direction $(\theta + \Delta\theta)$ that λ gave a minimum – figure J25(b). The total path difference between light leaving slit 1 and slit N in this direction must be the same for both wavelengths.

$$\text{For } (\lambda + \Delta\lambda)\text{, path difference} = N(\lambda + \Delta\lambda) \qquad \text{(maximum)}$$
$$\text{For } \lambda\text{, path difference} = N\lambda + \lambda \qquad \text{(minimum)}$$
$$\text{So, } N(\lambda + \Delta\lambda) = N\lambda + \lambda$$
$$\Rightarrow \frac{\Delta\lambda}{\lambda} = \frac{1}{N}$$

The larger the number of lines, the smaller the wavelength difference that can be resolved. It is easy to show that for the nth order spectrum $\Delta\lambda/\lambda = 1/(nN)$.

The effect of slit width

From the work on single-slit diffraction students know that each slit in the grating will send light in directions governed by its width, b, and the wavelength of light, λ. In particular, there will be *no* light at angles θ such that $\sin\theta = \lambda/b, 2\lambda/b, 3\lambda/b, \ldots$. If no slit sends out light at these angles there must be minima in the grating pattern at these angles.

This single-slit effect modifies the intensity produced by the grating, as shown in figure J16 in the *Students' guide*, and accounts for the missing orders in the grating pattern of demonstration J12.

Having established that the principal maxima occur at values of θ_n either side of the straight-through position, where $n\lambda = s\sin\theta_n$ ($n = 0$, 1, 2, etc.), students can use a grating to measure the average wavelength of light (transmitted by a filter), and briefly investigate an absorption band spectrum, experiment J13a. An alternative experiment, J13b, from REVISED NUFFIELD PHYSICS, *Year 5*, which uses a Young's fringes method to measure wavelength, is suggested for students who have not done it before.

EXPERIMENT

J13 Measuring the wavelength of light

J13a Using a grating

ITEM NO.	ITEM
94A	lamp, holder, and stand
27	transformer
191/2	fine grating, 300 lines mm^{-1}
1167/3R	red filter
501	metre rule and 0.5 metre rule
	optional
	sodium vapour lamp
	parallel-sided glass cell
	potassium manganate(VII) solution

Students set up the lamp above the 0 cm mark of the 0.5 metre rule and hold the grating against the edge of the metre rule positioned at right angles to it, as shown in figure J26.

Plasticine can be used to keep the 0.5 metre rule standing on its edge so that the scale is visible through the grating.

With the red filter in place, students note the position of the first maximum and measure its distance from the central filament, x. If the experiment is done in pairs, a partner may move a pointer along the 0.5 metre rule until stopped at the first maximum by the observer looking through the grating.

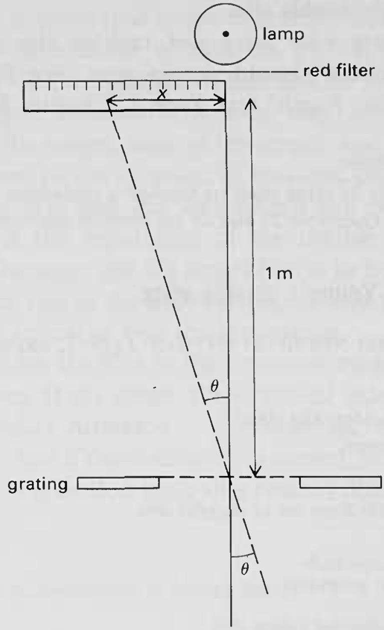


Figure J26

Using the stated grating spacing, s, the average wavelength for the filter may be calculated from

$n\lambda = s \sin\theta \qquad (n = 1)$ and $\sin\theta = x/\sqrt{1 + x^2}$
(x measured in metres)

Two extensions are possible for a fast and keen group:

i Use a sodium lamp, and the known average wavelength of the doublet, to check the value given for the grating spacing.
ii Introduce a parallel-sided glass cell containing potassium manganate(VII) solution in front of the grating, using the lamp filament, *without a filter*, as source. A strong absorption band will be visible if the solution is strong enough. Add potassium manganate(VII) solution to water in the cell until the solution is strong enough to move beyond a 'stepped' absorption band to a uniformly dark band, but not so strong that it is difficult to see the scale of the rule. Using the calibration wavelength from above, the absorption band wavelength limits may be found from simple ratios of the x distances.

Revised Nuffield Advanced Chemistry

Spectroscopy is dealt with in Topic 4.

Young's double slits

Students who have not tackled the theory of Young's double-slit experiment should do so now. See for example REVISED NUFFIELD PHYSICS, *Pupils' text Year 5*, Chapter 8.

Questions

Question 26 takes students through a derivation of the formula $\lambda/x = s/L$ for Young's fringes. Questions 27 and 28 are practice questions.

J13b Using Young's double slits

(REVISED NUFFIELD PHYSICS *Year 5*, experiment 69)

ITEM NO.	ITEM
94A	lamp, holder, and stand
27	transformer
	either
1167/2O	2-slit slide from set of parallel slits
	or
	microscope slide
	colloidal graphite
	needle
	slide holder for ruling slits
24	hand lens
1153	plastic ruler or 0.5 mm graticule
1167/1J	holder and two halves of razor blade
1167/1H	translucent screen
1167/3R	set of colour filters (red, green, blue)
501	0.5 metre rule
	optional
	slide projector

Detailed instructions on how to make a pair of slits on a microscope slide coated with colloidal graphite are given in the students' laboratory notes and in REVISED NUFFIELD PHYSICS *Teachers' guide Year 5*, page 103.

The lamp and screen should be at least 3 m apart with the 2-slit slide 1 m from the lamp. The single slit (a few millimetres wide) in front of the lamp ensures that only the area of the double slits is illuminated on the slide. Fringes are visible on the screen, and the lamp may be over-run to 14 V for short periods to make them brighter. See figure J27.

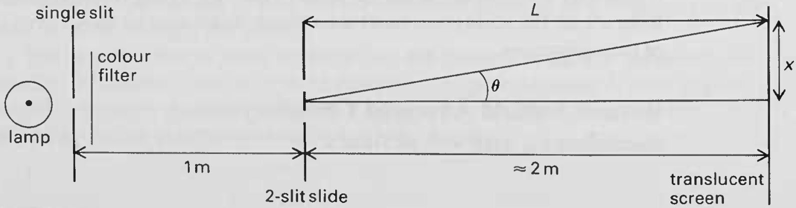


Figure J27

With a colour filter in place (red or green is best), students measure the separation of a given number of maxima on the screen. The total distance is divided by the number of maxima to give x, the average separation of fringes. Alternatively, they may use the eyepiece to view the fringes on the screen, remove the screen, and hold the graticule (plastic ruler) in front of the eyepiece to measure the fringes.

Since $L \gg x$, $x/L \approx \theta$, and the wavelength can be calculated from $n\lambda = s \sin\theta \approx sx/L$ if the separation of the double slit is known. A convenient way to measure the slit separation is to hold the slide and a small piece of plastic rule in the slide carriage of a projector and use the image on a distant screen to find the separation.

If the distance from the slits to the screen or eyepiece is halved, the fringe spacing halves. If the single slit is moved sideways, the pattern moves in the opposite direction – provided the single slit still illuminates both slits. And if the double slit is moved, the pattern moves in the same direction – provided both slits remain illuminated.

Questions

Questions 31 to 34 test understanding of grating theory.

Reading for teachers

LONGHURST, *Geometrical and physical optics.*

Computer programs

There are several programs which deal with diffraction and interference patterns at two slits and with gratings: 'Interference and diffraction of waves', Chelsea Science Simulations; 'Multiple slit interference', G.S.N. Physics Pack; 'Slits' and '2 Slits', Dynamic Modelling System; and 'Young's Slits', Heinemann Computers in Education.

Reflection gratings

Having established the theory and behaviour of the transmission grating, a gesture may be made in the direction of a more modern device by showing the spectrum from a concave reflection grating, and, incidentally, taking a first step in examining the electromagnetic spectrum (demonstration J14). Because light does not pass through a reflection grating it can cope with ultra-violet and infra-red radiation which would be absorbed by most transmission grating materials.

By shaping the grooves of a grating in a special way, ('blazing'), it is possible to direct much of the energy to one particular order and to increase the effective dispersion. Considerable advances have been made in constructing both reflection and transmission etalons, or stacks of offset plates, which behave similarly to gratings at glancing angles.

Reasonably priced reflection gratings are available which consist of replica grating material glued to the face of a concave lens or mirror.

DEMONSTRATION
J14 A spectrum using a concave reflection grating

ITEM NO.	ITEM
1167/1J	holder with two halves of a razor blade
1073	concave reflection grating
94A	lamp, holder, and stand
27	transformer
1153	white screen (non-fluorescent)
1153	fluorescent paper (green)
1046	infra-red and ultra-violet filters
68	phototransistor
1033	cell holder with one cell
1507	ammeter, 1 mA, d.c.
1000	leads

Put the screen adjacent to the lamp (position A in figure J28), with the vertical filament at the same height as the screen centre. The grating should also be at this height and positioned so that it produces an image of the filament on the screen. The grating is then about 0.5 m from the lamp and the screen.

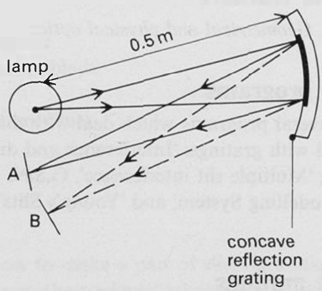


Figure J28
Concave reflection grating.

Moving the screen to position B and beyond should enable the first, second, and possibly the third order spectra to be shown. A slit put in front of the lamp will improve the purity of the spectra.

With the screen at position B (first order spectrum), the presence of ultra-violet and infra-red radiations may be demonstrated.

Ultra-violet: Pin the strip of fluorescent paper to the screen so that the lower half of the spectrum falls on it, the upper half still falling on the non-fluorescent screen. Some fluorescence can be seen beyond the visible spectrum, particularly if the lamp is over-run (14 V), but much of the fluorescence will be in the visible blue–violet region. The u.v. filter can be used to remove much of the visible radiation and so enhance the effect.

Infra-red: Students may have seen this before (REVISED NUFFIELD PHYSICS *Teachers' guide Years 1 and 2*, Demonstration 145).

Connect the phototransistor, cell, and milliammeter in series. Put the phototransistor just in front of the screen in the blue part of the first order spectrum. Rotate the reflection grating slowly to sweep the spectrum *from the blue end to the red* across the phototransistor: a peak response will be found beyond the visible red region. With the transistor positioned here, the effect of the filters may be demonstrated.

The overlap between ultra-violet radiation and visible light, and between infra-red radiation and light, will be useful in the discussion of the family of electromagnetic waves in Section J3.

By using a reflection grating at glancing angles of incidence, it is possible to measure wavelengths some thousand times smaller than the grating spacing. X-ray wavelengths may be measured using an optical reflection grating at glancing angles. This technique can be introduced using an old LP gramophone record or a metal rule as a grating for visible radiation.

DEMONSTRATION
J15 Reflection grating at glancing incidence

ITEM NO.	ITEM
1153	LP gramophone record (old)
1063	multiple light source and festoon lamp (fluorescent tube or horizontal filament lamp and holder would do)
27	transformer
24	hand lens

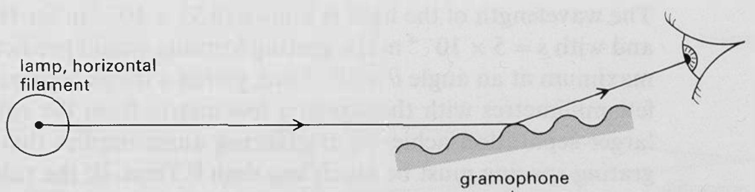


Figure J29

Mount the light source with the filament horizontal at one end of a darkened laboratory, about head height. (It is also possible to use a fluorescent strip light.)

Students stand about 3 m away and hold the record horizontally at arm's length, facing the lamp. By tilting the record, the reflected image of the lamp can be produced at the far edge of the record. By adjusting the height and tilt of the record, several orders of spectrum should be observable.

By examining the playing surface with a hand lens, the groove spacing may be estimated and compared with the wavelength of light. It is about $10^3 \times \lambda$.

ITEM NO.	ITEM
1505	laser
1153	metal rule with 0.5 mm divisions (0.5 mm graticule would do)
1153	white screen
1153	Plasticine
1522	adjustable height stand (Labjack)

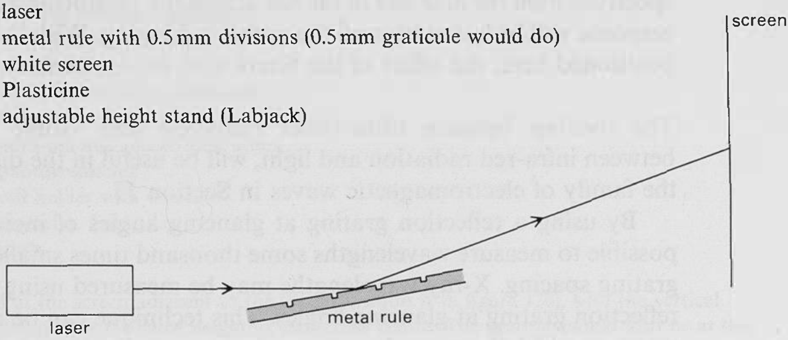


Figure J30

Use of laser: See the 'Safety note' on page 207.

Shine the laser without lenses on to a white screen several metres away.

Lay the metal rule or graticule on the adjustable stand and fix a piece of Plasticine under the end of the rule further from the laser so that it slopes slightly upwards (figure J30). Position the stand so that the rule lies a few millimetres below the laser beam and adjust its height until the beam strikes the rule at about its mid-point. Well-defined orders should be visible on the screen; adjust the angle of the rule for optimum effect.

The wavelength of the light is known (6.33×10^{-7} m for He–Ne lasers), and with $s = 5 \times 10^{-4}$ m the grating formula would predict a first order maximum at an angle $\theta \approx 10^{-3}$ rad, giving a fringe separation of only a few millimetres with the screen a few metres from the rule. The much larger separation achieved at glancing angle implies that the *effective* grating spacing must be much less than 0.5 mm. (If the rule, or a comb, is held in front of the eye and rotated, the rulings or the gaps between the teeth appear to be more crowded and their separation seems smaller.)

An optical reflection grating, for which s might typically be 10^{-6} m, could be used with wavelengths of 10^{-9} m, if orientated at glancing angle. This is approaching the X-ray region of the electromagnetic spectrum and the wavelength of X-rays can be determined employing this technique.

Reading for teachers

LONGHURST, *Geometrical and physical optics.*

Waves through complex gratings

We now suggest a brief look at the diffraction patterns produced by more complex gratings, that is, periodic arrays of apertures. This both extends the ideas of diffraction imaging of experiment J6 in preparation for a little work on holography and should give students an appreciation of how X-ray crystallography can provide information about the structures of a wide variety of materials.

No more than a single lesson should be spent on demonstration J16 and the pace should be brisk to ensure that the thread of the argument is not obscured by too much detail. Students can follow up the ideas in the article 'Imaging' in the Reader *Particles, imaging, and nuclei,* and in C. A. Taylor's book *Images.*

DEMONSTRATION
J16 Diffraction at complex gratings

ITEM NO.	ITEM
1505	laser
24	hand lens (or diverging, lens, – 20D)
1521/A	converging lens, + 4D
1153	white screen
1167/1B	2 holders for lenses
191/1	2 coarse gratings
1167/3S	black chiffon
1521/G	slides with various aperture arrays

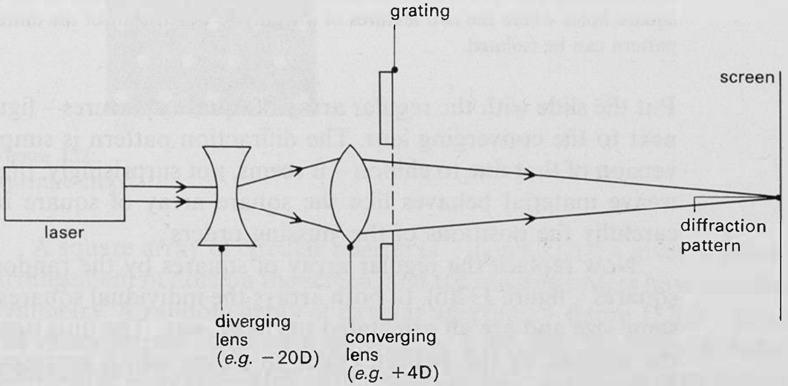


Figure J31

Use of laser: See the 'Safety note' on page 207. In this experiment *special care must be taken that none of the higher order beams are sent anywhere near students.*

The optical arrangement is the same as for demonstration J6a, but it is important to make sure that the beam is at least 5 mm wide at the converging lens.

Put one of the coarse gratings (80 lines mm^{-1}) next to the converging lens. A bright, equally spaced row of maxima (characteristic of the grating pattern) will be seen on the screen, with no missing orders. (Since the slit width is very small the central maximum envelope of the single-slit pattern extends across the whole screen.)

Now add a similar grating at right angles to the first. Instead of two rows of maxima at right angles, which students might expect, the screen is covered with a square array of dots. Students may recall seeing a similar pattern when they looked through chiffon in experiment J9, and if this is now substituted for the two gratings a much smaller square array of dots is formed on the screen (because the weave is much coarser than the lines in the grating). There is an additional feature: there seem to be regions, especially along the diagonals, where the dots are missing. In demonstration J12 the 'missing orders' in the grating pattern were due to the existence of minima in the diffraction pattern of the single slit – which is the repeated element in the array. Here the missing orders are also due to the diffraction pattern of the repeated element – which in this case is a square hole. This can be investigated using a slide with a regular square array of square holes.

Students may be able to suggest some reasons for the square array of dots in the diffraction pattern, such as the presence of additional regular detail of diagonals, but it is probably best to move on to the demonstration of diffraction by the arrays of square holes where the two features of the grating which control the diffraction pattern can be isolated.

Put the slide with the regular array of square apertures – figure J32(a) – next to the converging lens. The diffraction pattern is simply a scaled version of that due to chiffon – it seems, not surprisingly, that a square-weave material behaves like the square array of square holes. Note carefully the positions of the 'missing orders'.

Now replace the regular array of squares by the random array of squares – figure J32(b). In both arrays the individual squares are all the same size and are all orientated the same way. The dots disappear and the minima in the pattern occur in places which correspond to the 'missing orders' in the diffraction pattern of the regular array. It seems then that the spacing and geometry of the array determines the positions of the maxima, and that the size and shape of the repeated aperture determines the relative intensities of the maxima. (A useful test would be to compare the pattern produced by the random array of square apertures with that due to a single square aperture: we would expect them to be the same. However, it is not easy to compare these two patterns directly because the pattern given by a single aperture would be very faint.) The suggestion can, however, be confirmed using other slides.

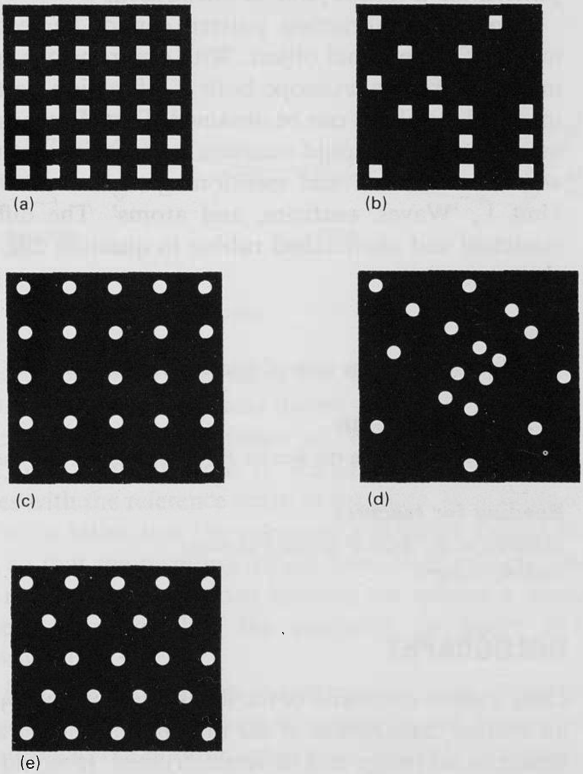


Figure J32
Aperture arrays.

A square array of circular apertures – figure J32(c) – gives a square arrangement of dots on the screen, and the missing orders have circular symmetry. A random array of circular apertures – figure J32(d) – gives the characteristic diffraction pattern for a circular aperture. A hexagonal array of circular apertures – figure J32(e) – gives a pattern of dots with hexagonal symmetry, and again the intensities of these dots correspond to the diffraction pattern of a circular aperture. (A reminder of question **21** – about a grating with lines obscured at random – would be useful here.)

The two general points, that *i* the repeat pattern or lattice of a structure determines the pattern of dots in its diffraction pattern, and *ii* the nature of the scattering centres or the contents of the unit cell determines the relative intensities of these dots, are fundamental to X-ray crystallography. Only one simple technique, the Laue transmission pattern for a thin crystal, is introduced.

An X-ray diffraction pattern cannot be recombined to form an image of the original object. With electrons of appropriate wavelength in an electron microscope both a diffraction pattern and a magnified image of the object can be obtained. That electrons can be focused using magnetic lenses should occasion no surprise, but their wave properties will seem unusual and mentioning them now will pave the way for Unit L, 'Waves, particles, and atoms'. The diffraction patterns for stretched and unstretched rubber in question **29d** were produced using electrons.

Question

Question 29 extends the ideas of demonstration J16 to X-ray crystallography.

Reading for students

The article 'Imaging' in the Reader *Particles, imaging, and nuclei.*

Reading for teachers

HARBURN *et al.*, *Atlas of optical transforms.*
TAYLOR, *Images.*

HOLOGRAPHY

Only a small extension of the ideas already developed is needed to offer an outline explanation of the making of a hologram and how a three-dimensional image can be reconstructed. It would be a pity if students were denied an introduction to this spectacular application of superposition patterns – one which promises to have important consequences for information technology and communication. Since the principles of holography may be unfamiliar to some teachers, we deal with these in rather greater detail than the single period suggested for presenting the topic might seem to justify. The treatment should be light, and if students do not grasp the explanation readily, it should not be laboured. A student who appreciates the peculiarity and beauty of a holographic image and sees that there is in its operation a link with the diffraction theory, will have gained most of what is intended.

Production of a hologram

Producing the diffraction image of a three-dimensional object about the size of a chess piece is impracticable with a school laser, because of the laser's low power and because of the rigorous mechanical conditions needed to produce constant path differences of the order of a wavelength of light between two widely separated beams. The outline given here shows how a hologram may be made *in principle*; a schematic arrangement of apparatus is shown in figure J33.

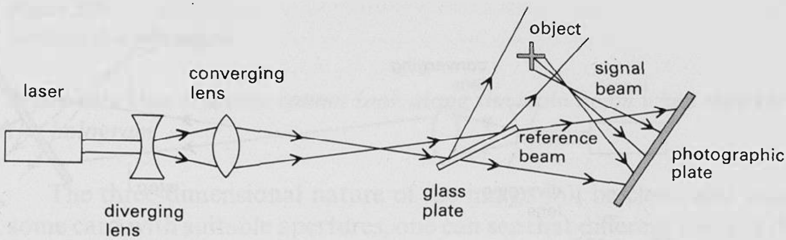


Figure J33
Production of a hologram (not to scale).

The piece of plane glass splits the laser beam into two divergent beams. One (the reference beam) travels to the photographic plate; the other (the signal beam) illuminates an object positioned to one side of the plate, not very far from it. Radiation scattered by the object superposes with the reference beam at the plate. In making a hologram care has to be taken that the reference and signal beams have similar intensity, so that the resulting fringes have clear maxima and minima. (For this reason the transmitted beam at the splitter is usually used to illuminate the object, but the geometry in figure J33 is more straightforward.)

Each illuminated point on the object acts as a source of spherical secondary wavelets which radiate outwards and superpose with the reference wave at the photographic plate. There is a sense in which the superposition pattern on the plate is a diffraction image of the object together with additional information about its orientation relative to the reference beam. Light from each point on the object is scattered to all points on the photographic plate, so each part of the hologram contains information about the complete object, though viewed from a different angle.

The plate is developed and a positive made so that maxima in the superposition pattern appear as apertures in the hologram.

Reconstructing the image

A reference beam, similar to that used to make the hologram, is shone on the hologram. A good deal of light is just transmitted, but some is diffracted as shown in figure J34. If the hologram is viewed from one side of the reference beam a virtual image of the object may be seen in three-dimensional detail. By moving the eye to look through different parts of the hologram, the object may be viewed from different perspectives.

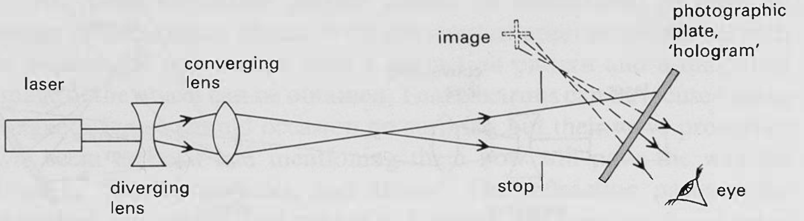


Figure J34
Reconstruction of an image from a hologram (not to scale).

We now suggest reconstructing the image of a ready-made transmission hologram so that students can see some of the effects described above.

DEMONSTRATION
J17 Reconstructing the image from a hologram

ITEM NO.	ITEM
1505	laser
24	hand lens (or diverging lens, -20D)
1521/C	converging lens, $+10$D
1153	sheet of card to act as a stop
1167/1H	translucent screen
1521/F	hologram

Use of laser: See the 'Safety note' on page 207. In this demonstration special care is needed to make sure that students do not look into the beam.

Use either the hand lens or the -20D lens to diverge the laser beam. Adjust the converging lens so that the focused point of light is produced not far from it and a beam large enough to illuminate most of the hologram is formed on the translucent screen. Insert the hologram, *leaving the screen in place to prevent students looking down the main beam.* Refer to the manufacturer's intructions for the recommended viewing angle. Students should look in from one side of the screen, the distance of the screen from the hologram being adjusted to make this convenient.

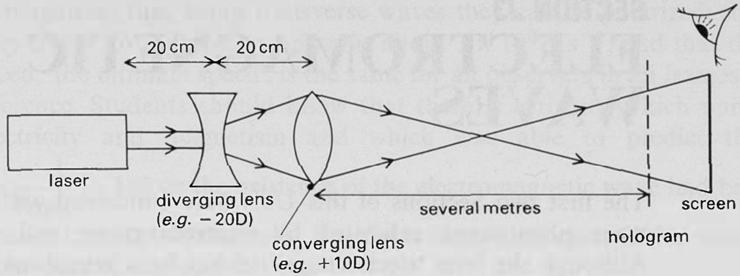


Figure J35
Looking at a hologram.

Make sure that students cannot look along the main beam when they view the hologram.

The three-dimensional nature of the image will be clear, and using some card with suitable apertures, one can see that different parts of the hologram give different views of the object.

The production and reconstruction of white-light holograms involves a more complex technique, but such holograms have the great advantage of not requiring a source of long coherence-length radiation.

Acoustical holography is at present an experimental technique, but its use for forming images at ultrasonic and audible frequencies is likely to grow.

The high density of information storage possible with a hologram is already stimulating research in the communications industry and has even been advanced by neurologists as an analogy of the way the human brain may record information.

Question

Question 30 deals with the formation of a hologram in more detail.

Reading for students

The article 'Imaging' in the Reader *Particles, imaging, and nuclei.*
'Some applications of holography' (*Students' guide* page 174).

Reading for teachers

HECHT and ZAJAC, *Optics.* Chapter 14.
HOLLIGAN, *An introduction to holography.*
TAYLOR, *Images.*
WALKER, *Light and its uses.*

SECTION J3
ELECTROMAGNETIC WAVES

The first two Sections of this Unit were concerned with a variety of wave phenomena exhibited by electromagnetic and other waves. Although the term 'electromagnetic' has been introduced, no attempt has been made to justify it, and the link between radio waves, microwaves, and light has been only phenomenological. We have not yet shown that these waves have any more in common than, say, light waves, sound waves, and water waves.

Some difficulties

Teaching about the nature of electromagnetic radiation is difficult, for several reasons. But the topic is important, and an attempt to convey something of both its theoretical and practical importance seems worth while at A-level.

Some of the difficulties concern the abstract nature of the topic. Students who are happy enough with the idea that electric and magnetic fields are caused by stationary and moving charges respectively, may find the idea of E- and B-fields in the absence of any nearby charges hard to accept. Another difficulty of the theory has to do with causality. In teaching about 'tied' waves in a waveguide, or the free wave, we should avoid giving the impression that the changing E-field *causes* a changing B-field which in turn *causes* a changing E-field and so on. A better statement would be that both are part of the same phenomenon.

Electromagnetic waves can be detected using differently shaped aerials: coils, as in a ferrite rod aerial, which lend themselves most directly to an explanation in terms of B-fields; and dipoles, where the direct explanation is in terms of E-fields. The wavelength often determines which of the two types is practically convenient, but we should avoid implying that a particular aerial detects one of the fields uniquely.

Aims

The aims of this Section are necessarily modest. We think that students should know that in general the origin of the electromagnetic wave is the acceleration of charge and, in particular, the oscillation of charge in a conductor such as a radio aerial or dipole; that the electromagnetic wave allows us to explain the 'action-at-a-distance' effect of an accelerating charge on another charge; that electromagnetic waves of vastly different wavelength are transverse waves in which associated E- and B-fields are perpendicular to each other and to the direction of propagation; that being transverse waves they can be polarized; that they travel through empty space at about $3 \times 10^8\,\mathrm{m\,s^{-1}}$; and that this speed, 'the ultimate speed', is the same for all observers in all frames of reference. Students should know that there *is* a theory which unites electricity and magnetism and which was able to predict that $c=\dfrac{1}{(\varepsilon_0\mu_0)^{1/2}}$ before the existence of the electromagnetic wave had been demonstrated experimentally. But we certainly do *not* intend that theory to be part of this course.

We hope the treatment outlined below will suggest useful strategies and will help teachers to overcome some of the difficulties discussed above.

'Tied' and 'free' waves

The first suggested demonstration is of a 'tied' or guided wave which is associated with charges on a conductor, so that a link can be made with earlier work. A free wave produced by the same source is shown to have similar properties, suggesting that the transverse geometry of the E- and B-fields might be common to both.

DEMONSTRATION
J18a Guided or 'tied' wave

ITEM NO.	ITEM
1050	15 cm dipoles and oscillator
1101	sensitive galvanometer
52K	2 crocodile clips
	either
504	2 long retort stand rods
	or
501, 1153	2 metre rules covered with aluminium foil
1000	leads

The 'waveguide' can be two long retort stand rods, or metre rules with aluminium foil strips stuck to their inner faces. These are supported at the right height and spacing to take the receiving dipole and diode unit from the 15 cm kit, between, but not touching them (figure J36). Slots melted in expanded polystyrene blocks form a useful support which reduces the risk of unwanted reflections. It is worth while making up a short length of coaxial cable, terminated in crocodile clips as shown in figure J36, to connect the oscillator to the waveguide.

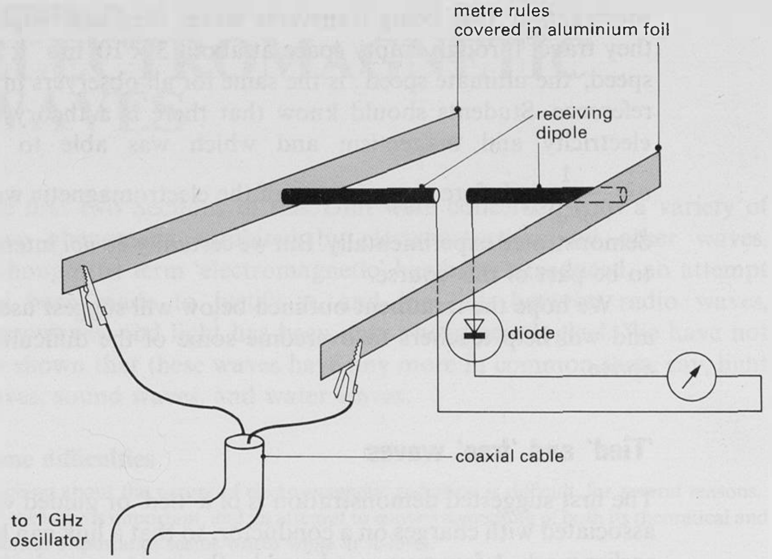


Figure J36
Waveguide made from retort stand rods or foil-covered metre rules.

As the receiving dipole is moved along between the conductors, the galvanometer reading reaches a maximum value every 15 cm or so. Explain that the oscillator generates an alternating p.d. of frequency about 10^9 Hz.

Questioning should elicit that we are dealing with a *standing wave* phenomenon, probably caused by *reflection* of the wave at the far end of the conductors; that the current in the galvanometer is caused by a p.d. between the ends of the receiving dipole. Without the diode the galvanometer gives no response – so the p.d. is alternating and may be associated with *changing concentration of charges* at the antinodes of the standing wave pattern (figure J37).

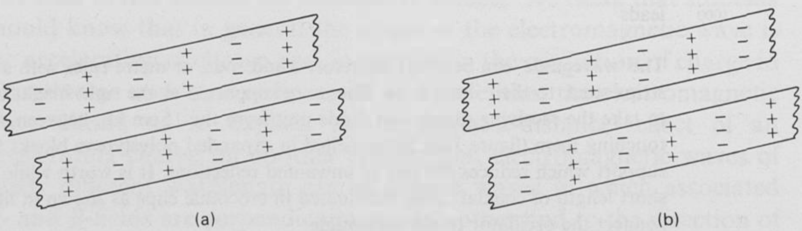


Figure J37
(a) Charge concentrations for tied standing wave.
(b) Half a period later.

With an assumed frequency of 10^9 Hz for the oscillator, a 15 cm separation of antinodes gives a wavelength of 3.0×10^{-1} m and a wave velocity of 3×10^8 m s^{-1}.

The wave is associated with the changing charge concentration in the conductors. It might be useful to discuss the geometry of the electric field which charge concentrations on the rods or strips would produce, and the likely existence of a *B*-field at right angles to it, associated with movement of charge in the conductors. But unless such questions arise, it is probably best to pass straight to the second part of the demonstration. The threads can be taken up again in demonstration J19, when the matter of how the wave is reflected can also be considered.

Questions

Question 35 is related to this demonstration. Question 36 extends the ideas to a 'tied' electromagnetic pulse.

DEMONSTRATION
J18b Free waves

Apparatus as for demonstration J18a plus

1153 large metal sheet or mirror, about 30 cm × 30 cm (for example capacitor plate, item 1025)

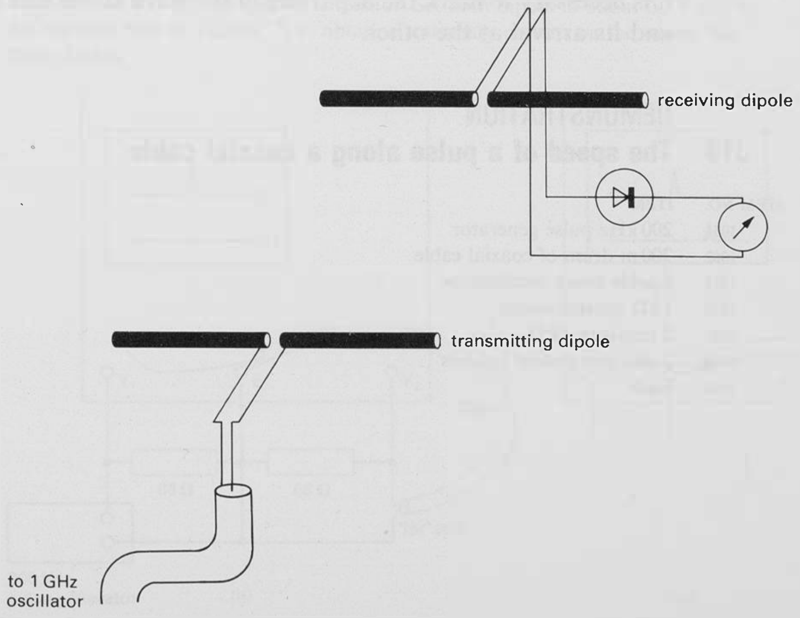


Figure J38
Detection of free travelling wave.

Disconnect the 'waveguide' and connect the dipole transmitter to the oscillator (figure J38).

The waveguide has been removed, but the receiving dipole still detects a strong signal. Charges in the transmitting dipole are accelerated by the rapidly alternating p.d. from the oscillator, and the charges in the distant receiving dipole are affected by the radiation from these accelerated charges.

The standing wave behaviour has disappeared because there is no longer a reflected wave. Students should be able to suggest that a metal plate might make a suitable reflector. When this is tried, a standing wave is formed and the separation of the antinodes is found to be the same as in demonstration J18a, indicating that although the 'free' wave is not associated with the nearby conductors of a waveguide, it has the same wavelength and hence speed, as the 'tied' wave.

To measure a wave speed of about $3 \times 10^8\,\mathrm{m\,s^{-1}}$ accurately by direct timing requires a rather long distance of travel if the time interval is to be measurable using the best instrument available in the school laboratory, the cathode ray oscilloscope. It is convenient to use a 'tied' wave in a conductor which can be coiled round to save space – a drum of coaxial cable. Pulses are used rather than a continuous wave, because it is necessary to mark the departure of the wave at one end of the cable and its arrival at the other.

DEMONSTRATION

J19 The speed of a pulse along a coaxial cable

ITEM NO.	ITEM
1031	200 kHz pulse generator
1062	200 m drum of coaxial cable
1511	double-beam oscilloscope
1510	1 kΩ potentiometer
1151	2 resistors, 68 Ω
1040	2 clip component holders
1000	leads

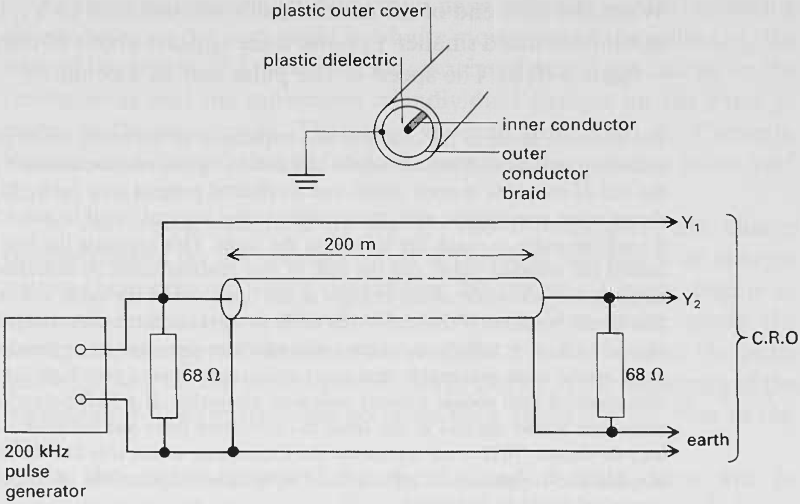


Figure J39
Speed of pulse in a cable.

Connect the 200 kHz pulse generator across a 68 Ω resistor and across the Y_1 input of the oscilloscope. *Connect both ends of the outer sheath of the coaxial cable to earth* (figure J39). Connect the far end of the inner conductor of the coaxial cable to the Y_2 input, across which is another 68 Ω resistor. Set both inputs to about 0.2 V div^{-1}. Set the time base to 1 μs div^{-1}, to obtain a stationary pattern on the screen. See figure J40(a).

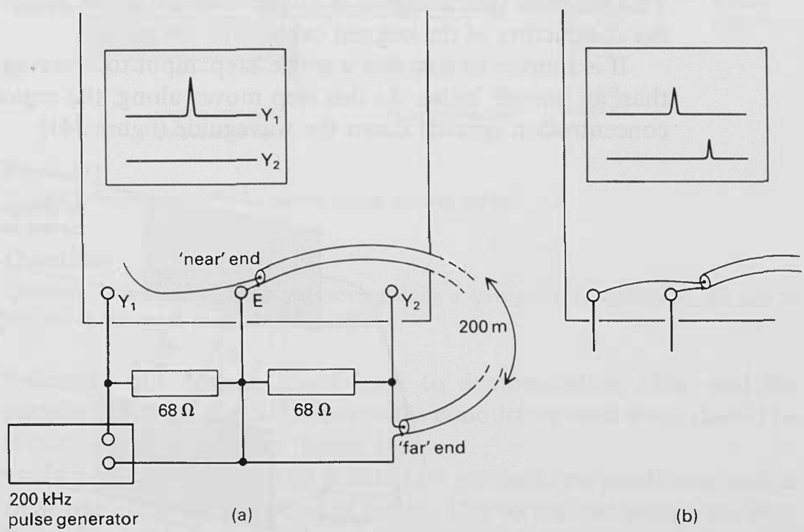


Figure J40
Two-beam display of the speed of a pulse along a cable.

When the near end of the coaxial cable is connected to Y_1, the Y_1 pulse diminishes and a smaller Y_2 pulse trace appears about 10 mm to the right – figure J40(b). The speed of the pulse can be calculated.

The resistors, equal to the characteristic impedance of the cable, are to prevent confusing reflections from both ends of the cable. If the resistor across Y_2 is removed, the end of the cable is open circuit and a reflected pulse is seen on Y_1. It is positive, shows some attenuation, and returns after a time interval equal to twice the time that it took the pulse to reach the far end of the cable. This supports the idea that it is indeed the reflected pulse, and the lack of sign reversal could be discussed both in terms of the behaviour of the charges at the open end of the cable and reflection at a 'less dense' medium. If the end of the cable is short-circuited, then the pulse suffers a change of sign on reflection, which could again be discussed as outlined above.

It would seem reasonable that some resistance between very high (open end) and zero (shorted end) should prevent reflection altogether. If a 1 kΩ potentiometer is connected across the end of the cable the transition from one behaviour to the other may be shown. 70 Ω or so represents the impedance which acts like an infinite length of cable and suppresses all reflections. The broader implications of 'impedance matching' could be indicated.

The speed obtained by calculation will depend on the type of coaxial cable used. Cables with a 'honeycomb' insulator will give values close to $3 \times 10^8\ \mathrm{m\,s^{-1}}$, solid insulator varieties will give a speed which could be as low as $2 \times 10^8\ \mathrm{m\,s^{-1}}$. The obvious differences between this and demonstration J18a are the geometry of the conductors and the medium between them.

The oscilloscope detects a sudden p.d. (on Y_1) when the pulse leaves the generator and another (on Y_2) when it arrives at the far end of the cable. This suggests that a region of charge concentration moves rapidly on the conductors of the coaxial cable with the pulse.

It is simpler to consider a single 'step' input to a 'waveguide' rather than an 'on-off' pulse. As this step moves along, the region of charge concentration spreads down the waveguide (figure J41).

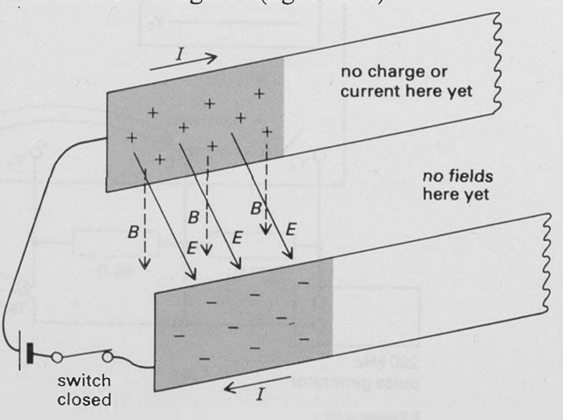


Figure J41
E- and *B*-fields as step pulse moves along waveguide.

In these discussions students may have difficulty in distinguishing clearly between the very rapid but finite movement of the pulse (*i.e.*, the edge of the region of *E*- and *B*-fields associated with a net charge on the conductors), and the movement of individual charges on the wires or plates of the waveguide. They should recall from Unit B, 'Currents, circuits, and charge', that individual charges in a conductor move very slowly, at about $10^{-6}\,\mathrm{m\,s^{-1}}$.

In the region occupied by the *E*- and *B*-fields there are charge concentrations on the conductors. It is the slow movement of charges within these regions which constitutes the current. Charge begins to move when the signal wavefront reaches it. One metre along the waveguide charge will begin to move about 3×10^{-9} s after the pulse leaves the source. But an individual charge carrier at the beginning of the waveguide will have travelled no more than about 3×10^{-15} m in this time.

In the region occupied by the *E*- and *B*-fields there will be concentrations of charge on the conductor and movement of charge within the charged regions (figures J41 and J42). The *E*- and *B*-fields are at right angles to each other at all points and perpendicular to the direction of travel of the pulse.

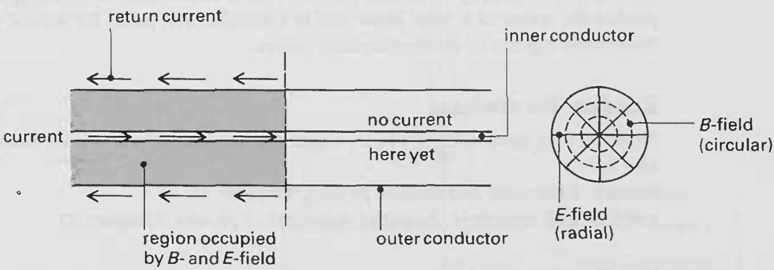


Figure J42
E- and *B*-fields as step pulse moves along coaxial cable.

Questions

Question 36 is about a step pulse moving along a waveguide. Questions 37, 38, and 39 are about the speed of signals in a cable.

Reference can now be made back to demonstration J18a, and the pattern of *E*- and *B*-fields between the conductors used there should be sketched after discussion (figure J43).

In both cases the *E*- and *B*-fields are mutually perpendicular and at right angles to the direction of travel. The waves and pulses we have been discussing consist of travelling *E*- and *B*-fields, and they are *transverse*. This geometry will be important in the circus of polarization experiments, J21, to which one could proceed immediately after the next demonstration.

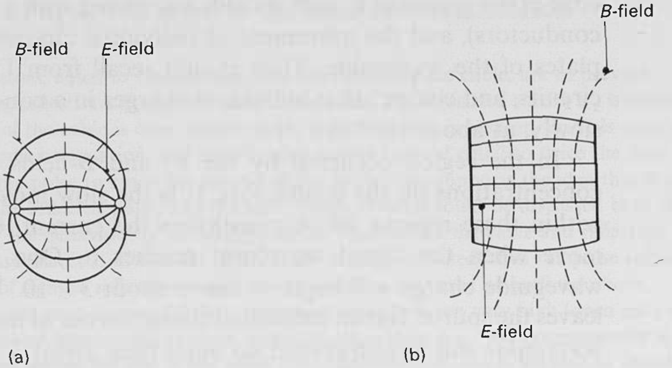


Figure J43
E- and *B*-fields between (a) parallel wires or rods and (b) parallel plates.

Questions

Question 40 is about the inverse-square law. Question 41 asks students to compare two ways of transmitting a television programme. Question 42 is a learning question leading to $c = 1/(\varepsilon_0\mu_0)^{1/2}$. It uses previous ideas about electric and magnetic fields to predict the speed of a 'tied' pulse and is a suitable end point for school work on the theoretical aspects of electromagnetic waves.

Reading for students

Students who want a slightly fuller treatment can be referred to their text books and also to:

BENNET, *Electricity and modern physics.* Chapter 15.
WHELAN and HODGSON, *Essential principles of physics.* Chapter 37.

Reading for teachers

FEYNMAN *et al.*, The Feynman lectures on physics *Volume 2.* Chapter 24.
SHIVE and WEBER, *Similarities in physics.* Chapter 8.

Although students may recognize $3 \times 10^8\ \mathrm{m\,s^{-1}}$ as the speed of light, they probably will not have seen it measured.

THE SPEED OF LIGHT

All direct methods for measuring the speed of light depend on being able to divide a beam of light into short pulses, and measuring the time of flight of one pulse over a known distance. Formerly this division was done mechanically, for example by a toothed wheel or a rotating mirror. A more modern method uses electronics to produce short pulses of light, and an oscilloscope to measure time.

Question

Question 43 is about an experiment which measures the speed of light by dividing the beam mechanically.

DEMONSTRATION

J20 Speed of light

ITEM NO.	ITEM
1523	apparatus to measure the speed of light
1511	double-beam oscilloscope, 0.1 μs div^{-1} or better
1033	2 cell holders with 6 cells (or other 9 V d.c. supply)
1000	leads

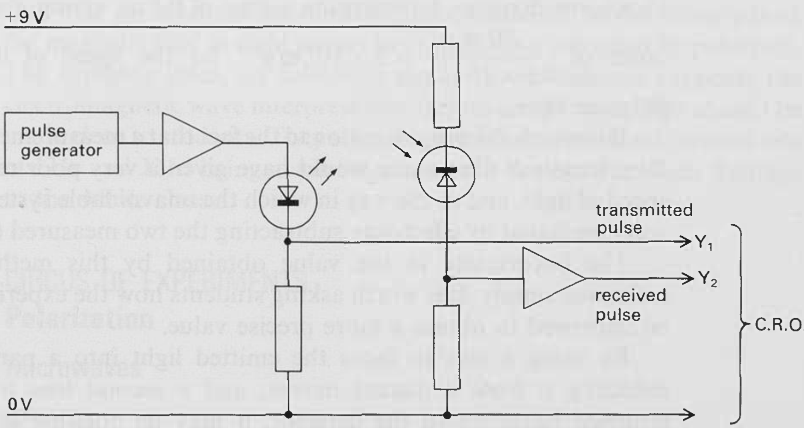


Figure J44
Measurement of the speed of light.

Set the time-base speed to 0.1 μs div^{-1} (probably 0.5 μs div^{-1} and '× 5'), with 'X-gain' or 'variable sweep' in the 'calibrated' position. Trigger the sweep from the transmitted pulse on Y_1, and adjust the Y_1 gain appropriately.

Item 1523 pulses a light-emitting diode at 1 MHz. The pulses of light emitted are fed via a fibre optics cable to the detector, a photodiode.

The transmitted pulse and the received pulse are displayed on the two traces of the double-beam oscilloscope.

Connect the output marked 'received pulse' to Y_2: there may be a signal on Y_2 before the optical fibre is connected, due to electrical pick up between the transmitting and receiving sides of the apparatus. It is important to be able to distinguish this signal from the received pulse due to light striking the detector.

Now connect the short (about 10 cm) length of optical fibre between the transmitting L.E.D. and the receiving photodiode. The received pulse will be delayed by about 0.2–0.3 μs. This time delay is the sum of the pulse's travel time plus the photodiode's response time and any delays in the electronic circuits, for instance the amplifier which amplifies the photodiode's output before it is sent to the oscilloscope. Use a water based felt-tipped pen to mark the position of the received pulse on the screen.

Now replace the short length of optical fibre by the long one (20 m). The pulse of light arriving at the detector may be attenuated by this length of fibre, making it necessary to adjust the Y_2 gain. The shift in position of the peak of the Y_2 pulse represents the extra time it takes the light pulse to travel through 20 m of optical fibre. The peak shifts by about one division, representing a time of 0.1 μs, giving a value of the order of $\frac{20\,\mathrm{m}}{0.1 \times 10^{-6}\,\mathrm{s}} = 2 \times 10^8\,\mathrm{m\,s^{-1}}$ for the speed of light in the polymer fibre.

It is worth drawing attention to the fact that a measurement using the 20 m length of fibre alone would have given a very poor result for the speed of light, and to the way in which the unavoidable systematic error was eliminated by effectively subtracting the two measured time delays.

The uncertainty in the value obtained by this method can be estimated simply. It is worth asking students how the experiment could be improved to obtain a more precise value.

By using a lens to focus the emitted light into a parallel beam, reflecting it from a distant mirror, and a second lens to focus the returned beam on to the detector, it may be possible to adapt the apparatus to measure the speed of light in air.

See also RUTHERFORD, S. J. 'Speed of light in a fibre' in *Optical fibres in school physics.*

Television or video

Two programmes in the Granada Television series *Experiment: Physics*, in which students take measurements from a filmed experiment are relevant. 'The determination of the velocity of light' uses a rotating mirror method; in 'The determination of the velocity of radio waves' students note the positions of nodes in a standing wave on Lecher wires.

All electromagnetic waves travel at the same speed

'Free' and 'tied' electromagnetic waves, including light, have been shown to travel with the same speed (within experimental uncertainty). Some students will have tried question **42**, which shows that theory predicts the speed of a tied wave in 'free space' to be $1/(\varepsilon_0\mu_0)^{1/2}$, and there are of course good theoretical reasons for expecting there to be a family of electromagnetic waves travelling in free space at that speed. (A diagram of wavelengths and frequencies in the electromagnetic spectrum is given in figure J28 of the *Students' guide*.) All students should verify that the expression has the dimensions of speed and some could be referred to an elementary account of Maxwell's equations and their solution, such as that in WENHAM *et al., Physics: concepts and models.*

The speed of electromagnetic waves has been measured at many wavelengths. High accuracy is easier to achieve at some wavelengths than others: see *Students' guide* table J1 (page 167).

Polarization of electromagnetic waves

The model developed previously for electromagnetic radiation is of a transverse wave (for example *Students' guide* figure J32). The electric field may be in any of the directions perpendicular to the velocity (and the magnetic field at right angles to that). Such a wave can be polarized. The evidence from the following circus of experiments supports the electromagnetic wave interpretation, it does not prove it. This should be made clear to students. They should work in twos or threes, several sets of J21c and J21d being available, and could demonstrate their findings to the whole class.

CIRCUS OF EXPERIMENTS

J21 Polarization

J21a Microwaves

ITEM NO.	ITEM
184/1	microwave transmitter
184/2	microwave receiver
181	general purpose amplifier
183	loudspeaker (if not part of item 181)
1524	polarization indicating grille

With the grille absent, the detected signal gradually reduces in intensity as the transmitter and receiver are rotated relative to each other until no signal is detected when they are 'crossed' at 90°. This indicates that there is 'directionality' associated with both the wave and the receiver.

If the grille is inserted with the transmitter and receiver both 'upright', as shown in figure J45, the received signal is zero. But the signal grows to full strength as the grille is rotated until the metal rods are horizontal. In this position they seem not to interact with the wave.

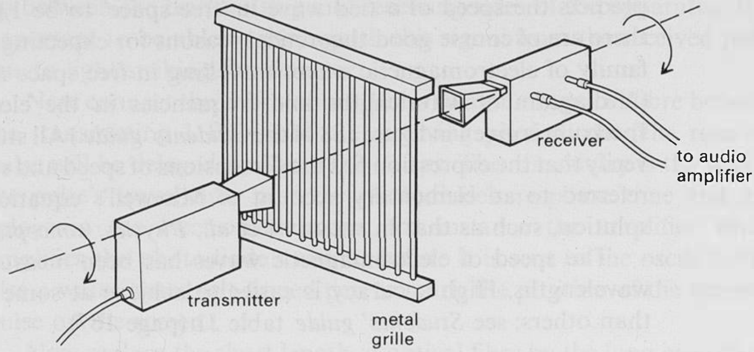


Figure J45
Polarization of microwaves.

A possible explanation is that the transmitter produces a wave with the electric field vertical and the receiver only detects the vertical component of the field, so that when they are 'crossed' (at 90° to each other) nothing is detected. The wave is said to be *plane polarized*, with the electric vector vertical. When the electric field of the wave is parallel to the rods, it is able to set the conduction electrons in the metal into oscillation, and the rods reflect the incident wave. But when the electric field vector and the rods are at right angles, this effect is vanishingly small, and the wave is transmitted.

If the transmitter and receiver are 'crossed', then a maximum signal can be received when the grille is rotated to 45°. The grille effectively 'rotates the plane of polarization' of the wave.

J21b 30 cm or gigahertz waves

ITEM NO.	ITEM
1050	15 cm dipoles and oscillator
1153	several 15 cm lengths of brass rod, about 4 mm diameter
1101	sensitive galvanometer
1153	40 cm square of hardboard (see note below)
1000	leads

The receiving dipole (with its diode) is connected to the galvanometer. All the effects seen in experiment J21a may be observed. The rod used in place of the grille may either be a piece of brass rod 15 cm long, about 4 mm in diameter; or a short retort stand rod; or two *metal* rods from dynamics trolleys joined end-to-end by tape.

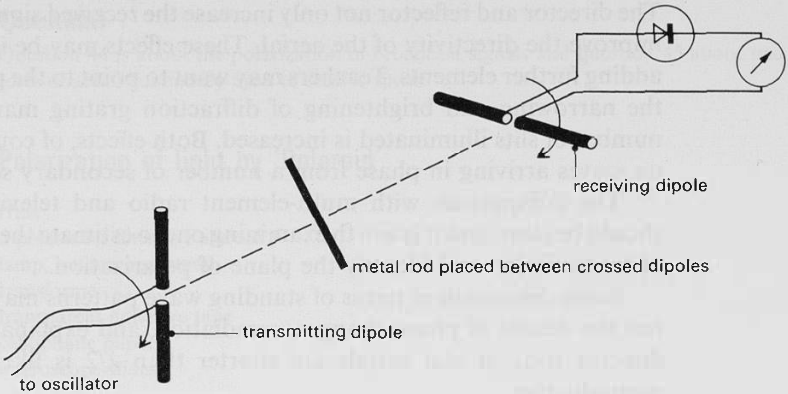


Figure J46
Polarization of 30 cm waves.

Aerials and the reception of radio waves

A small extension can give an insight into the design of aerials for the reception of radio and television signals.

Stand the receiving dipole – figure J47(a) – at the centre of a suitable sheet of material (*e.g.* hardboard) so that it and other metal rods can be rotated together as a unit. Put the transmitter about 1 m away on the edge of another bench. If two gigahertz kits are available, extra dipoles may be mounted in the special holders. Otherwise use a brass rod, with the transmitting and receiving dipoles vertical and the brass rods mounted vertically in large corks.

With the transmitting and receiving dipoles oriented so that the galvanometer detects the maximum signal, a rod (a reflector) is brought up behind the receiver, as in figure J47(b). The signal increases to a maximum when the reflector is about 7 cm ($\approx \lambda/4$) behind the receiver.

Bringing another rod up in front of the receiver (a director) can increase the signal still further – figure J47(c). In all these adjustments, the dipoles are best moved with a metre rule to avoid confusing reflections.

The whole array (director, receiving dipole, and reflector) is now rotated until the signal falls to one fifth the maximum value. If the extra rods are now removed the signal increases.

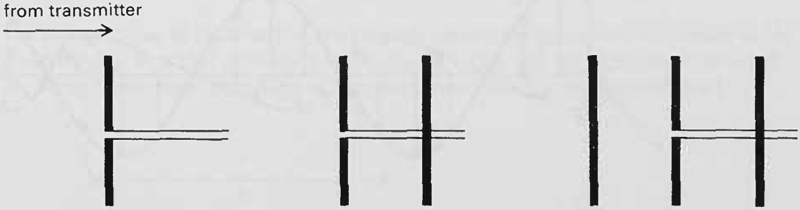


Figure J47
Elements of an aerial.

The director and reflector not only increase the received signal, they also improve the directivity of the aerial. These effects may be increased by adding further elements. Teachers may want to point to the parallel with the narrowing and brightening of diffraction grating maxima as the number of slits illuminated is increased. Both effects, of course, depend on waves arriving in phase from a number of secondary sources.

The comparison with multi-element radio and television aerials should be clear, and it is worth examining one to estimate the wavelength of transmission and identify the plane of polarization.

Some discussion in terms of standing wave patterns may be helpful, but the details of phase changes, reradiation, and explanation of why director rods in real aerials are shorter than $\lambda/2$ is likely to prove unproductive.

It is also worth examining and discussing the aerial systems of a v.h.f./medium wave radio. For receiving v.h.f. (about 100 MHz, or 3 m), a telescopic metal aerial – equivalent to one half of a dipole – is used. For medium wave (about 1 MHz, or 300 m), a dipole would clearly be impracticable. Instead, a coil of a few dozen turns wound on a ferrite rod is used. The coil and the ferrite rod (used because large field strengths are easily produced without damping eddy currents) suggest that the magnetic field of the wave is being detected. The orientation of the coil's axis and the half dipole at right angles to the direction of travel is consistent with a transverse wave. (Note, however, that the loop of a portable television is generally a 'folded dipole', giving a maximum signal when oriented with its axis towards the transmitter, and so is most easily considered an E-field detector.)

A picture of what a polarized radio wave may be like has now developed and figure J48, showing the wave frozen at a given moment, is a useful mental image to leave students with.

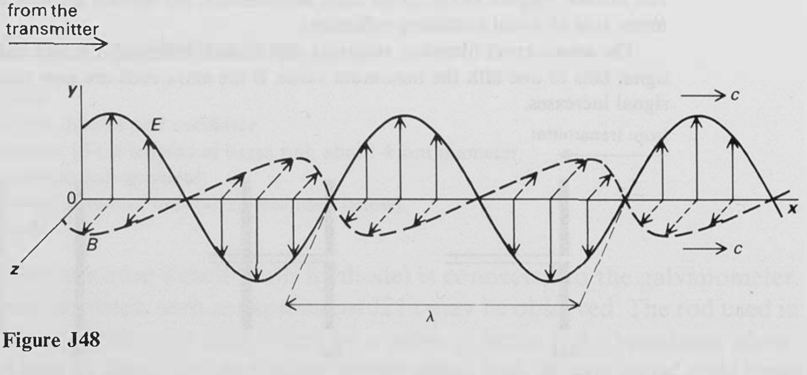


Figure J48

Questions

Question 44 is about the polarization of broadcast signals and question 45 about two sorts of aerial commonly used to receive them.

J21c Polarization of light by 'Polaroid'

ITEM NO.	ITEM
1183	3 polarizing filters ('Polaroid')
94A	lamp, holder, and stand
27	transformer
1153	transparent adhesive tape
1153	cellophane paper
1155	microscope slides

Explain that a single sheet of Polaroid is the optical equivalent of the metal grille used with microwaves. Ultra microscopic polarizing crystals are contained in a transparent film which is stretched to align the crystals. Only light with its electric vector in a fixed direction is transmitted, any component of the electric vector parallel to the length of the crystals being absorbed. The 'transmittance ratio' for these two components can be as high as 10^5.

When the lamp is viewed through a single sheet, there is some reduction in intensity due to absorption, but the intensity remains unchanged as the sheet is rotated, indicating that light leaving the lamp is not polarized. If a second sheet is held in front of the first and is rotated, the filament will disappear almost completely twice per rotation, indicating that light is plane-polarized after passing through the first sheet.

If the two sheets are 'crossed' so that no light is transmitted, and a third sheet rotated between them, the filament will be seen, dimly, four times during each rotation. This third sheet is 'rotating the plane of polarization' by passing a component of the polarized wave which leaves the first sheet in the direction which can be passed by the 'crossed' sheet. See figure J50.

The phrase 'plane of polarization' used in older books on optics is best avoided as the plane referred to is that containing the magnetic vector and the direction of travel of the wave. Most important effects of electromagnetic waves are associated more directly with the electric vector and it is now usual practice to describe a wave as polarized with the electric vector in a certain direction. The directions of the double-headed arrows on the sheets of Polaroid illustrated in figures J49 and J50 indicate the orientation of the electric vector passed by the sheet, and the solid-headed arrows on the rays show the direction of the electric vector of the wave.

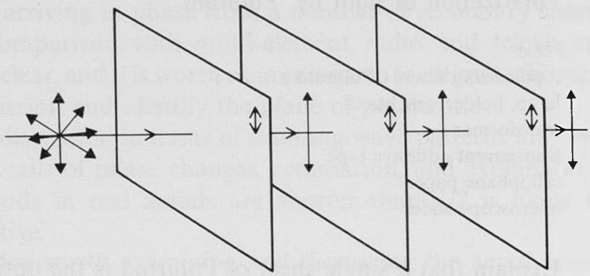


Figure J49

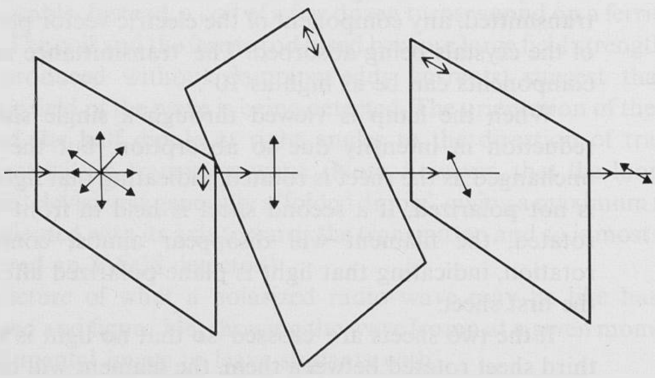


Figure J50

Students may already have seen the effect of stretching polythene between crossed Polaroids (Unit A, 'Materials and mechanics') and could now extend their experience by looking at cellophane wrapping material, transparent adhesive tape, and various mineral crystals. These are worth seeing for the beautiful colours produced and for an indication of how crystallographers and mineralogists can use the effect to make diagnostic measurements.

J21d Polarization of light by reflection and by scattering

ITEM NO.	ITEM
1183	2 polarizing filters ('Polaroid')
100/1	rectangular plastic tank
94A	lamp, holder, and stand
94B	housing shield
94G	2 barriers
94H	plano-cylindrical lens, + 7D
27	transformer
1153	small sheets of glass, polythene, and metal (*e.g.* aluminium foil)
	milk (or powdered milk) should be available; each experiment requires only a drop or two

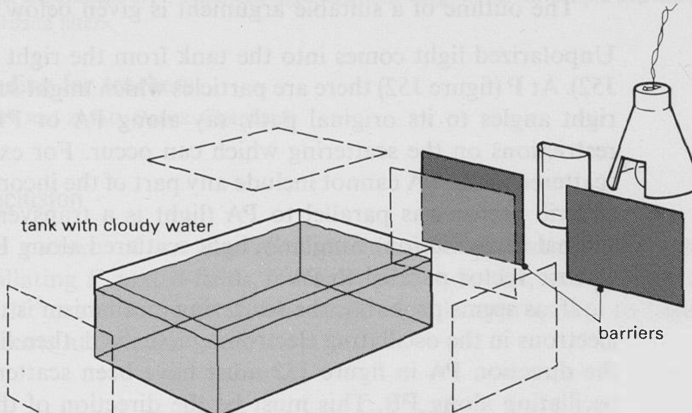


Figure J51
Scattering by cloudy water. (Broken rectangles show positions for a Polaroid filter.)

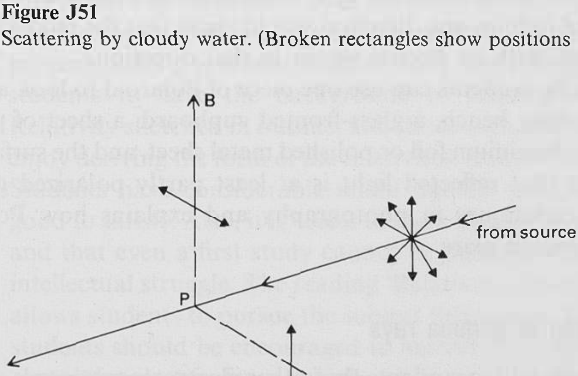


Figure J52
Polarization by scattering.

Students arrange the lamp, lens, and barriers to send a bright, narrow beam through the tank of water. Tap water may contain enough suspended matter to make the beam visible; if not, add a drop or two of milk. It is essential not to add too much or the light will be scattered more than once before emerging, and the argument below will not hold.

Students hold and rotate a Polaroid filter in the various positions indicated in figure J51. They should see that the transmitted light is not polarized, while that emerging from the top and sides of the tank *is* polarized. They may be able to explain this, given their understanding of radio waves, and that light is a transverse wave.

The outline of a suitable argument is given below.

Unpolarized light comes into the tank from the right (figures J51 and J52). At P (figure J52) there are particles which might scatter the light at right angles to its original path, say along PA or PB, but there are restrictions on the scattering which can occur. For example, the light scattered along PA cannot include any part of the incoming light whose electric vector was parallel to PA (light is a transverse, not a longitudinal wave motion). Similarly, light scattered along PB must have its electric vector parallel to PA.

If, as seems probable, the scattering mechanism is the oscillation of electrons in the oscillating electromagnetic field, then light travelling in the direction PA in figure J52 must have been scattered by electrons oscillating along PB. This must be the direction of the *E*-field of the light travelling along PA. The tank of cloudy water could, if one wished, be used to detect polarized light without using Polaroid. The absence of scattered light in one direction would show that the incoming light was polarized, with its electric vector in that direction.

Finally, students can use one piece of Polaroid to look at reflections from a shiny bench, a glass-fronted cupboard, a sheet of polythene, a piece of aluminium foil or polished metal sheet, and the surface of water. The fact that reflected light is at least partly polarized gives rise to special techniques in photography and explains how Polaroid sunglasses reduce glare.

Scattering of gamma rays

It is worth telling students that other electromagnetic waves, including X-rays and gamma rays, can be polarized by scattering, and their polarization can be detected even though there is no equivalent to Polaroid for such waves. If the tank is replaced by a target containing suitable atomic nuclei, the gamma rays scattered by the nuclei can then be scattered again from a second target.

The variation in intensity of gamma rays scattered in different directions gives information about the polarization on scattering, and so will yield information about how charges in the nuclei respond to electromagnetic forces.

Questions

Question 46 is about the intensity of light transmitted by a polarizing filter in various orientations; question 47 is a coded answer question about polarized light.

Home experiment

Home experiment JH4, Polarized light, contains suggestions for simple activities using polarizing filters.

Reading for teachers

HECHT and ZAJAC, *Optics.* Chapter 8.

Conclusion

The picture of an electromagnetic wave consisting of transverse, oscillating E- and B-fields, mutually perpendicular, and propagating at $3 \times 10^8 \, \mathrm{m\,s^{-1}}$ *in vacuo* will be as far as it is appropriate to take most students at A-level.

THE SPECIAL THEORY OF RELATIVITY

The surprising consequences of the finite speed of propagation of light, and the fact that it is constant as measured in all frames of reference, properly belong to a later stage of study. But it might be useful for some students to have the background to Einstein's Special Theory of Relativity sketched in outline, and those with mathematical ability may enjoy deriving the form of the relativistic transformation for themselves. Students have considerable initial interest in this subject which it is good to satisfy. Also, it is useful to show some of the scope of the theory, and that even a first study cannot be pursued without a great deal of intellectual struggle. The reading 'Relativity' (*Students' guide*, page 177) allows students to pursue the subject themselves. It is suggested that all students should be encouraged to answer the question about the cows along the electric fence which, while not concerned with relativistic effects, shows that the finite speed of an electromagnetic pulse has surprising consequences. There are plenty of good books which students can be referred to if they would like to pursue Special Relativity further.

Further reading for students

BENNET, *Electricity and modern physics.*
BONDI, *Relativity and common sense.*
EPSTEIN, *Relativity revisualized.*
HOFFMANN, *Relativity and its roots.*
LANDAU and RUMER, *What is relativity?*
SANDAGE, 'The red shift'.
SHANKLAND, 'The Michelson–Morley experiment'.

Reading for teachers

FRENCH, *Special relativity.*
KITTEL, KNIGHT, and RUDERMAN, Berkeley Physics Course *Volume 1.*
LILLEY, *Discovering relativity for yourself.*
PURCELL, Berkeley Physics Course *Volume 2.*
SHERWIN, *Basic concepts in physics.*